

*Blackwell  
Companions to  
Philosophy*

# A COMPANION TO EPISTEMOLOGY

*Second Edition*



*Edited by*  
JONATHAN DANCY, ERNEST  
SOSA, and MATTHIAS STEUP

 WILEY-BLACKWELL



# A Companion to Epistemology

## Blackwell Companions to Philosophy

---

This outstanding student reference series offers a comprehensive and authoritative survey of philosophy as a whole. Written by today's leading philosophers, each volume provides lucid and engaging coverage of the key figures, terms, topics and problems of the field. Taken together, the volumes provide the ideal basis for course use, representing an unparalleled work of reference for students and specialists alike.

### Already published in the series:

1. The Blackwell Companion to Philosophy, Second Edition  
*Edited by Nicholas Bunnin and Eric Tsui-James*
2. A Companion to Ethics  
*Edited by Peter Singer*
3. A Companion to Aesthetics, Second Edition  
*Edited by Stephen Davies, Kathleen Marie Higgins, Robert Hopkins, Robert Stecker, and David E. Cooper*
4. A Companion to Epistemology, Second Edition  
*Edited by Jonathan Dancy, Ernest Sosa, and Matthias Steup*
5. A Companion to Contemporary Political Philosophy (two-volume set), Second Edition  
*Edited by Robert E. Goodin and Philip Pettit*
6. A Companion to Philosophy of Mind  
*Edited by Samuel Guttenplan*
7. A Companion to Metaphysics, Second Edition  
*Edited by Jaegwon Kim, Ernest Sosa and Gary S. Rosenkrantz*
8. A Companion to Philosophy of Law and Legal Theory  
*Edited by Dennis Patterson*
9. A Companion to Philosophy of Religion  
*Edited by Philip L. Quinn and Charles Taliaferro*
10. A Companion to the Philosophy of Language  
*Edited by Bob Hale and Crispin Wright*
11. A Companion to World Philosophies  
*Edited by Eliot Deutsch and Ron Bontekoe*
12. A Companion to Continental Philosophy  
*Edited by Simon Critchley and William Schroeder*
13. A Companion to Feminist Philosophy  
*Edited by Alison M. Jaggar and Iris Marion Young*
14. A Companion to Cognitive Science  
*Edited by William Bechtel and George Graham*
15. A Companion to Bioethics, Second Edition  
*Edited by Helga Kuhse and Peter Singer*
16. A Companion to the Philosophers  
*Edited by Robert L. Arrington*
17. A Companion to Business Ethics  
*Edited by Robert E. Frederick*
18. A Companion to the Philosophy of Science  
*Edited by W. H. Newton-Smith*
19. A Companion to Environmental Philosophy  
*Edited by Dale Jamieson*
20. A Companion to Analytic Philosophy  
*Edited by A. P. Martinich and David Sosa*
21. A Companion to Genetics  
*Edited by Justine Burley and John Harris*
22. A Companion to Philosophical Logic  
*Edited by Dale Jacquette*
23. A Companion to Early Modern Philosophy  
*Edited by Steven Nadler*
24. A Companion to Philosophy in the Middle Ages  
*Edited by Jorge J. E. Gracia and Timothy B. Noone*
25. A Companion to African-American Philosophy  
*Edited by Tommy L. Lott and John P. Pittman*
26. A Companion to Applied Ethics  
*Edited by R. G. Frey and Christopher Heath Wellman*
27. A Companion to the Philosophy of Education  
*Edited by Randall Curren*
28. A Companion to African Philosophy  
*Edited by Kwasi Wiredu*
29. A Companion to Heidegger  
*Edited by Hubert L. Dreyfus and Mark A. Wrathall*
30. A Companion to Rationalism  
*Edited by Alan Nelson*
31. A Companion to Ancient Philosophy  
*Edited by Mary Louise Gill and Pierre Pellegrin*
32. A Companion to Pragmatism  
*Edited by John R. Shook and Joseph Margolis*
33. A Companion to Nietzsche  
*Edited by Keith Ansell Pearson*
34. A Companion to Socrates  
*Edited by Sara Ahbel-Rappe and Rachana Kamtekar*
35. A Companion to Phenomenology and Existentialism  
*Edited by Hubert L. Dreyfus and Mark A. Wrathall*
36. A Companion to Kant  
*Edited by Graham Bird*
37. A Companion to Plato  
*Edited by Hugh H. Benson*
38. A Companion to Descartes  
*Edited by Janet Broughton and John Carriero*
39. A Companion to the Philosophy of Biology  
*Edited by Sahotra Sarkar and Anya Plutynski*
40. A Companion to Hume  
*Edited by Elizabeth S. Radcliffe*
41. A Companion to the Philosophy of History and Historiography  
*Edited by Aviezer Tucker*
42. A Companion to Aristotle  
*Edited by Georgios Anagnostopoulos*
43. A Companion to the Philosophy of Technology  
*Edited by Jan-Kyrre Berg Olsen, Stig Andur Pedersen, and Vincent F. Hendricks*
44. A Companion to the Philosophy of Literature  
*Edited by Garry L. Hagberg and Walter Jost*
45. A Companion to Latin American Philosophy  
*Edited by Susana Nuccetelli, Ofelia Schutte, and Otávio Bueno*

### Also under contract:

- A Companion to Schopenhauer  
*Edited by Bart Vandenabeele*
- A Companion to Relativism  
*Edited by Steven D. Hales*

# A Companion to Epistemology

Second Edition

*edited by*

JONATHAN DANCY,  
ERNEST SOSA,

*and*

MATTHIAS STEUP

 **WILEY-BLACKWELL**

A John Wiley & Sons, Ltd., Publication

This second edition first published 2010

© 2010 Blackwell Publishing Ltd except for editorial material and organization © Jonathan Dancy, Ernest Sosa, and Matthias Steup

Edition history: Blackwell Publishing Ltd (1e, 1992)

Blackwell Publishing was acquired by John Wiley & Sons in February 2007. Blackwell's publishing program has been merged with Wiley's global Scientific, Technical and Medical business to form Wiley-Blackwell.

*Registered Office*

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

*Editorial Offices*

350 Main Street, Malden, MA 02148-5020, USA

9600 Garsington Road, Oxford, OX4 2DQ, UK

The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, for customer services, and for information about how to apply for permission to reuse the copyright material in this book, please see our website at [www.wiley.com/wiley-blackwell](http://www.wiley.com/wiley-blackwell).

The right of Jonathan Dancy, Ernest Sosa, and Matthias Steup to be identified as the authors of the editorial material in this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

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.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

*Library of Congress Cataloging-in-Publication Data*

A companion to epistemology / edited by Jonathan Dancy, Ernest Sosa, and Matthias Steup. – 2nd ed.

p. cm. – (Blackwell companions to philosophy)

Includes bibliographical references and index.

ISBN 978-1-4051-3900-7 (hardcover : alk. paper) 1. Knowledge, Theory of. I. Dancy, Jonathan.

II. Sosa, Ernest. III. Steup, Matthias.

BD161.C637 2009

121–dc22

2009020185

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

Set in 9.5/11pt Photina by Graphicraft Limited, Hong Kong

Printed in Singapore

# Contents

---

List of Contributors	xiii
Preface to the First Edition	xix
Preface to the Second Edition	xxiii
 Part I Ten Review Essays	 1
<b>Skepticism and Closure</b> Anthony Brueckner	3
<b>Contextualism</b> Richard Feldman	12
<b>Foundations and Coherence</b> Michael Huemer	22
<b>Recent Work on the Internalism–Externalism Controversy</b> Laurence Bonjour	33
<b>A Priori Knowledge</b> Albert Casullo	43
<b>The Common Sense Tradition</b> Noah Lemos	53
<b>The Power of Perception</b> Peter Markie	62
<b>Virtue Epistemology</b> John Greco	75
<b>Social Epistemology</b> Alvin Goldman	82
<b>Bayesian Epistemology</b> Alan Hájek and Stephan Hartmann	93
 Part II Twenty Epistemological Self-profiles	 107
<b>Robert Audi</b>	109
<b>Laurence Bonjour</b>	114
<b>Stewart Cohen</b>	118
<b>Earl Conee and Richard Feldman</b>	123
<b>Fred Dretske</b>	130
<b>Richard Foley</b>	134
<b>Richard Fumerton</b>	139
<b>Alvin I. Goldman</b>	144
<b>Gilbert Harman</b>	152
<b>Peter Klein</b>	156
<b>Hilary Kornblith</b>	163
<b>Keith Lehrer</b>	168
<b>Alvin Plantinga</b>	173
<b>John Pollock</b>	178
<b>Ernest Sosa</b>	185
<b>Barry Stroud</b>	190
<b>Michael Williams</b>	194
<b>Timothy Williamson</b>	199
<b>Crispin Wright</b>	204
<b>Linda Zagzebski</b>	210

Part III Epistemology from A to Z (Entries without author list cross references)	217
A	219
<b>a priori/a posteriori</b> Albert Casullo	219
<b>abduction</b> Christopher Hookway	221
<b>absurdity</b> Robert S. Tragesser	221
<b>Academy (Plato)</b> Nicholas P. White	221
<b>act/object analysis</b> Michael Pendlebury	222
<b>adverbial theory</b> Michael Pendlebury	222
<b>agnosticism</b> Alvin Plantinga	223
<b>Alston, William P. (1921–2009)</b> Robert Audi	223
<b>analyticity</b> Jerrold J. Katz	224
<b>anamnesis</b> Nicholas P. White	230
<b>antinomy</b> Robert S. Tragesser	230
<b>apodeictic</b> Robert S. Tragesser	231
<b>aporia</b> Robert S. Tragesser	231
<b>apperception</b> David McNaughton	231
<b>Aquinas, Thomas (1225–74)</b> Scott MacDonald	231
<b>argument</b> Albert Casullo	235
<b>argument from analogy</b> Jonathan Dancy	235
<b>argument from illusion</b> Richard Fumerton	236
<b>Aristotle (384–322 bc)</b> Terence Irwin	240
<b>Armstrong, David M. (1926– )</b> Adrian Heathcote	244
<b>association</b> Barry Stroud	245
<b>ataraxia</b> Charlotte Stough	245
<b>Augustine, St (354–430)</b> Scott MacDonald	245
<b>Austin, John L(angshaw) (1911–60)</b> Graham Bird	247
<b>avowals</b> David Pears	249
<b>axiomatization, axiomatics</b> Robert S. Tragesser	249
<b>Ayer, A(lfred) J(ules) (1910–89)</b> Graeme Forbes	250
B	253
<b>Bacon, Francis, Lord Verulam (1561–1626)</b> G. A. J. Rogers	253
<b>basic belief</b> Matthias Steup	253
<b>behaviourism</b> Edward Erwin	254
<b>belief</b> John Heil	254
<b>belief in and belief that</b> John Heil	259
<b>Bergmann, Gustav (1906–87)</b> Laird Addis	260
<b>Berkeley, George (1685–1753)</b> Michael Ayers	261
<b>Blanshard, Brand (1892–1987)</b> Michael Williams	264
<b>Brentano, Franz (1838–1917)</b> R. M. Chisholm	265
<b>burden of proof</b> Douglas N. Walton	266
C	267
<b>Carnap, Rudolf (1891–1970)</b> Wesley C. Salmon	267
<b>Cartesianism</b> John Cottingham	268
<b>causal theories in epistemology</b> Carl Ginet	268
<b>certainty</b> Peter D. Klein	272
<b>Chisholm, Roderick (1916–99)</b> Richard Foley	275



<b>circular reasoning</b>	Douglas N. Walton	277
<b>cogito</b>	John Cottingham	277
<b>coherentism</b>	Keith Lehrer	278
<b>collective belief</b>	Margaret Gilbert	281
<b>commonsensism and critical cognitivism</b>	Noah H. Lemos	282
<b>concepts</b>	Christopher Peacocke	285
<b>Continental epistemology</b>	Linda Alcoff	287
<b>convention</b>	Robert S. Tragesser	292
<b>conversational implicature</b>	Duncan Pritchard	293
<b>criteria and knowledge</b>	Bruce Hunter	294
<b>criterion, canon</b>	Bruce Hunter	298
<b>critical cognitivism</b>		299
<b>D</b>		300
<b>Davidson, Donald (1930–2003)</b>	Ernest Lepore	300
<b>death of epistemology</b>	Michael Williams	301
<b>defeasibility</b>	Bruce Hunter	304
<b>definition</b>	Robert S. Tragesser	305
<b>Derrida, Jacques (1930–2004)</b>	Samuel C. Wheeler III	306
<b>Descartes, René (1596–1650)</b>	John Cottingham	306
<b>Dewey, John (1859–1952)</b>	Peter H. Hare	310
<b>dialectic (Hegel)</b>	Kenneth R. Westphal	312
<b>dialectic (Plato)</b>	Nicholas P. White	313
<b>different constructions in terms of “knows”</b>	Jaakko Hintikka	313
<b>direct realism</b>	L. S. Carrier	318
<b>disposition</b>	John Heil	322
<b>dogmatism</b>	Alvin Plantinga	322
<b>doxastic voluntarism</b>	Sharon Ryan	322
<b>Dutch book argument</b>		325
<b>E</b>		326
<b>empiricism</b>	Bruce Hunter	326
<b>Epicurus (c.341–271 bc)</b>	Phillip Mitsis	331
<b>epistemic deontology</b>	Sharon Ryan	332
<b>epistemic luck</b>	Mylan Engel	336
<b>epistemic supervenience</b>	John Turri	340
<b>epistemic virtue</b>	Guy Axtell	343
<b>epoche</b>	Dagfinn Follesdal	347
<b>essence (Husserl)</b>	Dagfinn Follesdal	347
<b>essence (Plato)</b>	Nicholas P. White	348
<b>essentialism</b>		348
<b>ethics and epistemology</b>	Jonathan Dancy	348
<b>evidence</b>	Richard Feldman	349
<b>evolutionary argument against naturalism</b>	Omar Mirza	351
<b>evolutionary epistemology</b>	Edward Stein	354
<b>existence</b>	Christopher Hookway	356
<b>experience, theories of</b>	Michael Pendlebury	357
<b>explanation</b>	Wesley C. Salmon	361
<b>external world</b>		364
<b>externalism/internalism</b>	Laurence Bonjour	364

<b>F</b>	369
<b>fact/value</b> Robert Audi	369
<b>fallibilism</b> Adam Leite	370
<b>feminist epistemology</b> Lorraine Code	375
<b>first philosophy</b> Christopher Hookway	380
<b>Firth, Roderick (1917–87)</b> John Troyer	380
<b>Foucault, Michel (1926–84)</b> Linda Alcoff	381
<b>foundationalism</b> William P. Alston	382
<b>Frege, Gottlob (1848–1925)</b> David Bell	385
<b>Freud, Sigmund (1856–1939)</b> Jim Hopkins	387
<b>G</b>	389
<b>Gadamer, Hans-Georg (1900–2002)</b> Linda Alcoff	389
<b>genetic epistemology</b> Richard F. Kitchener	389
<b>genetic fallacy</b> Douglas N. Walton	393
<b>geometry</b> Wesley C. Salmon	393
<b>Gettier problem</b> Paul K. Moser	395
<b>given, the</b> Alan H. Goldman	397
<b>Goodman, Nelson (1906–98)</b> Catherine Z. Elgin	400
<b>Grice, H. Paul (1913–88)</b> Richard E. Grandy	402
<b>H</b>	404
<b>Habermas, Jürgen (1929– )</b> Hans-Johann Glock	404
<b>Hegel, Georg Wilhelm Friedrich (1770–1831)</b> Kenneth R. Westphal	405
<b>Heidegger, Martin (1889–1976)</b> Charles Guignon	408
<b>Hempel, Carl Gustav (1905–97)</b> Jaegwon Kim	409
<b>Hempel's paradox of the ravens</b> Peter Achinstein	410
<b>hermeneutics</b> Charles Guignon	412
<b>Hintikka, Jaakko (1929– )</b> Risto Hilpinen	413
<b>historical knowledge</b> Jack W. Meiland	414
<b>historicism</b> David Bakhurst	416
<b>Hobbes, Thomas (1588–1679)</b> Tom Sorell	417
<b>holism</b> Christopher Hookway	420
<b>Hume, David (1711–76)</b> Barry Stroud	420
<b>Husserl, Edmund (1859–1938)</b> Dagfinn Follesdal	423
<b>I</b>	425
<b>idea</b> R. S. Woolhouse	425
<b>idealism</b> Nicholas Rescher	425
<b>ideology</b> David Bakhurst	429
<b>illusion</b>	431
<b>immediacy, presence</b> Alan H. Goldman	431
<b>in itself/for itself</b> M. Okrent	432
<b>incorrigibility</b> William P. Alston	433
<b>indeterminacy of reference</b> Christopher Hookway	433
<b>indeterminacy of translation</b> Christopher Hookway	434
<b>Indian epistemology</b> J. N. Mohanty	434
<b>indubitability</b> William P. Alston	438
<b>induction, problem of</b>	438
<b>induction: enumerative and hypothetical</b> Gilbert Harman	438
<b>infallibility</b> William P. Alston	444

<b>inference</b> Robert S. Tragesser	444
<b>inference to the best explanation</b> Richard Fumerton	445
<b>infinite regress argument</b> John F. Post	447
<b>informal fallacies</b> Douglas N. Walton	450
<b>innate ideas</b> G. A. J. Rogers	454
<b>introspection</b> Vrinda Dalmiya	456
<b>intuition and deduction</b> Robert S. Tragesser	460
<b>intuition in epistemology</b> Bruce Russell	464
<b>isostheneia</b> Charlotte Stough	468
<b>J</b>	469
<b>James, William (1842–1910)</b> Peter H. Hare	469
<b>judgement</b> John Heil	470
<b>justification</b> Matthias Steup	471
<b>K</b>	472
<b>Kant, Immanuel (1724–1804)</b> James van Cleeve	472
<b>KK-thesis</b> Risto Hilpinen	476
<b>knower paradox</b>	476
<b>knowledge</b>	476
<b>knowledge and belief</b> Steven Luper	476
<b>knowledge by acquaintance/by description</b> David B. Martens	479
<b>knowledge how, who, why etc.</b>	482
<b>knowledge-seeking by questioning</b> Jaakko Hintikka	483
<b>L</b>	487
<b>Leibniz, Gottfried Wilhelm (1646–1716)</b> R. C. Sleigh, JR	487
<b>Lewis, Clarence Irving (1883–1964)</b> John Troyer	489
<b>limits of human knowledge</b> Nicholas Rescher	490
<b>linguistic understanding</b> Christopher Peacocke	492
<b>literature and knowledge</b> Paisley Livingston	497
<b>Locke, John (1632–1704)</b> R. S. Woolhouse	500
<b>logical construction</b> R. M. Sainsbury	503
<b>logical empiricism</b>	504
<b>logical positivism</b> Barry Stroud	504
<b>logicism</b> David Bell	506
<b>lottery paradox</b> Jonathan Vogel	506
<b>luck, epistemic</b>	511
<b>M</b>	512
<b>Marxism</b> David Bakhurst	512
<b>mathematical knowledge</b> Mark Steiner	514
<b>memory</b> Tom Senor	520
<b>Merleau-Ponty, Maurice (1908–61)</b> John J. Compton	524
<b>methodology</b> Wesley C. Salmon	525
<b>Mill, John Stuart (1806–73)</b> John Skorupski	525
<b>Molyneux's problem</b> Steve Smith	527
<b>Montaigne, Michel de (1533–92)</b> Richard H. Popkin	528
<b>Moore, G(eorge) E(dward) (1873–1958)</b> Thomas Baldwin	529
<b>Moore's paradox</b> Roy A. Sorensen	531
<b>moral epistemology</b> Jonathan Dancy	532
<b>myth of the given</b>	537

N	538
<b>naïve realism</b>	538
<b>natural science, epistemology of</b> Wesley C. Salmon	538
<b>naturalism</b> Philip Pettit	542
<b>naturalized epistemology</b> Hilary Kornblith	543
<b>necessary/contingent</b> David Blumenfeld	546
<b>necessity, modal knowledge</b> Albert Casullo	547
<b>Neoplatonism</b> S. Strange	548
<b>Neurath, Otto (1882–1945)</b> Andy Hamilton	549
<b>Nietzsche, Friedrich (1844–1900)</b> Alexander Nehamas	551
<b>nihilism</b>	552
<b>noumenal/phenomenal</b> James van Cleve	552
<b>Nozick, Robert (1940–2002)</b> Sherrilyn Roush	554
<b>Nyāya</b> J. N. Mohanty	556
O	558
<b>objective/subjective</b> Robert Audi	558
<b>objectivity</b> David Bell	559
<b>Ockham, William of (c.1285–1347)</b> Marilyn McCord Adams	562
<b>ontological commitment</b> Christopher Hookway	563
<b>ontological relativity</b> Christopher Hookway	564
<b>ostensive definition</b> P. M. S. Hacker	564
<b>other minds</b> Akeel Bilgrami	566
P	572
<b>paradox</b> Jonathan Vogel	572
<b>paradox of the knower</b> C. Anthony Anderson	572
<b>paradoxes of analysis</b> Felicia Ackerman	574
<b>paranormal knowledge</b> Patrick Grim	577
<b>Peirce, Charles S. (1839–1914)</b> Christopher Hookway	580
<b>perceptual knowledge</b> Fred Dretske	581
<b>perspectivism</b>	586
<b>phenomenalism</b> Richard Fumerton	586
<b>phenomenology</b> Charles Guignon	590
<b>philosophical knowledge</b> Felicia Ackerman	590
<b>Plato (c.429–347 bc)</b> Nicholas White	593
<b>Popper, Karl (1902–94)</b> Anthony O'Hear	597
<b>positivism</b>	599
<b>pragmatism</b> Susan Hack	599
<b>preface paradox</b> Earl Conee	604
<b>presence</b>	606
<b>presocratic epistemology</b> J. H. Lesher	606
<b>prima facie reasons</b> Jonathan Dancy	609
<b>primary and secondary qualities</b> Barry Stroud	609
<b>principle of charity</b> Ernest Lepore	613
<b>principle of contradiction</b> Robert S. Tragesser	613
<b>principle of credulity</b> William P. Alston	614
<b>principle of identity</b> Robert A. Tragesser	615
<b>private language argument</b> P. M. S. Hacker	616
<b>probability, theories of</b> Brian Skyrms	622
<b>problem of the criterion</b> Matthias Steup	626

<b>problem of the external world</b>	George Pappas	628
<b>problem of induction</b>		634
<b>problem of other minds</b>		634
<b>problem of rule-following</b>	Philip Pettit	634
<b>problems of induction</b>	Laurence Bonjour	638
<b>projection, projectibility</b>	Catherine Z. Elgin	643
<b>proof</b>	Robert S. Tragesser	643
<b>propositional knowledge</b>	Robert K. Shope	644
<b>protocol sentences</b>	Andy Hamilton	649
<b>psychologism</b>	David Bell	649
<b>psychology and epistemology</b>	Edward Erwin	649
<b>Putnam, Hilary (1926– )</b>	Thomas Tymoczko	654
<b>Pyrrhonism</b>	Charlotte Stough	655
<b>Q</b>		657
<b>Quine, Willard Van Orman (1908–2000)</b>	Christopher Hookway	657
<b>R</b>		659
<b>rationalism</b>	Edwin Curley	659
<b>rationality</b>	L. Jonathan Cohen	663
<b>realism</b>	Philip Pettit	668
<b>reasons/causes</b>	Robert Audi	672
<b>Reichenbach, Hans (1891–1953)</b>	Wesley C. Salmon	673
<b>Reid, Thomas (1710–96)</b>	Keith Lehrer	674
<b>reification, hypostatization</b>	Christopher Hookway	676
<b>relativism</b>	Harvey Siegel	676
<b>relevant alternatives</b>	Stewart Cohen	678
<b>reliabilism</b>	Alvin Goldman	681
<b>religious belief, epistemology of</b>	Alvin Plantinga	692
<b>religious belief, epistemology of – recent developments</b>	Michael Bergmann	697
<b>representation</b>	Robert Cummins	699
<b>representative realism</b>	Frank Jackson	702
<b>Rescher, Nicholas (1928– )</b>	Robert Almeder	705
<b>Rorty, Richard (1931–2007)</b>	Michael Williams	706
<b>rule-following</b>		707
<b>Russell, Bertrand Arthur William (1872–1970)</b>	R. M. Sainsbury	707
<b>Ryle, Gilbert (1900–76)</b>	Graham Bird	709
<b>S</b>		
<b>safety</b>		712
<b>Santayana, George (1863–1952)</b>	John Lachs	712
<b>Sartre, Jean-Paul (1905–80)</b>	Gregory McCulloch	713
<b>scepticism</b>	Peter D. Klein	714
<b>scepticism, contemporary</b>	Peter D. Klein	715
<b>scepticism, modern</b>	Richard Popkin	719
<b>Schlick, Moritz (1882–1936)</b>	Andy Hamilton	722
<b>scientia media</b>		723
<b>self-consciousness</b>	Vrinda Dalmiya	723
<b>self-evidence</b>	Alan H. Goldman	723
<b>self-knowledge and self-identity</b>	Sydney Shoemaker	724
<b>self-presenting</b>	Alan H. Goldman	726

# CONTENTS

<b>Sellars, Wilfrid (1912–89)</b>	Jay Rosenberg	727
<b>sensation/cognition</b>	Fred Dretske	729
<b>sense-data</b>	Alan H. Goldman	732
<b>sensitivity and safety</b>	Duncan Pritchard	732
<b>Sextus Empiricus</b>	Charlotte Stough	736
<b>simplicity</b>	Elliott Sober	738
<b>social sciences, epistemology of</b>	Fred D’Agostino	739
<b>sociology of knowledge</b>	David Bloor	744
<b>Socrates</b>		747
<b>solipsism</b>	Clive Borst	747
<b>Spinoza, Benedict (<i>also</i> Baruch) (1632–77)</b>	Don Garrett	749
<b>Stoic epistemology</b>	J. V. Allen	750
<b>Strawson, Peter Frederick (1919–2006)</b>	P. F. Snowdon	753
<b>strong programme</b>	David Bloor	754
<b>Suárez, Francisco (1548–1617)</b>	Jorge J. E. Gracia	755
<b>subjectivism</b>	Richard Foley	756
<b>subjectivity</b>	Christopher Peacocke	758
<b>supervenience</b>		760
<b>surprise examination paradox</b>	Roy A. Sorensen	760
<b>T</b>		763
<b>tabula rasa</b>	R. S. Woolhouse	763
<b>testimony</b>	Jennifer Lackey	763
<b>theory</b>	Wesley C. Salmon	768
<b>transcendental arguments</b>	A. C. Grayling	768
<b>tripartite definition of knowledge</b>	Paul K. Moser	771
<b>truth, theories of</b>	Paul Horwich	772
<b>truths of reason/truths of fact</b>	David Blumenfeld	777
<b>U</b>		779
<b>underdetermination of theory</b>	Christopher Hookway	779
<b>use/mention</b>	Matthias Steup	779
<b>V</b>		780
<b>verificationism</b>	Barry Stroud	780
<b>Vico, Giambattista (1668–1744)</b>	Leon Pompa	780
<b>Vienna Circle</b>	Andy Hamilton	782
<b>W</b>		783
<b>Whitehead, Alfred North (1861–1947)</b>	Donald W. Sherburne	783
<b>Wittgenstein, Ludwig (1889–1951)</b>	David Pears	783
<b>Z</b>		788
<b>zebras and cleverly disguised mules</b>	Mylan Engel	788
<b>Index</b>		793

# Contributors

---

Peter Achinstein  
*Johns Hopkins University*

Felicia Ackerman  
*Brown University*

Laird Addis  
*University of Iowa*

Linda Martín-Alcoff  
*Hunter College*

J. V. Allen  
*University of Pittsburgh*

Robert F. Almeder  
*Georgia State University*

William P. Alston<sup>†</sup>

C. Anthony Anderson  
*University of California, Santa Barbara*

Robert Audi  
*University of Notre Dame*

Guy Axtell  
*University of Nevada Reno*

Michael Ayers  
*University of Oxford*

David Bakhurst  
*Queen's University at Kingston, Canada*

Thomas Baldwin  
*University of York*

David Bell  
*University of Sheffield*

Michael Bergmann  
*Purdue University*

Akeel Bilgrami  
*Columbia University*

Graham Bird  
*University of Manchester (Emeritus)*

David Bloor  
*University of Edinburgh*

David Blumenfeld

Laurence Bonjour  
*University of Washington*

Clive Borst

Anthony Brueckner  
*University of California, Santa Barbara*

L. S. Carrier  
*University of Miami*

Albert Casullo  
*University of Nebraska-Lincoln*

R. M. Chisholm<sup>†</sup>

Lorraine Code  
*York University, Ontario*

L. Jonathan Cohen<sup>†</sup>

Stewart Cohen  
*University of Arizona, Tucson*

John J. Compton  
*Vanderbilt University*

Earl Conee  
*University of Rochester*

## CONTRIBUTORS

John Cottingham  
*University of Reading*

Robert Cummins  
*University of Illinois at Urbana-Champaign*

Edwin Curley  
*University of Michigan*

Fred D'Agostino  
*University of Queensland*

Vrinda Dalmiya  
*University of Hawai'i at Manoa*

Jonathan Dancy  
*University of Reading and  
University of Texas, Austin*

Fred Drestke  
*Duke University*

Catherin Z. Elgin  
*Harvard University*

Mylan Engel  
*Northern Illinois University*

Edward Erwin  
*University of Miami*

Richard Feldman  
*University of Rochester*

Richard Foley  
*New York University*

Dagfinn Føllesdal  
*Stanford University*

Graeme Forbes  
*University of Colorado at Boulder*

Richard Fumerton  
*University of Iowa*

Don Garrett  
*New York University*

Margaret Gilbert  
*University of Connecticut*

Carl Ginet  
*Cornell University*

Hans-Johann Glock  
*Universität Zürich*

Alan H. Goldman  
*College of William and Mary*

Alvin I. Goldman  
*Rutgers University*

Jorge J. E. Gracia  
*University at Buffalo*

Richard E. Grandy  
*Rice University*

A. C. Grayling  
*Birkbeck, University of London*

John Greco  
*St. Louis University*

Patrick Grim  
*State University of New York at Stony Brook*

Charles Guignon  
*University of South Florida*

Susan Haack  
*University of Miami*

P. M. S. Hacker  
*St John's College, Oxford*

Alan Hájek  
*Australian National University*

Andy Hamilton  
*Durham University*

Peter H. Hare<sup>†</sup>

Gilbert Harman  
*Princeton University*

Stephan Hartmann  
*Tilburg University*

Adrian Heathcote  
*University of Sydney*

John Heil  
*Washington University*



Risto Hilpinen  
*University of Miami*

Jaakko Hintikka  
*Boston University*

Christopher Hookway  
*University of Sheffield*

Jim Hopkins  
*King's College London*

Paul Horwich  
*New York University*

Michael Huemer  
*University of Colorado at Boulder*

Bruce Hunter  
*University of Alberta*

Terence Irwin  
*Keble College, University of Oxford*

Frank Jackson  
*Australian National University*

Jerrold J. Katz<sup>†</sup>

Jaegwon Kim  
*Brown University*

Richard F. Kitchener  
*Colorado State University*

Peter D. Klein  
*Rutgers University*

Hilary Kornblith  
*University of Massachusetts*

Jennifer Lackey  
*Northwestern University*

John Lachs  
*Vanderbilt University*

Keith Lehrer  
*University of Arizona (Emeritus)*

Adam Leite  
*Indiana University*

Noah M. Lemos  
*College of William & Mary*

Ernest LePore  
*Rutgers University*

J. H. Lesher  
*University of North Carolina at Chapel Hill*

Paisley Livingston  
*Lingnan University*

Steven Luper  
*Trinity University*

Scott MacDonald  
*Cornell University*

Peter Markie  
*University of Missouri*

David B. Martens  
*University of the Witwatersrand, Johannesburg*

Marilyn McCord Adams  
*University of North Carolina at Chapel Hill*

Gregory McCulloch<sup>†</sup>

David McNaughton  
*Florida State University*

Jack W. Meiland<sup>†</sup>

Omar Mirza  
*St. Cloud State University*

Phillip Mitsis  
*New York University*

J. N. Mohanty  
*Temple University*

Paul K. Moser  
*Loyola University Chicago*

Alexander Nehamas  
*Princeton University*

Anthony O'Hear  
*University of Buckingham*

## CONTRIBUTORS

M. Okrent  
*Bates College*

George Pappas  
*The Ohio State University (Emeritus)*

Christopher Peacocke  
*Columbia University*

David Pears<sup>†</sup>

Michael Pendlebury  
*North Carolina State University*

Philip Pettit  
*Princeton University*

Alvin Plantinga  
*University of Notre Dame*

John Pollock<sup>†</sup>

Leon Pompa

Richard H. Popkin<sup>†</sup>

John F. Post  
*Vanderbilt University*

Duncan Pritchard  
*University of Edinburgh*

Nicholas Rescher  
*University of Pittsburgh*

G. A. J. Rogers  
*Keele University*

Jay Rosenberg<sup>†</sup>

Sherrilyn Roush  
*University of California, Berkeley*

Bruce Russell  
*Wayne State University*

Sharon Ryan  
*West Virginia University*

R. M. Sainsbury  
*University of Texas at Austin*

Wesley C. Salmon<sup>†</sup>

Tom Senor  
*University of Arkansas*

Donald W. Sherburne  
*Vanderbilt University (Emeritus)*

Sydney Shoemaker  
*Cornell University*

Robert K. Shope  
*University of Massachusetts Boston*

Harvey Siegel  
*University of Miami*

John Skorupski  
*University of St. Andrews*

Brian Skyrms  
*University of California*

R. C. Sleigh, Jr  
*University of Massachusetts (Emeritus)*

Steve Smith  
*Yale University*

P. F. Snowdon  
*University College London*

Elliott Sober  
*University of Wisconsin – Madison*

Tom Sorell  
*University of Birmingham*

Roy A. Sorensen  
*Washington University in St. Louis*

Ernest Sosa  
*Rutgers University*

Edward Stein  
*Benjamin N. Cardozo School of Law*

Mark Steiner  
*The Hebrew University of Jerusalem*

Matthias Steup  
*Purdue University*

Charlotte Stough<sup>†</sup>

S. Strange

Barry Stroud  
*University of California*

Robert S. Tragesser

John Troyer  
*University of Connecticut*

John Turri  
*Huron University College*

Thomas Tymoczko<sup>†</sup>

James Van Cleve  
*University of Southern California*

Jonathan Vogel  
*Amherst College*

Douglas N. Walton  
*University of Windsor*

Kenneth R. Westphal  
*University of Kent*

Samuel C. Wheeler III  
*University of Connecticut*

Nicholas P. White  
*University of California (Emeritus)*

Michael Williams  
*Johns Hopkins University*

Timothy Williamson  
*University of Oxford*

R. S. Woolhouse  
*University of York (Emeritus)*

Crispin Wright  
*New York University*

Linda Zagzebski  
*University of Oklahoma*



# Preface to the First Edition

---

JONATHAN DANCY

The present Companion, like the majority of the other volumes in the Blackwell Companions to Philosophy series, is organized as a standard reference book, with alphabetically arranged articles of varying length (anything from 250 to 3500 words) on leading theories, thinkers, ideas, distinctions and concepts in epistemology. It aims for a broad readership, while recognizing that the nature of contemporary epistemology inevitably imposes restrictions on this. In some other areas of philosophy it remains feasible to design a book which is largely accessible to the general reader; in epistemology, however, the main readership is likely to be students from undergraduate level upwards, as well as professional philosophers, and it is to them that the Companion is primarily addressed. A minority of topics resist treatment other than at an advanced level: they have not for that reason been excluded, lest coverage of the area become incomplete. But the vast majority are accessible to all levels of the intended readership.

Not all entries will be comprehensible on their own: at least, not to the inexperienced reader. This is where the cross-referencing system comes in. I have used two interrelated methods of guiding readers from one entry to another. Within the text itself terms or names occur in small capitals; this will often occur where reference is made to DESCARTES, or to REALISM, for example. This means that there is an entry on this person or topic, and that it would be worthwhile having a look at it for present purposes. The mere fact that there is an entry on this person or topic, however, is not sufficient for me to flag it. Not all references to Descartes or to realism are significant. What is more, a person or topic may not be flagged in this way on its first

occurrence in an entry; I may wait for the best moment, as it were. And sometimes one and the same person or topic is flagged more than once in the same entry, where there has been a long gap or I think it particularly appropriate for some other reason.

Most of the flagging that is done within the body of the text is of this form; a word or phrase is highlighted in the sentence, as I highlighted DESCARTES above. In doing this, I have not insisted slavishly that the word highlighted be *exactly* the same one as the headword that the reader is effectively being referred to. For instance, I may direct the reader to an entry on realism by flagging the remark that Santayana was a REALIST. Sometimes, however, I was unable to work the cross-reference into the text in this way. On these occasions it is inserted at the end of the relevant sentence or paragraph.

There are also cross-references to be found at the end of most entries. These fulfil two functions. First, they enable me to point out areas to which the present entry is related, but which have not occurred significantly in the text. Second, they enable me to insist a bit that you should consider again looking at an entry that has already been flagged in the text. So if you see a person or topic flagged both within and at the end of an entry, you can take it that I think you really should have a look at it.

The Blackwell series of Companions is conceived as related primarily to Anglo-American philosophy. The topics the editors chose to cover were selected with this in mind. But this does not mean that other traditions are completely ignored. There is an entry on Indian epistemology and, as well as a general entry on Continental epistemology,

there are many entries on individual thinkers in that tradition. We do not pretend, however, to give that tradition as detailed coverage as we give to the one which is our main focus.

It might be thought that the jacket illustration is symptomatic of our general approach.<sup>1</sup> Here we have the solitary thinker working in private. Isn't he a wonderful example of the *CARTESIAN* approach to epistemology which is so characteristic of the Anglo-American analytic tradition, and which is so vehemently rejected on the Continent? There is some truth in this, which we will come to in a moment. There are two points to be made against it. First, the attempt to escape from the clutches of the Cartesian paradigm is as common within the analytic tradition as it is outside. Second, our solitary thinker is not as solitary as all that. He is reading a book, which could be taken to show that he is not relying entirely on his own resources, as the Cartesian mind is supposed to do (*see* REID; TESTIMONY). Against this, one could point out that the picture exemplifies a conception of knowledge as something to be gained by rational enquiry and perception rather than in practical life and action. This "logocentrism" may be a more insidious feature of the Cartesian approach, and certainly the emphasis on practice and action is distinctive of Continental epistemology (*see* for example HEIDEGGER), as is an emphasis on social considerations.

One difficulty the editors faced in deciding which topics to cover derived from the interconnectedness of philosophical areas. Epistemology can be to some extent separated from adjacent areas, but only with a justified sense of artificiality. The nearest areas are metaphysics, philosophy of mind and philosophy of science. These gave us two problems, one theoretical and one practical. The practical one was that in considering whether to include an entry on a topic, we had to ask ourselves whether there would be an entry on it in one of the other *Companions*, and if so how our entry should be related to that one. At the limit, we have an entry on natural science, an area which will on its own occupy a large part of one *Companion*. But there are many other occasions where the shortness of our cov-

erage here is caused by our sense that the major entry on this topic should not appear in a *Companion to Epistemology*. The theoretical one was that there are many occasions where views in epistemology are dependent on views in metaphysics or in the philosophy of mind, and we could not hope to cover everything equally well. Contributors were asked to concentrate on epistemology, and the entries have been written accordingly. When reading entries on individual thinkers, therefore, you should bear in mind that these entries do not pretend to be complete accounts of their subject's work in philosophy; they are concentrating on the epistemology as far as that is possible. The same applies to topics. The entry on natural science is concerned only with the epistemology of science, the entry on religious belief limits itself to epistemological considerations, and so on. The limitation to epistemology is normally implicit rather than explicit; otherwise every entry would have to be headed "X's epistemology" or "the epistemology of Y".

This *Companion* has two editors, divided by the Atlantic (and rejoined by electronic mail). Its general shape was conceived during a very pleasant weekend which I spent in Providence, RI, in Spring 1989. Thereafter, I relied on Ernest Sosa for a constant stream of suggestions about who in the US we might approach as potential contributors – a stream that was evidence of his enviable knowledge of the profession. UK contributors were my responsibility. Beyond that, the detailed editing of contributions has been my province, though I am very grateful to my co-editor for help and advice on the occasional knotty points that arose. I am, of course, equally grateful to our contributors for being willing to undertake what in many cases was a fairly thankless and far from easy task – and for the openness with which so many of them received my suggestions for changes to suit my own idea of how things should be. I have had many occasions to express my appreciation of the professionalism of the profession.

Finally, I want to thank my wife Sarah, who helped me with various aspects of the editing process, and my son Hugh, who spent

two weeks last autumn turning entries into computer-readable form. For a while this Companion was a family affair.

*Keele, February 1992*

NOTE

- 1 This paragraph refers to the cover illustration from the first edition, Georg Friedrich Kersting's *Lesender bei Lampenlicht*.





# Preface to the Second Edition

---

The second edition of the *Companion to Epistemology* differs from the first edition in being divided into three parts. Part I consists of ten new review essays that analyze and discuss recent developments in the main areas of epistemology. Part II presents the reader with a novel approach to portraying a philosophical discipline: twenty epistemological self-profiles. The purpose of these self-profiles is to let their authors give succinct, authoritative accounts of their views and chief arguments in support of them. Reading through all twenty of them will give students and practitioners of epistemology an excellent view of the range of theories and the high level of sophistication characteristic of the current state of the art in the field. Deciding whom to invite to contribute a self-profile obviously raised some difficult issues. The editors' choices were guided by the thought that authors of self-profiles should have produced a stream of significant publications spanning the past two decades, publications addressing the main issues of epistemology and advancing the state of play in the discipline. This approach slanted our choices in favor of seniority, excluding several more junior epistemologists who, using different criteria, would have deserved inclusion.

The first edition of the *Companion* makes up most of Part III, Epistemology from A–Z. The following new or significantly revised entries have been added to this part: a priori (Albert Casullo), Alston, William (Robert Audi), belief (John Heil), defeasibility (Bruce Hunter), empiricism (Bruce Hunter), con-

versational implicature (Duncan Pritchard), doxastic voluntarism (Sharon Ryan), epistemic deontology (Sharon Ryan), epistemic luck (Mylan Engel), epistemic supervenience (John Turri), epistemic virtue (Guy Axtell), ethics and epistemology (Jonathan Dancy), evolutionary argument (Omar Mirza), fallibilism (Adam Leite), feminist epistemology (Lorraine Code), the given (Alan Goldman), intuition in epistemology (Bruce Russell), lottery paradox (Jonathan Vogel), memory (Tom Senor), Robert Nozick (Sherilyn Roush), moral epistemology (Jonathan Dancy), naturalized epistemology (Hilary Kornblith), relativism (Harvey Siegel), reliabilism (Alvin Goldman), religious belief – recent developments (Michael Bergmann), sensitivity (Duncan Pritchard), testimony (Jennifer Lackey), zebras and cleverly disguised mules (Mylan Engel).

I wish to thank David Coss and Tanya Hall (at the time, philosophy majors at St. Cloud State University) for proofreading the scanned first edition files, Erin Kealey at Purdue for correcting the entire set of page proofs for the second edition and preparing the index, and Nick Bellorini and Liz Cremona at Blackwell for their editorial help and guiding the second edition through the production process. I also wish to thank my co-editors, Jonathan Dancy and Ernest Sosa, for their expert advice on planning and putting together this volume.

Matthias Steup  
Purdue University



Part I

Ten Review Essays



# Skepticism and Closure

## 1. PRELIMINARY FORMULATIONS OF CLOSURE

I know the conjunction of some axioms (say, those of Peano Arithmetic), and I prove a theorem T on their basis. My belief of T presumably amounts to knowledge. That is how knowledge is generated in the so-called deductive sciences. As a first shot, we might formulate a generalized *Closure Principle* for knowledge as follows:

CL1: Necessarily, for all S,  $\phi, \psi$ : if S knows that  $\phi$ , and  $\phi$  entails  $\psi$ , then S knows that  $\psi$ .

Call this principle *closure of knowledge under entailment*. The meaning of the word “closure” here follows its mathematical usage, according to which a set  $\alpha$  is closed under a function F defined on its members just in case F when applied to a member (or members) of  $\alpha$  always yields a member of  $\alpha$  as its value. So the set of even natural numbers is closed under addition, since the sum of even addends is always even. Similarly, CL1 says that, for any member  $\phi$  of the set K of propositions known by some S, each entailed consequence  $\psi$  of  $\phi$  will also be a member of K. But CL1 is obviously false. There are infinitely many entailed consequences of the conjunction of the Peano axioms of which I am unaware. Since I do not believe these propositions, I lack knowledge of them.

Suppose that we try to solve the problem by restricting CL1 to those entailed consequences of a known proposition that are *known to be such*:

CL2: Necessarily, for all S,  $\phi, \psi$ : if S knows that  $\phi$ , and S knows that  $\phi$  entails  $\psi$ , then S knows that  $\psi$ .

Call this *closure of knowledge under known entailment*. There are two problems with CL2. First, suppose that I know the conjunction of the Peano axioms and also know that this conjunction entails T\*. However, owing to a sort of cognitive disconnect, I fail to put these

two pieces of knowledge together and thus fail to deduce T\*. Suppose, then, that I fail to believe T\*. Then I shall fail to know T\*. Second, suppose, again, that I fail to put my two pieces of knowledge together but nevertheless believe T\* because I think that the ghost of Elvis has testified to the truth of T\*. Then I again fail to know T\*.

In order to deal with these problems, we shall restrict CL2 to those entailed consequences of a known proposition that are (1) known to be such, and (2) believed on the basis of an appropriate inference:

CL3: Necessarily, for all S,  $\phi, \psi$ : if S knows that  $\phi$ , and S knows that  $\phi$  entails  $\alpha$ , and S believes that  $\psi$  on the basis of an inference from  $\phi$  and ( $\phi$  entails  $\psi$ ), then S knows that  $\psi$ .<sup>1,2</sup>

## 2. THE CLOSURE-BASED CARTESIAN SKEPTICAL ARGUMENT

CL3 can be used in the construction of a Cartesian skeptical argument concerning knowledge of propositions about the external world. Let us choose a *target proposition* from the set of those propositions about the external world which I take myself to know. Let us choose  $H = I \text{ have hands}$ . Let SK be the proposition that *I am a brain in a vat with experiences just like those I actually have, produced by a supercomputer's stimulation of my brain*. Let “K(S,  $\phi$ )” stand for *S knows that  $\phi$* . Let “B(S,  $\psi, \alpha, \beta$ )” stand for *S believes that  $\psi$  on the basis of an inference from  $\alpha$  and  $\beta$* . Let “ab” stand for *Anthony Brueckner*. Here is the skeptical argument:

- (1) If K(ab, H), and K(ab, [H entails  $\sim$ SK]), and B(ab,  $\sim$ SK, H, [H entails  $\sim$ SK]), then K(ab,  $\sim$ SK).
  - (2) {K(ab, [H entails  $\sim$ SK]) and B(ab,  $\sim$ SK, H, [H entails  $\sim$ SK])}.
  - (3)  $\sim$ K(ab,  $\sim$ SK).
- So:
- (4)  $\sim$ {K(ab, H), and K(ab, [H entails  $\sim$ SK]), and B(ab,  $\sim$ SK, H, [H entails  $\sim$ SK])}.
- So:
- (5)  $\sim$ K(ab, H).

Premise 1 is just an instance of CL3. Premise 2 is in fact true of me: I know that the pertinent entailment holds, and I believe that  $\sim$ SK on the basis of an inference from my beliefs that H and that the entailment holds. What of premise 3? My sensory evidence for believing  $\sim$ SK does not favor  $\sim$ SK over SK, since I would have that same evidence regardless of whether I am in a world in which  $\sim$ SK is true (a normal world) or, instead, in a world in which SK is true (a vat world). In other words, my sensory evidence fails to discriminate between  $\sim$ SK and SK. Thus, it seems to follow that my evidence fails to *justify* me in believing that  $\sim$ SK. Hence I do not know that  $\sim$ SK, supposing that justification is a necessary condition for knowledge. This reasoning in support of premise 3 can be seen as resting upon the following *Underdetermination Principle*:

- (UP) If S's evidence for  $\phi$  does not favor  $\phi$  over an incompatible hypothesis  $\psi$ , then S is not justified in believing  $\phi$  and rejecting  $\psi$ .<sup>3</sup>

The antecedent of UP says, in effect, that S's evidence *underdetermines* a choice between  $\phi$  and  $\psi$ . As for the rest of the argument, 4 follows from 1 and 3 by Modus Tollens, and the conclusion 5 follows from 2 and 4 by propositional logic.

The foregoing argument would seem to generalize to all putative knowers and to just about all putatively known propositions about the external world.<sup>4</sup>

### 3. DENYING CLOSURE: NOZICK

One way to block the skeptical argument is to deny CL3. If CL3 is not an exceptionless general principle about knowledge, then the skeptic cannot appeal to this principle as the rationale underlying premise 1. In a famous discussion, Robert Nozick argues that, on his *Tracking Analysis of Knowledge*, CL3 fails.<sup>5</sup> Let " $\phi \Box \rightarrow \psi$ " stand for the subjunctive conditional *If  $\phi$  were the case, then  $\psi$  would be the case*. According to Nozick's Tracking Analysis, S knows that P if and only if:

- (I) P is true.
- (II) S believes that P.
- (III)  $\sim P \Box \rightarrow \sim (S \text{ believes that } P)$ .
- (IV)  $P \Box \rightarrow S \text{ believes that } P$ .

Let us assume that the truth-conditions for  $\phi \Box \rightarrow \psi$  are these: the  $\phi$ -worlds (possible worlds in which  $\phi$  is true) *closest* to (most similar to) the actual world are  $\psi$ -worlds. According to Nozick's analysis, S knows that P just in case S *tracks P's truth-value* through the following range of possible worlds: the closest  $\sim P$ -worlds, the actual world, and the closest (non-actual) P-worlds. Condition III has come to be known as *Sensitivity*. It explains failures to know in Gettier cases, cases of lucky guesses, and cases of wishful thinking.<sup>6</sup> Condition IV captures another dimension of the non-accidentality of knowledge. It explains failures to know in Harman's assassination case<sup>7</sup> and in cases in which a brain in a vat arrives at a rare true belief that P owing to a passing whim of his vat-keeper (who does not happen to give him that true belief in close P-worlds).

Sensitivity reveals that the skeptic is right about one thing: his premise 3. In order to know  $\sim$ SK, S must satisfy this logical equivalent of the pertinent instance of condition

- III:  $SK \Box \rightarrow \sim (S \text{ believes that } \sim SK)$ .

But S does not satisfy this condition, because the closest SK-worlds are worlds where S *does* believe that  $\sim$ SK (just as he actually believes). No one satisfies the condition in question for knowing  $\sim$ SK, because everyone who is in a vat world mistakenly believes that he is *not* in a vat world.

I myself do not at present claim to know whether I am in a normal world or a vat world, given that I am in the middle of an evaluation of the skeptical argument and Closure's role in it.<sup>8</sup> However, I think that there are some subjects in some possible worlds who satisfy Nozick's conditions for knowing H (our "hands" proposition), in virtue of inhabiting normal, non-vat worlds. Call one such subject Ace. H is true in Ace's normal world; Ace correctly believes H; in the  $\sim$ H-worlds close to Ace's world Ace

loses his hands, say, in a terrible accident and accordingly does not mistakenly believe H; and in the H-worlds close to Ace's world, Ace again believes H via perception. So we have  $K(\text{Ace}, H)$ . Assume that Ace knows that H entails  $\sim SK$  and believes  $\sim SK$  on the basis of an appropriate inference. But as we have seen,  $\sim K(\text{Ace}, \sim SK)$ . Given Nozick's Tracking Analysis, then, CL3 fails, because Ace constitutes a counter-example. We have shown this without making any appeal to assumptions about what the actual world is like.<sup>9</sup>

Many philosophers simply find it difficult to accept the claim that one can know that one has hands while *failing* to know that one is *not* a handless brain in a vat (even while believing the latter). Keith DeRose calls this the *abominable conjunction*.<sup>10</sup> Indeed, many criticisms of Nozick's Tracking Analysis have been made. I will mention just two.

First, consider the proposition  $SE = \text{The cause of my series of sense-experiences is physical in nature}$ . According to the Tracking Analysis, I do not know SE. In the closest worlds in which SE is false, either there is no cause of my sense-experiences, or the cause is not physical in nature (for example, a Cartesian Evil Genius). In such worlds, though, I again believe SE. So Sensitivity is violated, and we get a skeptical result from the Tracking Analysis: I do not know SE.<sup>11</sup>

Second, Saul Kripke has presented a case with the following structure.<sup>12</sup> There are 1,001 barn lots in Many Blue Fake Barns Country. For bizarre religious reasons, the county planners want to erect 1,000 blue fake barns and one red real barn on the lots. They have erected 1,000 blue fake barns on Lots 2–1,001. I am looking at the red real barn, located on Lot 1. I know that  $RB = \text{Lot 1 has a red barn}$ . This is because I have a true belief of RB, and because the following subjunctive conditionals are true:

IIIrb  $\sim RB \square \rightarrow \sim(I \text{ believe } RB)$ .

IVrb  $RB \square \rightarrow I \text{ believe } RB$ .

IIIrb is true in virtue of the fact that in the close  $\sim RB$ -worlds a fake blue barn is located on Lot 1 and the red real barn is located on one of the other lots. So I do not believe RB. IVrb is true

in virtue of the fact that in the close RB-worlds I again believe RB on the basis of perception. But I do *not* know that  $B = \text{Lot 1 has a barn}$ , because the following instance of Sensitivity is false:

IIIb.  $\sim B \square \rightarrow \sim(I \text{ believe } B)$ .

IIIb is false because in the closest  $\sim B$  worlds the county planners place a blue fake barn on Lot 1 and the real red barn on some other of the 1,001 lots, and I accordingly believe B. This seems to most people to be a particularly egregious violation of Closure: I know that Lot 1 has a red barn but fail to know that Lot 1 has a barn.<sup>13</sup>

#### 4. DENYING CLOSURE: DRETSKE

Let us now turn to another famous attack on Closure. Nozick's attack centered on Sensitivity, an alleged condition for knowing that is in some sense an *externalist* condition: one can satisfy the condition without being aware that one does. On my understanding of Fred Dretske's first-wave attack on Closure, the focus is upon an *internalist* condition for knowing: having evidential justification.<sup>14</sup> Dretske's story begins with his famous zebra example. In a zoo, I see a zebra in a pen marked with a sign saying "Zebra". My evidence E for believing  $Z = \text{This animal is a zebra}$  consists in my seeing a zoo animal that looks just like a zebra in a pen marked "Zebra". On the basis of this ordinary evidence E, we have  $J(\text{ab}, E, Z)$  (*Anthony Brueckner's belief that Z is justified on the basis of his evidence E*), and we have  $K(\text{ab}, Z)$  as well, since the animal is in fact a zebra. Dretske's intuition is that, after all, people *do* come to know propositions about various animals while strolling through zoos, on the basis of evidence like E. We can also suppose that I know that Z entails  $\sim CD = \text{This animal is not a cleverly disguised mule}$ . Do I know  $\sim CD$ ? According to Dretske, E does not enable me to know  $\sim CD$ . He says that E does not count towards  $\sim CD$ . So Dretske holds that  $\sim J(\text{ab}, E, \sim CD)$  and hence that  $\sim K(\text{ab}, \sim CD)$ .

Dretske has a theory which explains how it is that I can know Z while failing to know

the entailed consequence  $\sim$ CD (even while believing  $\sim$ CD via deduction from Z). The following principle enunciates the *Relevant Alternatives Approach to Knowledge*:

RA: If S knows that  $\phi$ , then S's evidence for  $\phi$  is strong enough to rule out the relevant alternatives to  $\phi$ , i.e., strong enough to justify a belief of the contradictories of these alternatives, enabling S to know these contradictories. S's evidence for  $\phi$ , though, need not be strong enough to rule out the irrelevant alternatives to  $\phi$ . Thus, S will not be able to know the contradictories of those alternatives.

In the zebra case, the following alternatives to Z are relevant: *This animal is a wombat*, *This animal is a horse*, *This animal is a lion*, and so on. But CD is an irrelevant alternative to Z. My evidence E is strong enough to rule out the foregoing relevant alternatives, but E is *not* strong enough to rule out the irrelevant alternative CD. Hence, by RA we can consistently hold that  $J(ab, E, Z)$ ,  $K(ab, Z)$ ,  $\sim J(ab, E, \sim CD)$ , and  $\sim K(ab, \sim CD)$ .

The application of these ideas to our skeptical argument is obvious. If we can hold that SK is an irrelevant alternative to H, then given RA we can deny the Closure-based premise 1 of the argument. I can know H and infer  $\sim$ SK given my appreciation of the pertinent entailment and yet fail to know  $\sim$ SK. So my inability to rule out SK is consistent with  $K(ab, H)$ , just as my inability to rule out CD on the basis of E is consistent with  $K(ab, Z)$ .

It is equally obvious that the success of the RA-based *anti-Closure* strategy depends upon the possibility of providing a principled way of drawing the distinction between relevant and irrelevant alternatives. Further, the success of the RA-based *anti-skeptical* strategy depends upon skeptical alternatives such as SK turning out to be *irrelevant* on whatever is the proposed way of drawing the crucial distinction.

To make it plausible that there is a genuine relevant/irrelevant distinction present in ordinary thought about knowledge and justification, consider a case in which an

LAPD detective is investigating the murders of Nicole Simpson and Ron Goldman. The detective believes OJS = *O. J. Simpson is the murderer*, and his body of evidence rules out to his satisfaction the following alternative to OJS: *O. J. Simpson's blood was planted at the murder scene by overzealous police officers seeking to frame the innocent football star*. The detective's superior is also satisfied that there is a chain-of-evidence case that rules out the foregoing alternative, but the superior says, "Have you ruled out the following alternative to OJS: *Queen Elizabeth II committed the murders because she was having an affair with the famous running back?* Have you checked the Queen's alibi?" The absurdity of the superior's demand highlights the fact that crazy, off-the-wall alternatives that we ordinarily do not consider are irrelevant alternatives which we need not rule out in order to know things such as OJS. Surely skeptical alternatives such as SK fall in the off-the-wall category.

As against this, one can reply that the detective's evidence must in fact be strong enough to rule out the Queen alternative even though the detective never considers that off-the-wall alternative. Otherwise the detective does not know OJS, since his evidence does not justify him in believing his hypothesis about the murders. Similar remarks apply to an ordinary person who claims to know that he has chapped hands but who does not consider off-the-wall handless-brain-in-a-vat alternatives and hence does not try to rule them out.

Here is another way for the RA theorist to go. One summer I drive through Normal Barn Country and get a good look at a barn by the side of the road. My evidence does not enable me to discriminate real barns from fake barns, but we would presumably want to say that  $K(ab, B)$  given the absence of tricky fake barns from the countryside. The following summer finds me driving through Many Fake Barns Country, in which every other barn along the roadside is a fake. Once again, I get a good look at a real barn. My evidence is exactly similar to that of the previous summer, and my powers of real/fake discrimination are unchanged. In this situation, we have  $\sim K(ab, B)$ . This is because the fake barn



alternative to B has become relevant owing to the presence of the many fake barns. My inability to rule out the now-relevant fake barn alternative precludes me from knowing B. That discriminative limitation was compatible with my knowing B the previous summer, in virtue of the fact that the fake barn alternative was then irrelevant. My evidence for B was not then required to have sufficient strength to enable me to rule out the then-irrelevant alternative in order to justify my belief of B and to enable me to know B.<sup>15</sup>

This pair of examples suggests a *probabilistic* way of drawing the relevant/irrelevant distinction. But RA so conceived cannot be brought to bear against the skeptical argument. This is because it is extremely problematic to hold that SK is an irrelevant alternative to H in virtue of its low probability. Are we to suppose, for example, that we have scouted the neighborhood for brains in vats, found none, and thus reasonably concluded that the hypothesis that I am a brain in a vat has a very low probability? Or are we to suppose that we can somehow assign a low *prior probability* to SK that is not conditional upon our body of evidence?

In recent work, Dretske argues against Closure by emphasizing that Closure-denial is really the only game in town.<sup>16</sup> He first maintains that when, for example, little Jimmy peeks into the cookie jar his visual evidence for  $C = \text{There are cookies in the jar}$  is not also evidence for  $P = \text{There is a physical reality independent of little Jimmy's mind}$ , even though C entails P. Dretske calls P a *heavyweight implication* of C, and he holds that, in general, one's justifying evidence for believing an ordinary proposition like C does not *transmit* to the heavyweight implications of the proposition. So, when Ivan tells me over the phone that he is in the middle of mowing the lawn, this testimonial evidence does not constitute a reason for me to believe that I am not a victim of Evil Genius deception who inhabits a grass-less world. According to Dretske, our sources of good evidence for ordinary beliefs do not provide evidence for their heavyweight implications, such as that I am not a brain in a vat. So, if we insist on holding on to Closure, the price is skepticism. Since we do

not have any evidence for believing the heavyweight anti-skeptical implications of ordinary propositions such as H, we lack knowledge of these implications (such as  $\sim SK$ ). If Closure is a true principle, then it follows that we also lack knowledge of the ordinary implying propositions. So we must choose between denying Closure and embracing skepticism. The former option, says Dretske, is clearly the only game in town!

#### 5. SUSTAINING CLOSURE: STINE, COHEN, KLEIN AND WRIGHT

There is a variant on RA theory on which Closure is sustained. According to Gail Stine, whenever I know that  $\phi$ , my justifying evidence  $e$  will be strong enough to rule out the relevant alternatives to  $\phi$ , so that for such an alternative  $\psi$ ,  $J(ab, e, \sim\psi)$ , thus allowing for  $K(ab, \sim\psi)$ .<sup>17</sup> Stine further agrees with Dretske that  $e$  does not transmit to the contradictories of the irrelevant alternatives to  $\phi$ , such as  $\alpha$ , and so we have  $\sim(J, ab, e, \sim\alpha)$ . But according to Stine, I have *non-evidential* justification for believing  $\sim\alpha$ . So we have  $J^*(ab, \sim\alpha)$  (*Anthony Brueckner has justification for believing  $\sim\alpha$* ).<sup>18</sup> Further, supposing that I believe  $\alpha$ , we have  $K(ab, \sim\alpha)$ . The same holds for all the other irrelevant alternatives to  $\phi$ : I have non-evidential justification for believing their contradictories in virtue of their irrelevance, and I am thus in a position to know these contradictories. So, on Stine's tweak of RA, Closure holds good.

Let us turn to some non-RA-based approaches to Closure which, like RA, are motivated by Dretskean intuitions about the apparent limitations of our justifying evidence for ordinary propositions like Z and H. Stewart Cohen, for example, is strongly inclined to endorse Closure but is uneasy when confronted with the question: What is my evidence for believing  $\sim SK$ ?<sup>19</sup> Cohen thinks that in general we lack evidence that justifies rejecting radical skeptical hypotheses such as SK. But he thinks that we *can* know  $\sim SK$  given the *non-evidential, a priori rationality* of believing  $\sim SK$ . Cohen is not sanguine, though, about this way of sustaining Closure,

since he finds the notion of an a priori belief of a *contingent* proposition to be puzzling.<sup>20</sup>

Peter Klein agrees with part of Dretske's take on the zebra case.<sup>21</sup> Klein agrees that my ordinary justifying evidence E for Z does not constitute an *adequate source of justification* for believing  $\sim$ CD. Yet, once Z becomes justified for me in virtue of my coming to possess E, the proposition Z *itself* becomes available as an adequate source of justification for believing further propositions that are appropriately related to Z. Since Z entails  $\sim$ CD, Z itself constitutes an adequate source of justification for believing  $\sim$ CD. So even though  $\sim$ J(ab,E, $\sim$ CD), as Dretske maintains and Klein agrees, we nevertheless have J(ab, $\sim$ CD), contrary to what Dretske maintains. Closure is thus not violated. A *prima facie* worry about Klein's view is that it seems that we are somehow *manufacturing* justification: in moving from E to Z to  $\sim$ CD, it seems that I am wringing more out of my evidence than is there, since E does not adequately justify  $\sim$ CD.

Though he does not challenge Closure, Crispin Wright thinks that there are situations in which one's justification for believing  $\phi$  fails to transmit to some entailed consequence  $\psi$ .<sup>22</sup> Suppose that I reason as follows:

- a That man kicked a soccer ball into the net.
- So:
- b That man scored a goal.
- So:
- c There is a soccer game in progress.

Wright holds that, since b entails c, it follows that if I am justified in believing b, then I am also justified in believing c. Again, Closure is not violated. However, my justification for believing b fails to *transmit* from b to c in the following sense: if I lack justification for believing c at the beginning of my reasoning, then my reasoning will not enable me to *acquire* justification for c that I antecedently lacked. This is because my premise a provides justification for b only in the presence of an *antecedent* justification for believing c. So my reasoning from a to c cannot serve to *generate* a justification for c that was not already present before the reasoning occurred.

Now consider the following G. E. Moore-inspired reasoning:

- MOORE I: I am having an experience as of seeing my hands.
- So:
- MOORE II: I have hands.
- So:
- MOORE III: I am not a handleless brain in a vat.

As with the soccer reasoning, there is a failure of transmission of justification from MOORE II to MOORE III, even though, if I have justification for the former, I also have justification for the latter. Wright thinks that MOORE I justifies MOORE II only if I have antecedent justification for MOORE III. That is why the Moorean reasoning cannot *generate* justification for its conclusion. Wright thinks that I *do* have antecedent justification for the conclusion. As with Cohen, Wright holds that the justification in question is *non-evidential* in character. Wright calls the form of justification in question an *a priori entitlement*. This is an unearned, unacquired justification that I possess for believing Wittgensteinian *hinge propositions* such as: MOORE III, the proposition that the universe has existed for more than three minutes, the proposition that my senses are reliable, and so on.<sup>23</sup> So Closure will not be contravened in virtue of my knowing MOORE II *without* knowing MOORE III, even though justification is not transmitted from the former to the latter in the course of the Moorean reasoning.

A worry about the views of Stine, Cohen, and Wright: one cannot help but think that these philosophers are simply helping themselves to a hefty chunk of unearned, non-evidential, a priori justification for believing the contradictories of skeptical hypotheses – propositions which our ordinary evidence, on their views, simply does not reach!<sup>24</sup>

## 6. MOOREAN APPROACHES TO CLOSURE AND SKEPTICISM: SOSA AND PRYOR

Ernest Sosa defends what he calls a Moorean approach to Closure and skepticism.<sup>25</sup> Consider the following inconsistent triad:

- (x)  $K(ab, H)$ .
- (y)  $\sim K(ab, \sim SK)$ .
- (z) If  $K(ab, H)$ , then  $K(ab, \sim SK)$ .

All three propositions are plausible, but not all three can be true. Sosa's Moorean stance is to embrace (x) and the Closure-based (z), and to deny (y). In order to rationalize his Moorean stance, Sosa denies that Nozick's Sensitivity is a necessary condition for knowledge. We have seen that if Nozick were right on this score, then (y) would be true. Sosa offers an alternative to Nozick's Tracking Analysis:

S knows that P iff

- (i) P is true;
- (ii) S believes that P;
- (iii) S believes that  $P \square \rightarrow P$ .

Condition (iii) is known as *Safety*. It requires that S's belief that P could not easily have been mistaken. Even though Safety is the subjunctive contrapositive of Sensitivity, its properties are significantly different. Suppose once more that Ace inhabits a normal world and has a true belief of H. We shall therefore have  $K(Ace, H)$  since, in the worlds close to Ace's in which he again believes H, H is true. Does Ace also know  $\sim SK$ , or is Closure violated as it was on Nozick's analysis? Ace has a true belief of  $\sim SK$ ; and, in the close (so again normal) worlds in which Ace believes  $\sim SK$ ,  $\sim SK$  is true. The worlds that Safety has us consider are not vat-worlds in which Ace's belief of H is mistaken. So Sosa's Safety-based analysis of knowledge appears to *sustain* Closure, thereby clearing the way for the Moorean stance.

However, Peter Murphy has shown that a variant of Kripke's barn example can be used to illustrate how Safety engenders a *violation* of Closure.<sup>26</sup> Recall that in Many Blue Fake Barns Country, there are 1,000 blue fake barns and just one real barn – the red one that I see. So I have a true belief of RB. This belief is safe because, in the close worlds in which I again believe RB on the basis of perception, RB is true: there are no close worlds in which I believe RB on the basis of seeing a *red* fake barn. So we have  $K(ab, RB)$ . Do we also have  $K(ab, B)$ , as Closure requires? No, because the close worlds in which I believe B on the basis of seeing a

barn-like structure include many worlds in which I see one of the predominant blue fake barns – so B is false in those worlds. So Closure turns out to be violated on Sosa's analysis.

Another Moorean is James Pryor, who holds that a visual experience as of hands provides *prima facie* justification for my belief of H (justification which can be defeated in various ways).<sup>27</sup> According to Pryor, this justification unproblematically transmits to my belief of H's entailed consequence  $\sim SK$ . *Contra* Wright, I need not possess an *antecedent* justification for believing  $\sim SK$  in order for my hand-experience to do its justifying work. As we have seen, though, there are intuitively powerful worries about how hand-experience can justify rejecting the skeptical hypothesis that I am a handless brain in a vat having vivid hand-experience.

The materials for Pryor's answer to this worry must be found somewhere in his account of how my hand-experience provides *prima facie* justification for my belief of H. According to Pryor, perceptual experiences have *propositional content* and thus provide justification for the associated perceptual beliefs which share their content. So my hand-experience has the content *I have hands* and accordingly justifies my belief of H. However, my belief of  $\sim SK$  obviously differs in content from my belief of H. So it is not at all clear how my hand-experience, which shares its content with my belief of H but not with my belief of  $\sim SK$ , can provide justification for the *latter* belief.<sup>28</sup>

## 7. A PROBLEM FOR THE CLOSURE-BASED SKEPTICAL ARGUMENT

Suppose that we grant the truth of the version of Closure that is employed in the skeptical argument that we have been considering:

- CL3: Necessarily, for all  $S, \phi, \psi$ : if S knows that  $\phi$ , and S knows that  $\phi$  entails  $\alpha$ , and S believes that  $\psi$  on the basis of an inference from  $\phi$  and ( $\phi$  entails  $\psi$ ), then S knows that  $\psi$ .

Marian David and Ted A. Warfield have noted that our skeptical argument suffers from

what we might call a *restriction problem*.<sup>29</sup> The difficulty is that, since we are relying on CL3 in our formulation of the argument, we needed to include the following premise:

- (2)  $\{K(ab, [H \text{ entails } \sim SK]) \text{ and } B(ab, \sim SK, H, [H \text{ entails } \sim SK])\}$ .

Now, while premise 2 is in fact true of me, a corresponding premise will not be true of the man in the street – call him *Street* – who has not considered SK. So we shall not be able to formulate a sound argument that (a) parallels the one that shows that  $\sim K(ab, H)$ , and (b) shows that  $\sim K(\text{Street}, H)$ . Our form of skeptical argument is restricted to putative knowers such as me, who have considered the skeptical hypothesis SK, recognized that H entails  $\sim SK$ , and have come to believe  $\sim SK$  on the basis of the pertinent deduction. So our skeptical argument cannot be generalized to *all* putative knowers.

Accordingly, let us consider a closure principle for *justification*:

CLJ: Necessarily, for all  $S, \phi, \psi, e$ : if S is justified in his belief that  $\phi$  by his evidence e, and  $\phi$  entails  $\psi$ , then S has justification for believing  $\psi$ .

I can *have justification* for believing  $\phi$  in a case in which I do not believe  $\phi$  and hence lack a *justified* belief of  $\phi$ . For example, my evidence strongly points towards the conclusion that  $M = \text{Smith is the murderer}$ , but I have not put my evidence together in such a way as to see this implication and so do not believe M. In this case, I *have justification* for believing M in virtue of possessing good evidence for this proposition, though I do not believe M and a fortiori lack a *justified* belief of M. In this case, if I were to believe M on the basis of my good evidence for this proposition, then that belief would be justified, holding fixed the rest of my epistemic situation. The idea behind CLJ, then, is that if my evidence e justifies my belief that  $\phi$ , and  $\phi$  in fact entails  $\psi$  (which entailment I may or may not notice), then I *have justification* for believing the entailed  $\psi$ , though I may or may not believe  $\psi$ . If I do not believe  $\psi$ , then it is still true that if I were to

believe  $\psi$  on the basis of e and  $\phi$ 's relation to  $\psi$ , then I would be justified in believing  $\psi$ .

It is important to note that if we set  $\psi = \sim SK$ , then the appropriate instance of CLJ does not suffer from the restriction problem: its application is not restricted to putative justified believers who have considered SK, know the pertinent entailment, and appropriately deduce  $\sim SK$ . So what would a skeptical argument employing CLJ look like? Let " $J(S, E, \phi)$ " continue to stand for S's *belief that is justified by his evidence E*, and let " $J^*(S, \phi)$ " stand for S *has justification for believing that  $\phi$* :

- (1') If  $J(ab, e, H)$  and  $(H \text{ entails } \sim SK)$ , then  $J^*(ab, \sim SK)$ .
- (2')  $(H \text{ entails } \sim SK)$ .
- (3')  $\sim J^*(ab, \sim SK)$ .
- (4')  $\sim [J(ab, e, H) \text{ and } (H \text{ entails } \sim SK)]$ .
- (5')  $\sim J(ab, e, H)$ .
- (6')  $\sim K(ab, H)$ .

Premise 1' is an instance of CLJ. Premise 2' is obvious. Premise 3' says that I do not *have justification for believing*  $\sim SK$  and so does not imply that I in fact believe  $\sim SK$ . The pertinent UP-based reasoning for 3' is given above in section 2. 4' follows from 1' and 3' by Modus Tollens. 5' says that I am not justified in my belief that I have hands by my evidence (whatever it is). 5' follows from 4' and 2' by propositional logic. The conclusion 6' follows from 5' given the assumption that knowledge requires justification.

So we have constructed a viable revised Closure-based skeptical argument that generalizes to all putative knowers. As we have seen, though, there are Dretskean considerations about our evidence for our ordinary beliefs of propositions like H that militate against acceptance of the argument's underlying Closure principle. The skeptic, of course, will embrace CLJ in the face of such considerations and simply draw the dreaded skeptical conclusion.<sup>30</sup>

## NOTES

- 1 CL3 needs to be further restricted in view of the fact that, for any  $\phi$  you choose, every necessary

$\psi$  will be entailed by  $\phi$ . We cannot solve this problem by simply restricting CL3 to non-necessary  $\phi$ s and  $\psi$ s, for then the restricted principle will not apply to, e.g., reasoning in mathematics and logic. Another restriction: we need to assume that when S makes his inference to  $\psi$  he retains his knowledge of  $\phi$  and of the pertinent entailment.

2 We can also formulate *Multiple Premise Closure*

MPC. Necessarily, for all  $S, \phi_1 - \phi_n, \psi$ : if S knows each of the members of  $\{\phi_1 - \phi_n\}$ , and S knows that those premises jointly entail  $\psi$ , and S believes that  $\psi$  on the basis of an inference from those premises to  $\psi$ , then S knows that  $\psi$ .

For discussion of MPC, see Hawthorne, 2004.

- 3 For discussion of UP, see Brueckner, 1994.
- 4 We shall return to the issue of generalization in the last section of this paper.
- 5 See Nozick, 1981.
- 6 Application of Condition III to such cases is an exercise left to the reader.
- 7 See Harman, 1973. In the case, a dictator is assassinated on Friday, and I believe that the assassination has occurred on the basis of widespread news reports. We can fill out the example so that, by a stroke of luck, I am offered to house-sit my friend's backwoods cabin for the weekend. While I am out of town, the news media present false denials of the assassination, which are accepted by everyone in town as true, though I do not get wind of the denials. Many have the intuition that I do not know that the assassination has occurred when I am out of town, even though I have a true belief on the matter. This intuition can be explained by the fact that, in the close worlds in which the assassination occurs, I do not leave town and accordingly mistakenly believe that the assassination did *not* occur.
- 8 As I proceed, I will simply speak of *Closure*, having in mind principles which are in the neighborhood of CL3.
- 9 See Brueckner, 1991.
- 10 See DeRose, 1995.
- 11 See Brueckner, 1984.
- 12 The example is given in unpublished lectures on Nozick. I have filled in the details in my own fashion.
- 13 Thanks to Matthias Steup for help in reconstructing Kripke's example.
- 14 See Dretske, 1970.
- 15 For a similar view of the crucial distinction, see Dretske, 1981.

- 16 See Dretske, 2005.
- 17 See Stine, 1976.
- 18 We shall discuss  $J^*$  in more detail below.
- 19 See Cohen, 1999.
- 20 Cohen notes that his contingent a priori beliefs cannot be accounted for by Kripkean considerations about reference-fixing.
- 21 See Klein, 1995 and 1981. For a critical discussion, see Brueckner, 2000.
- 22 See Wright, 2000 and 2003; and Davies, 2000.
- 23 For a recent discussion of how to argue that we have justification for believing hinge propositions, see Hazlett, 2006. For a critical discussion, see Brueckner, forthcoming a.
- 24 This worry does not apply equally to Klein's view. As we have seen, he offers a theoretical basis for the claim that we have justification for believing things like  $\sim CD$  even though our evidence for the entailing does not reach to the entailed consequence.
- 25 See Sosa, 1999.
- 26 See Murphy, 2005.
- 27 See Pryor, 2000.
- 28 For a critical discussion, see Brueckner, forthcoming b.
- 29 See David and Warfield, forthcoming.
- 30 For further discussion of these issues, see Brueckner, 1985.

# BIBLIOGRAPHY

- Brueckner, A.: "Why Nozick Is a Sceptic," *Mind* 92 (1984), 259–64.
- Brueckner, A.: "Skepticism and Epistemic Closure," *Philosophical Topics* 13 (1985), 89–117.
- Brueckner, A.: "Unfair to Nozick," *Analysis* 51 (1991), 61–4.
- Brueckner, A.: "The Structure of the Skeptical Argument," *Philosophy and Phenomenological Research* 54 (1994), 827–35.
- Brueckner, A.: "Klein on Closure and Skepticism," *Philosophical Studies* 98 (2000), 139–51.
- Brueckner, A.: "Hinge Propositions and Epistemic Justification," *Pacific Philosophical Quarterly* (forthcoming [a]).
- Brueckner, A.: "Experiential Justification," in *Epistemology: New Essays*, ed. Q. Smith (forthcoming [b]).
- Cohen, S.: "Contextualism, Skepticism, and the Structure of Reasons," *Philosophical Perspectives* 16 (1999), 57–89.
- David, M. and Warfield, T. A.: "Knowledge-closure and Skepticism," in *Epistemology: New Essays*, ed. Q. Smith (forthcoming).



- Davies, M.: "Externalism and Armchair Knowledge," in *New Essays on the A Priori*, ed. P. Boghossian and C. Peacocke (Oxford: Oxford University Press, 2000), 384–414.
- DeRose, K.: "Solving the Skeptical Problem," *Philosophical Review* 104 (1995), 1–52.
- Dretske, F.: "Epistemic Operators," *Journal of Philosophy* 67 (1970), 1007–23.
- Dretske, F.: "The Pragmatic Dimension of Knowledge," *Philosophical Studies* 40 (1981), 363–78.
- Dretske, F.: "The Case against Closure," in *Contemporary Debates in Epistemology*, ed. M. Steup and E. Sosa (Malden, MA: Blackwell Publishing, 2005), 13–26.
- Harman, G.: *Thought* (Princeton, NJ: Princeton University Press, 1973).
- Hawthorne, J.: *Knowledge and Lotteries* (Oxford: Oxford University Press, 2004).
- Hazlett, A.: "How to Defeat Belief in the External World," *Pacific Philosophical Quarterly* 87 (2006), 198–212.
- Klein, P.: *Certainty* (Minneapolis, MN: University of Minnesota Press, 1981).
- Klein, P.: "Skepticism and Closure: Why the Evil Genius Argument Fails," *Philosophical Topics* 23 (1995), 213–36.
- Murphy, P.: "Closure Failures for Safety," *Philosophia* 33 (2005), 331–4.
- Nozick, R.: *Philosophical Explanations* (Cambridge, MA: Harvard University Press, 1981).
- Pryor, J.: "The Skeptic and the Dogmatist," *Noûs* 34 (2000), 517–49.
- Sosa, E.: "How to Defeat Opposition to Moore," *Philosophical Perspectives* 13 (1999), 141–54.
- Stine, G.: "Skepticism, Relevant Alternatives, and Deductive Closure," *Philosophical Studies* 29 (1976), 249–61.
- Wright, C.: "Cogency and Question-begging: Some Reflections on McKinsey's Paradox and Putnam's Proof," *Philosophical Issues* 10 (2000), 140–63.
- Wright, C.: "Some Reflections of the Acquisition of Warrant by Inference," in *New Essays on Semantic Externalism, Scepticism, and Self-Knowledge*, ed. S. Nuccetelli (Cambridge, MA: MIT Press, 2003).

ANTHONY BRUECKNER

## Contextualism

Contextualism in epistemology is a view according to which the truth conditions for sentences containing the word "knows" vary from one context to another. A context is the

setting or situation in which a sentence is uttered. It is easiest to appreciate the significance of contextualism by first briefly describing the opposing and more traditional view. The traditional, or invariantist, view holds that the truth conditions for sentences containing the word "knows" are fixed. Accordingly, if a person says, "I know that I ate cereal for breakfast this morning", the person says something true just in case certain conditions are satisfied. What those conditions are is a matter of debate. Nearly all philosophers agree that it must be true that the speaker ate cereal for breakfast and that the person believes that he ate cereal for breakfast. What else is required is more controversial. Some say that the belief must be based on very strong evidence, others say that it must result from a sufficiently reliable belief-forming process, and others propose variations on these themes. But what all the invariantist theories agree on is that these standards do not change from one conversational setting to another. Contextualism in epistemology challenges tradition on this point. It holds that certain features of a context, or a setting or a situation, help to determine the truth conditions for sentences containing the word "knows". What is good enough evidence, or sufficient reliability, in one context may not be good enough in another. Versions of contextualism differ over the ways in which contextual factors affect the truth conditions for such sentences. This essay will examine some of the principal versions of contextualism, the considerations offered in its support, and some of the problems it faces.

### 1. SEMANTIC CONTEXTUALISM

Since the view that epistemologists most routinely call "contextualism" is primarily a view about the way the word "knows" and its cognates work, it is appropriate to call it "semantic contextualism". It will be useful to distinguish this view from another, somewhat similar, view which will be discussed towards the end of the article. This other view will be called "subject contextualism".

According to semantic contextualism, sometimes the word “knows” is used in such a way that it takes a great deal for it truly to apply. Other times it is used in a way that is more readily satisfied. In this respect, contextualism is similar to the view that there are two senses of the word “knows”. One is the strong sense and the other is the weak sense (*see* Malcolm, 1952). In its strong sense, “knows” requires something like absolute certainty, which is something we rarely achieve. In its weak sense, it requires strong reasons of the sort we often do have. Defenders of the two-senses view contend that it is often the strong sense that is in use in the philosophy classroom. Given that we are absolutely certain of very little about the world around us, in that setting, when someone says, “I know that I ate cereal for breakfast this morning”, what the person says is false. However, in ordinary settings, “knows” has its weak sense and much of what people claim using that word is true.

Contextualism differs from the two-senses view about “knows” in the following way. The two-senses view holds that “knows” has two distinct meanings. Contextualism holds that it has a whole range of systematically related meanings. Rather than shifting between two fixed meanings, the standards for the application of the word “knows” can vary along a spectrum. Exactly what the relevant varying condition is, and what determines what is required in any particular setting, is something about which contextualists differ. Some proposals are considered below.

A feature of contextualism that merits emphasis is that it focuses on the contexts of the speakers using the word “knows”, not on the contexts of the people they are talking about. To see the significance of this, consider two conversations going on at the same time, one where high standards prevail and another where lower standards prevail. In each conversation, a speaker utters the sentence, “Jones knows that Jones ate cereal for breakfast today”. Even though both speakers refer to the same person and proposition, contextualism implies that the speaker in the high standards context might say something

false while the one in the low standards context says something true. This would happen if the proposition that Jones ate cereal for breakfast that day is true, Jones believes it, and Jones does satisfy the lower standards but does not satisfy the higher standards. Notably, the context Jones himself is in does not matter. It is the ascriber’s context that matters.

There are numerous sentences that seem to function somewhat in the way contextualists claim “knows” functions. Earl Conee (2005) provides a useful example. Consider the sentence “Smith’s thimble is large enough”. Imagine a person looking for a thimble that will fit on his finger. He is told “Smith’s thimble is large enough”. Assuming it will fit on the person’s finger, this sentence is true. Suppose that later the person is looking for something large enough to hold a big ball of yarn. Thinking that Smith has a truly gigantic thimble, someone says, “Smith’s thimble is large enough”. This sentence expresses a falsehood. Both sentences are about the same thimble, and it has not changed size. Rather, what is required to count as “large enough” has shifted.

Other words that seem to show a similar kind of variability are adjectives such as “tall” and “flat”. Suppose that people from a small town in Kansas are looking for a place to get a good view of the surrounding area. They decide to go to the top of the four-story bank building saying, “The bank building is tall”. Everyone agrees, noting that there is nothing for many miles that matches it in height. As they approach the building they strike up a conversation with a stranger who turns out to be a banker from New York City. Looking around the town, the stranger says, “No building around here is tall”. Everyone agrees. A plausible view about these utterances is that when the locals were speaking they used the word “tall” in such a way that it applied to buildings among the tallest in their area. Being three or four stories qualified. When the banker spoke, the standards were much higher. Perhaps only buildings fifty stories high, or even more, qualified. It seems that the conditions for the application of the word “tall” have shifted.

There are a variety of ways in which one might attempt to spell out the details of this shift. One possibility is that sentences of the form “X is tall” are shorthand for something of the form “X is tall for a Y”. We do not bother to say the “for a Y” part in typical conversations since the context usually enables hearers to figure out what is intended. On one version of this view, the word “tall” expresses a relation that could also be expressed by phrase along the lines of “is greater in length than”. The word itself does not shift meaning, but the sentences it is in do shift owing to shifts in the unstated second term of the relation. A variation on this view holds that the word “tall” expresses different one-place properties on different occasions of use. Sometimes it expresses the property of being a tall building, sometimes it expresses being a tall Kansas building, and sometimes it expresses other properties in the family of tallness properties. Other words that are at least somewhat like “tall” are “healthy”, “wealthy”, and “wise”, as well as many others.

A different category of words to which “knows” is sometimes likened includes indexicals such as “I” and “now”. If you say “I am hungry now”, then “I” refers to you and “now” refers to the time at which you speak. If I then say, “Well, I am not hungry now”, then my use of “I” refers to me, not to you. And, if after eating, you say, “Now I am not hungry”, then “now” refers to the later time. None of these three utterances contradicts another. They could all be true. The referents of “I” and “now” shift around.

The variation displayed by the words in the examples just described is something that we routinely and unproblematically cope with in ordinary conversation. No one would react to your earlier and later statements about your hunger by angrily demanding that you finally make up your mind about things. Even politicians, who often view changing their minds as a weakness not to be admitted, are not bothered by this.

Epistemological contextualism is of interest to epistemologists not just because it is about the word “knows”. A second key part of the view is the claim that the context sensitivity of “knows” is relevant to the resolution of

interesting epistemological puzzles. This will be discussed in the next section.

## II. ARGUMENTS FOR CONTEXTUALISM

The main consideration cited in support of contextualism is that it provides a good way to respond to a difficult philosophical puzzle (see Cohen, 1988). The puzzle arises from consideration of three jointly inconsistent propositions, each of which seems to be true. The propositions are about some ordinary empirical propositions, such as the proposition that the lights are on or the proposition that there are trees in the backyard. Skeptical alternatives to these ordinary propositions are hypotheses such as the proposition that the lights are not on (but I am merely hallucinating that they are on) or the proposition that there are no trees (but I am a brain in a vat who is being fed tree perceptions). The puzzle then involves these claims:

- C1: We know some ordinary empirical propositions to be true.
- C2: We do not know the skeptical alternatives to these propositions to be false.
- C3: If we do not know the skeptical alternatives to be false, then we don’t know the ordinary empirical propositions to be true.

Each of these three propositions seems to be true, but they cannot all be true. A good response to the puzzle will explain which of the three is false, and also explain why we are tempted to think that it is true.

It is not disputable that each of (C1)–(C3) has at least some initial plausibility. All of us think we know ordinary propositions such as that you have hands or that there are tables and chairs, and so on. So (C1) seems to be true. Consider (C2). The skeptical alternatives include hypotheses such as that you are a brain in a vat, or a victim of some massive deception, etc. If asked whether you know that these hypotheses are false, you are likely to feel reluctant to assert that you do. After all, it is not easy to say how you can tell the difference between being an ordinary person



(a brain in a head, we might say) and a brain in a vat being caused to have your seemingly ordinary experiences. So (C2) seems to be true. But now consider (C3). If you do not know that you are no brain in a vat, then, it seems, you must fail to know all (or most) of the ordinary things that in some way depend on your not being a brain in a vat. For example, if you are a mere brain in a vat, then you do not have hands. If you do not know that you are not a handless brain in a vat, then, it seems, you do not know that you do have hands.

There is a complication regarding (C3) that requires brief examination here. (C3) is often taken to be true because it is a consequence of a very plausible principle about knowledge known as the “closure principle”. The general idea of the closure principle is that if a person knows one proposition, and knows that this proposition logically entails some other proposition, then the person knows, or at least is in a position to know, the entailed proposition. (C3) is not exactly an instance of this principle, since (C3) does not include anything corresponding to the point about knowledge of the entailment. For present purposes, it will suffice to assume that “we” in (C1)–(C3) includes only people who do know the relevant logical connection between ordinary empirical propositions and properly formulated skeptical hypotheses. Perhaps not everyone does have that logical knowledge, and perhaps (C3) would not be true if it were taken to apply to people lacking that knowledge. But (C3) does seem to be true when restricted to people who do know that ordinary propositions entail the falsity of the skeptical alternatives. Thus, the puzzle most clearly arises with respect to people with that knowledge. Since many people do have that knowledge, the puzzle does arise with respect to many people. With respect to them, one of (C1)–(C3) must be rejected.

The puzzle involving (C1)–(C3) can also be formulated as an argument for skepticism (see DeRose, 1995). If we have good reason to believe that both (C2) and (C3) are true, then we can use them as premises in an argument that has the denial of (C1) as its conclusion. This will be a valid argument, it will have seemingly plausible premises, and it will have

a distressing skeptical claim as its conclusion. In order to deal with the complication mentioned in the previous paragraph, it will be best to assume the argument is only about people who do know the logical connections between ordinary propositions and skeptical alternatives. The conclusion of the argument, then, is not the fully general skeptical claim that no one knows any empirical propositions, but the more limited skeptical conclusion that people with this knowledge of logical connections do not know empirical propositions to be true. Since many of us do know that the ordinary propositions imply the falsity of the properly formulated skeptical hypotheses, the skeptical argument calls into question our knowledge of the ordinary propositions.

Contextualists contend that their theory provides the best analysis of the inconsistent triad and the best response to the argument for skepticism. To appreciate the merits of contextualism, it will be helpful to understand the liabilities of rival views. Non-contextualist responses are of three sorts. The skeptical response denies (C1). It holds that we do not in fact know much. Defenders of this response owe us an explanation of why it seems so obviously correct to ascribe knowledge in ordinary contexts. The word “knows” seems to do a job for us, a job that will have to be done by some word in any case, yet skeptics are denying that it does this job. For many, this is too much to accept.

Many philosophers reject (C2). They say that we do know that the skeptical hypotheses are false. Critics contend that this is dogmatism. They will ask, with the appropriate tone of disbelief, “How can you know that you are not a brain in vat? After all, things would look exactly as they do if you were a brain in a vat”.

The third non-contextualist response to the argument denies (C3). This response relies upon the idea that the closure principle to which the argument appeals is incorrect. There is a large literature on the closure principle which decisively illustrates that it is difficult to formulate a precise version of the principle that is immune to worrisome objections. Nevertheless, to many, it seems clear that some sort of closure principle is true. At the

very least, many philosophers say, the particular application of the closure principle used in a suitably restricted version of this argument is beyond dispute. To say otherwise is to say that a person who knows that the proposition that he has hands entails that he is not a handless brain in a vat can be in the following situation: he knows that he has hands yet he does not know that he is not a handless brain in a vat. DeRose (1995) terms this “the abominable conjunction”. It is an implausible result, whatever the general merits of the closure principle.

The contextualist response is that the truth values of the sentences in the inconsistent triad, and in the skeptical argument, depend upon the context. According to contextualism, in ordinary situations, when ordinary standards are in place, (C1) is true and (C2) is false. Relative to those ordinary contexts, the skeptical argument has a false premise. Contextualism thus preserves our ordinary claims using the word “knows”. But when the standards are raised, attributions of the word “know” have more stringent standards, and (C1) is false while (C2) is true. Relative to those contexts, the argument is sound. Thus, contextualism recognizes the power of the skeptical considerations, and it admits that the denials made in the epistemology classroom, or in other situations in which higher standards have been invoked, are true as well. Contextualists are thus free to say that (C3) is true in all contexts, thus avoiding the abominable conjunctions. In no situation are all three elements of the inconsistent triad true. The puzzle arises because we fail to keep track of changes in context.

The central argument for contextualism is that the contextualist response to this puzzle is the best response. The fact that “knows” is being claimed to function in a way similar to other familiar words, such as “large enough”, “tall”, and indexicals, lends credibility to the response.

### III. DETAILS

A well-developed semantic contextualist theory will include two main components.

One component will be some general account of knowledge. This account will include among its proposed truth conditions for sentences containing “knows” an element suitable for the desired kind of variation. This might be degree of evidential support or reliability. The second component of a well-developed contextualist theory will be some account of what contextual features bring about these variations in truth conditions.

A contextualist can defend nearly any general theory of knowledge. Cohen (1988, 1999) defends a fairly traditional evidentialist (good reasons, justificationist) account of knowledge. He holds that a person has knowledge only if the person has adequate reasons for his true belief. But how good one’s reasons have to be in order to be justified, and thus to know, varies.

Lewis (1996: 554) defended the following account of knowledge:

*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.*

The contextually variable element here involves which possibilities we can “properly ignore”. Before going into that, it will be helpful to understand the idea of a possibility being eliminated by one’s evidence. On Lewis’s view, “the uneliminated possibilities are those in which the subject’s entire perceptual experience and memory are just as they actually are” (p. 553). To illustrate this, consider a typical person, Jones, looking into his backyard. He believes that he sees a maple tree. Among the alternatives to his seeing a maple tree are possibilities in which he sees an oak tree or a pine tree. These alternatives are eliminated by his evidence: the (normal) possibilities in which he sees trees of those kinds are ones in which his experiences differ from his actual experiences. There are, of course, possibilities in which the person has just the perceptual and memorial experience he has but is not actually seeing a maple tree. These are possibilities in which he is the victim of some sort of hallucination or deception. However, in ordinary settings, such possibilities are properly ignored. In some extraordinary settings,

they are not. Thus, in the ordinary setting, “Jones knows that he sees a maple tree” comes out true; but, in settings in which these other possibilities are not properly ignored, the same sentence comes out false.

The context sensitivity associated with “properly ignore” has its analogues in other cases. Suppose the chair of a committee calls a meeting to order by saying, “Everyone is here now, so we can begin”. Only a wise guy would point out that lots of people, in fact almost all the people in the world, are not present. In evaluating her statement, we are properly ignoring all these people and considering only the people on the committee. In a different context, of course, we could not properly ignore some of these people. Lewis thinks that the same is true of the possibilities that must be eliminated in order to have knowledge. Ordinarily, ridiculous alternatives concerning brains in vats and deceptive demons are properly ignored. But sometimes they must not be ignored.

Lewis provides a detailed set of rules describing which possibilities are properly ignored. Among them are the rule of actuality, which says that the possibility that is actual cannot be ignored. In addition, he says that a possibility that is believed by the subject cannot be properly ignored, nor can others that saliently resemble a possibility that is not ignored, nor can any possibility that is actually being attended to. He provides additional rules that will not be described here.

Keith DeRose’s (1995) contextualist theory works a little differently. Ignoring some complexity that will not matter for the present discussion, DeRose’s idea is as follows. Some of the true propositions that a person believes are such that the person would believe them if they were false. Suppose that wishful thinking leads you to believe that it will be sunny tomorrow. This belief is in fact true. But, if it were false, you would believe it anyway. DeRose says that beliefs like this are *insensitive*. In ordinary situations, if a true belief is sensitive, then it is a case of knowledge. Your belief that you have hands is sensitive – if you did not have hands, then you would not believe that you have hands. So you know that you have hands. Next, consider your belief that

you are not a handless brain in a vat. Granting some plausible assumptions, this true belief is not sensitive: if it were false – if you were a handless brain in a vat – things would seem just as they actually do and you would believe that you are not a handless brain in a vat. So you do not know that you are not a handless brain in a vat.

DeRose’s account may seem to be a recipe for denying the closure principle described earlier, but DeRose (1995: 36) avoids that result by appeal to the following rule:

When it is asserted that some subject knows (or does not know) some proposition P, the standards for knowledge tend to be raised, if need be, to such a level as to require S’s belief in that particular proposition P to be sensitive for it to count as knowledge.

DeRose means here that the standards for “knows” with respect to any proposition are raised. Thus, once someone says, “You do not know that you are not a handless brain in a vat”, the standards for properly being said to “know” any proposition are raised. In that context, “You know that you have hands” comes out false because you would believe that you do have hands in the now relevant worlds in which you are a handless brain in a vat. Since the more demanding standards thus apply to all attributions involving the word “knows”, knowledge sentences in those contexts will typically come out false.

There are additional ways in which contextualism can be developed, and the defenders of these versions have provided additional details. However, the accounts provided here give an adequate sense of the way the theory works.

#### IV. OBJECTIONS

Objections to semantic contextualism fall into several categories. Some apply primarily to one or another of the specific contextualist theories. One might argue that particular views about what affects contexts are incorrect or that a particular version of contextualism relies on an underlying view about knowledge that is mistaken. Thus, for example,

those who think that traditional evidentialist theories of knowledge and justification are mistaken will most likely find that their objections carry over to Cohen's evidentialist contextualist theory. Objections of this sort will not be considered here. Other objections to contextualism are designed to apply equally to all versions of the theory. Several of these objections will be considered here.

Some of the puzzling consequences of contextualism involve cases in which speakers in a context with one sort of standard for attributions of "knows" report, discuss, or think about claims made in contexts with different standards (see Hawthorne, 2004, ch. 2). Suppose that Professor Jones is in her epistemology classroom discussing skeptical arguments. According to all the contextualist theories described previously, when she and her students use the word "knows" in this context the more-difficult-to-meet standards are in place, and most sentences in which they attribute "knows" to someone will come out false. Suppose that, in an effort to make clear to her students the profoundly counter-intuitive implications of skepticism, she describes a common example. She tells a little story about her colleague who had made a simple claim based on memory. She reports that:

Professor Smith said that he knows that he taught his logic class this morning.

Assume that Professor Smith was in a more standard context when he uttered his sentence, "I know that I taught my logic course this morning". The objection to contextualism is that it implies an incorrect interpretation of what Jones tells her class about what Smith said. Smith, recall, made his claim at ordinary standards. But Jones is in a high-standards context. It seems to follow that, when she uses the word "knows" in reporting what Smith said, the word has its high-standards interpretation. After all, she is the one using the word at that time, and she is in her high-standards context. But then her report says, incorrectly, that Smith made the high-standards claim.

This objection is similar to another one that suggests that contextualism implies that

there are not conflicts or differences of opinion in cases when there really are differences (see Feldman, 2001). Suppose that one of Professor Jones's students is convinced by the arguments for skepticism. He thinks back to his earlier, pre-skeptical, remarks, and notes that earlier in the day he had said, "I know quite a lot about the solar system". Thinking back on this remark, he now thinks that it was mistaken. However, contextualism seems to imply that this student makes a mistake when he says that his earlier remark was mistaken. Although the same sentence would be false in his current context, it expressed a truth back in its original setting. This implies that his later self – the one who says "I do not have knowledge" – does not disagree with his earlier self – the one who said "I know a lot". This is because, according to contextualism, his later high-standards claim does not conflict with his earlier low-standards claim. Yet one would think that the student is in a position to determine whether he has changed his mind. Suppose that he were to say to the teacher: "You have convinced me that skepticism is correct. I used to think that I knew things, but I have changed my mind". The contextualist analysis of this situation implies that he has *not* changed his mind. The objector thinks contextualism has gone wrong here. The student could not be mixed up about whether he changed his mind.

Underlying these objections is the idea that contextualism attributes to people an unacceptable amount of *semantic blindness*. That is, it says that they fail to realize the meanings of their own words (see Schiffer, 1996). In this respect, "knows" seems to be unlike some of the other context-sensitive terms to which it has been compared. No one who understands the language thinks that a person contradicts himself when he first says, "I am hungry now", then eats a large meal, and then says, "I am not hungry now". This is because we all realize, at least implicitly, that indexicals like "now" shift reference. Words like "tall" may be a little more confusing, but it is easy to sort out and understand seemingly conflicting uses by adding qualifiers such as "for a building in Kansas" or "for a building in New York City" to sentences such as "That

building is tall". No such qualifiers are normally used with "knows", and we seem unaware of its alleged contextual shifts.

There are other sentences that potentially pose puzzles for contextualists. If the phrase "skepticism about topic N" means the same as "the view that there is no knowledge about topic N", then the word "skepticism" may display context sensitivity as well. But then peculiar things follow. Suppose that while in ordinary standards a student looks at her course syllabus and notices that the topic is "skepticism". She thinks, "I am going to learn about skepticism in class today". But then when she goes to class the teacher introduces skeptical hypotheses and changes the truth conditions for "knows" and thus the view referred to by "skepticism". He then discusses this second view rather than what the student thought he was going to discuss. Contextualism seems to have the odd implication that the teacher did not discuss the view the student thought that he would discuss. And if, later that day and back in ordinary standards, the student says, "I learned about skepticism today", the theory implies that she is mistaken. Similarly, suppose the teacher begins the discussion by writing (C1) – we know some ordinary empirical propositions – on the blackboard, and then says that skepticism denies this. This seems to be a low-standards assertion, since skeptical hypotheses have not been mentioned. But if the teacher then goes on to mention the skeptical hypotheses, and changes the context, the ensuing discussion will not be about the view introduced at the beginning of class. This seems to be an incorrect analysis of the situation.

Contextualists are not without possible responses to puzzles like these. To the latter claims, they may argue that skepticism is not a view about whether people have knowledge but instead a view about the truth value of sentences containing the word "knows". They are not forced to say that "skepticism" inherits the context sensitivity of "knows". However, even if they are right about the word "skepticism", it does seem that they are stuck with the result that much of what the teacher says at the beginning of class about

what he will discuss may prove to be false because the context shifts once the skeptical hypotheses are introduced.

With respect to the earlier points, it is possible to develop qualifiers to add to sentences with "knows" just as we do for other sentences. For example, one could imagine locutions such as "knows for practical purposes" or "knows with scientific certainty" or "knows with absolute certainty" gaining clearer meanings and more widespread use. Perhaps, then, it is possible that the language could evolve in such a way that "knows" is more like "tall".

Furthermore, it must be admitted that every theory has to acknowledge something like semantic blindness. Clearly, people regularly say that they and others know things, and that people seem to take these sentences to be literally true. Skeptics have to say that they are routinely mistaken. Thus, they must say that the ordinary claims invoking "knows" are exaggerations or approximate truths that we allow to go by unchallenged in polite discussion. But people are typically oblivious to all this. And dogmatists must explain away the reluctance so many people have to predicate "knows" of themselves and others in the face of skeptical hypotheses. This reluctance reveals that people must be making some kind of error, either about what is required to apply "knows" correctly or about what obtains in the situation, to explain their unwillingness to say what is, according to the theory, actually true. Presumably, dogmatists will say that people are unduly swayed by skeptical considerations when they become salient, and they will attribute our varying willingness to apply "knows" to vacillating thoughts about the merits of the considerations for and against a single proposition.

A different kind of worry about contextualism concerns its epistemological significance. (See Sosa, 2000, for discussion of this general theme.) Contextualism purports to help us understand our puzzlement about skepticism. And it does propose an explanation of some shifting inclinations to apply "knows" to people and propositions. However, in another way, it seems to evade the issue. When the most powerful skeptical arguments



are presented, they point out potential weaknesses in our reasons for believing many of the things we ordinarily take ourselves to know. In such settings, people often find themselves confused. On the one hand, they think they know quite a lot, but the worries about the merits of their evidence are troubling. Contextualism, as it is typically presented, resolves this puzzlement clearly and decisively in favor of skepticism. After all, according to contextualism as it has been developed, the skeptical argument is sound when it is fully articulated in the philosophy classroom. In that setting, the right response to the puzzle is to reject (C1), “We know ordinary empirical propositions to be true”. Those who think that there are flaws in the arguments for skepticism will see this as mistaken capitulation to skepticism, even if the word “knows” is context sensitive.

Critics of contextualism find the appeal to shifting standards for “knows” as a part of the solution to epistemological puzzles as analogous to a contextualist solution to what may seem to be a difficult moral puzzle. Suppose that you go through life thinking that some behavior – say, eating meat – is morally permissible. And then you go into an ethics class and contemplate considerations about this practice that are not ordinarily taken into account. You now are puzzled about the moral permissibility of eating meat. A contextualist response to this puzzle, analogous to the stated response to skepticism, would hold that the sentence “Eating meat is not morally permissible” is true in the ethics classroom but false in ordinary circumstances. This seems to be more of an evasion than a solution to the question about eating meat. Similarly, the critics might say, when skepticism is raised in the philosophy classroom, there is a reasonably clear question being considered. It is a question about how good a basis we have for our ordinary beliefs. This question is not about whether we are absolutely certain of ordinary propositions. We clearly are not. The question is about whether we do satisfy some other stringent standards. The skeptical arguments purport to show that our basis for our beliefs is not

very good at all. A good response, as contextualism’s critics see it, requires showing that we do satisfy these standards. The fact, if it is a fact, that the conditions for the applications of the word “knows” shift with contexts is not itself a key part of this response to skepticism.

The significance of these criticisms of contextualism remains a live issue.

## V. SUBJECT CONTEXTUALISM

Semantic contextualism implies that the truth conditions for sentences containing the word “knows” vary according to factors in the speaker’s context. A rival view holds that factors in the context of the subjects of knowledge attributions affect the truth conditions of these sentences. Although epistemologists often restrict the label “contextualism” to the former view, it will be useful to examine briefly this rival view, which can be called “subject contextualism”.

One version of subject contextualism highlights factors very much like those that semantic contextualists emphasize, but contends that these factors in the context of the subject of potential knowers matter. Thus, for example, a subject contextualist might hold that a factor affecting whether a subject knows a proposition is what alternatives to that proposition are relevant in the subject’s context (*see* Hawthorne, 2004). In ordinary circumstances, people do not think about skeptical hypotheses, and they are not relevant alternatives. But when they do think of them they must have good enough basis to rule them out. Since they lack such a basis, they lack knowledge in those circumstances.

Philosophers have proposed a second way in which the subject’s context may affect knowledge (*see* Stanley, 2005). The following sort of example (*see* DeRose, 1992) illustrates the idea. Suppose that Mr A is at the airport waiting for his flight from New York to Los Angeles. He has a printed itinerary saying that the flight stops in Chicago. If asked, Mr A would say that he knows that the plane stops in Chicago. Ms B is sitting next to him.

She is planning to take the same flight, and she looks at the same itinerary. However, she plans to get off the plane in Chicago in order to attend an extremely important meeting. Since the printed itinerary might be mistaken, and so much turns on the matter, Ms B might believe that the plane stops in Chicago but deny that she does know that the plane stops in Chicago. Subject contextualism implies that Mr A's claim to knowledge and Ms B's denial of knowledge can both be correct, even though they have the same evidence for the same proposition. (We are assuming that the proposition is true, and that they both believe it.) The idea is that whether a person has knowledge depends in part upon how important it is that the person be right about the proposition in question. The stakes for Ms B are much higher than they are for Mr A, and thus she needs better reasons in order to know. This view adds a pragmatic element to the conditions on knowledge.

Subject contextualism departs from the traditional invariant view in a different way from semantic contextualism. It has the implication, somewhat similar to an implication of semantic contextualism, that two people who believe the proposition on the basis of the same evidence may be such that one but not the other has knowledge. The corresponding implication of contextualism is that two speakers in different contexts may utter "S knows p", speaking of the same person and proposition, and one of them is right and the other wrong.

Defenders of subject contextualism think that the view has notable advantages. Some think that whether it is appropriate for a person to assert a proposition depends in part upon whether the person knows it. These philosophers contend that it does seem acceptable for Mr A simply to say, "The plane stops in Chicago". In contrast, they say that after wondering about the accuracy of the itinerary it would be somehow inappropriate for Ms B to say (flat out), "The plane stops in Chicago". Anything she says should be more hedged. But, if she did have knowledge, it would be hard to see why she ought not make the assertion. Subject contextualism

explains this by saying that she does lack knowledge, and that is why she should not make the assertion.

Depending upon details of its development, subject contextualism can explain why sensible, thoughtful, competent speakers of English are reluctant to say "I know that I have hands" when they are focusing on skeptical hypotheses such as the hypothesis that they are handless brains in vats. The explanation is that in those circumstances they do not know such things. However, when someone happily realizes that he has escaped injury after the fall in his woodworking shop and says, "I know that I still have hands", this can be true as well. And when the worried traveler described above says that she had better check the flight schedule because "she does not know whether the flight stops in Chicago" this remark is true. But so can the routine comment of the unworried traveler who looks at his printout and says, "I know that the flight stops in Chicago". Defenders of the view find these to be genuine advantages of the theory. Defenders of traditional views have to find ways to explain away these familiar utterances.

Whatever its advantages, subject contextualism has some surprising consequences. Suppose a person is aware of all the facts concerning our story about Mr A and Ms B. According to the theory, that person would be right to say something like "Mr A and Ms B believe exactly the same thing about the flight schedule for exactly the same reasons. But Mr A knows that the flight stops in Chicago and Ms B does not. This is because Ms B has a very important meeting to attend". This strikes many as absurd. Similarly, the theory seems to imply that Mr A could flaunt his greater knowledge to Ms B. While looking at their itineraries, he could truthfully say, "Well, I know that it stops in Chicago, but you don't". Of course, defenders of invariantist theories do need to find a way to explain away our reluctance to say that Ms B does have knowledge. If they think that she ought to double check the flight schedule, then they are committed to accepting the possible truths of things such as "She knows that the plane stops in Chicago, but she ought to

check on it anyway". Defenders of pragmatic theories find this unacceptable.

## BIBLIOGRAPHY

- Cohen, S.: "Contextualism, Skepticism, and the Structure of Reasons," *Philosophical Perspectives* 13, *Epistemology* (1999), 57–89.
- Cohen, S.: "How to Be a Fallibilist," *Philosophical Perspectives* 2 (1988), 581–605.
- Conee, E. and Cohen, S.: "Is Knowledge Contextual?," in *Contemporary Debates in Epistemology*, ed. Matthias Steup and Ernest Sosa (Malden, MA: Blackwell Publishing, 2005), 47–71.
- DeRose, K.: "Solving the Skeptical Puzzle," *Philosophical Review* 104 (1995), 1–52.
- DeRose, K.: "Contextualism and Knowledge Attributions," *Philosophy and Phenomenological Research* 52 (1992), 913–29.
- Feldman, R.: "Skeptical Problems, Contextualist Solutions," *Philosophical Studies* 103 (2001), 61–85.
- Hawthorne, J.: *Knowledge and Lotteries* (Oxford: Oxford University Press, 2004).
- Lewis, D.: "Elusive Knowledge," *Australasian Journal of Philosophy* 74 (1996), 549–67.
- Malcolm, N.: "Knowledge and Belief," *Mind* 61 (1952), 178–89.
- Schiffer, S.: "Contextualist Solutions to Skepticism," *Proceedings of the Aristotelian Society* 96 (1996), 317–33.
- Sosa, E.: "Contextualism and Skepticism," *Philosophical Issues* 10 (2000), 1–18.
- Stanley, J.: *Knowledge and Practical Interests* (Oxford: Oxford University Press, 2005).

RICHARD FELDMAN

## Foundations and Coherence

### 1. THE EPISTEMIC REGRESS PROBLEM

Suppose I believe that P, and I am asked why I believe it. I might respond by citing a reason, Q, for believing P. I could then be asked why I believe Q. I might respond by citing a reason, R, for believing Q. This process might continue for some time. If at any point I admit that one of the reasons I have given itself has no justification, then my whole chain of reasons would seem to collapse, leaving P unjustified. If I continue to supply reasons for my beliefs,

there is a threat of infinite regress. This problem is known as the *epistemic regress problem*. It raises the following question: when one has a justified belief, what is the overall structure of the reasons that one has supporting that belief?<sup>1</sup>

There are four possible answers to this question: (i) The series of reasons might have one or more stopping points, that is, reasons for which there are no further reasons. (ii) The series might be circular, that is, at some point in following back the chain of reasons, we might repeat a reason that appeared earlier. (iii) The series might go on infinitely without repeating. (iv) There might be no justified beliefs at all, in which case there is no series of reasons. At least one of these answers must be correct;<sup>2</sup> this is a logical truth independent of the particular nature of reasons for belief. There are four epistemological theories, corresponding to these possibilities:

- (i) *Foundationalism* holds that, when one has a justified belief, this belief either is itself non-inferentially justified or is based, directly or indirectly, on one or more non-inferentially justified beliefs. A belief is non-inferentially justified if it is justified to some degree in some way that does not depend on one's having reasons for the belief.<sup>3</sup>
- (ii) *Coherentism* holds that all justified beliefs are justified by virtue of the way they fit together with ("cohere with") the rest of one's belief system. This view is commonly interpreted as embracing the legitimacy of circular series of reasons.<sup>4</sup>
- (iii) *Infinitism* holds that, when one has a justified belief, there is an infinite series of reasons supporting it.<sup>5</sup>
- (iv) *Skepticism* holds that there are no justified beliefs.<sup>6</sup>

Foundationalism and coherentism are the two leading views in the field, and it is on them that I shall focus.

### 2. COHERENTISM

The coherence theory of justification locates the source of all justification for belief in the



relation of *coherence*. Typically, a system of beliefs is said to cohere well when it is consistent, many of the beliefs in the system are mutually supporting (that is, entail each other or render each other probable), and the system contains few or no anomalies (claims that have no explanation within the system).<sup>7</sup> While foundationalists may grant that coherence plays a role in enhancing the justification of some beliefs, coherentists hold the stronger thesis that coherence can *by itself* provide justification for belief.

### 2.1 The Probabilistic Argument

Perhaps the central idea behind coherentism is that it would be unlikely that a system of beliefs would cohere if it were not by and large correct; therefore, if a system of beliefs coheres, we have reason to think it is by and large correct. To illustrate the idea, suppose that two witnesses claim to have seen someone commit a crime. The witnesses independently (without consulting each other) report that the culprit had spiked green hair. We need not take the witnesses to be particularly reliable – even if each witness is more likely to make a false report than a true one in matters of this sort, it is nevertheless highly unlikely that the two would agree on such a specific and unusual description of the culprit unless the description were true. The best explanation for the coherence of the witnesses' reports is that they are true.<sup>8</sup>

This sort of example is intuitively forceful. Undoubtedly, the coherence of the witnesses' reports in situations of this kind enhances their credibility. But the question remains whether coherence *by itself* can provide justification for belief, or whether the force of the example depends on the background assumption that the witnesses have at least some degree of individual credibility, that is, that individual witness reports produce at least some justification for believing what is reported. Assume that a witness has individual credibility just in case his making a report raises the probability of what is reported; for example, the probability of the culprit having spiked green hair given that the witness reports this is greater than the prior probability

of the culprit having spiked green hair. It can be shown in terms of probability theory that if each witness lacks individual credibility in this sense, and their testimonies are entirely independent of each other, then their agreement on a single story also does nothing to increase the probability of that story's being correct.<sup>9</sup> The intuitive idea is this: if neither witness has any independent credibility, then the hypothesis that a particular claim is true would *not* serve to any extent to explain why the witnesses agreed on it, since they would be just as likely to agree on that story if it were false.

A second objection to the probabilistic argument for coherentism is that the argument requires some pre-existent knowledge in order to show that the coherence of a set of claims increases their probability of truth. In the above example, one must first know that each witness did in fact *say* that the culprit had spiked green hair before one can argue that the coherence of the reports is evidence of their truth. Similarly, before one can make an analogous argument for the likely accuracy of a coherent belief system, one must first know that the belief system is in fact coherent, which would presumably require also knowing what beliefs the system contains.<sup>10</sup> Most foundationalists would posit foundational knowledge of the contents of one's own beliefs, but a coherentist cannot take this route. The point here is not that this proves coherentism to be false; rather, the point is that the coherentist's central argument fails. The examples used to motivate coherentism do not in fact show how coherence alone can provide justification, since in those examples coherence can provide justification for some beliefs only given that some other beliefs are already justified.

### 2.2 The Input Objection

A common objection to the coherence theory is that the theory seems to provide no role for *experiences* (where these are distinct from beliefs) in the justification of our beliefs. The coherence of a belief system is a function solely of the contents of those beliefs; once the contents of one's beliefs are fixed, no other mental states that one has are relevant

to how coherent one's belief system is. Since the coherence theory traditionally claims that coherence among beliefs is the sole source of justification, the theory is committed to denying that any mental states other than one's beliefs are relevant to justification.<sup>11</sup> This is counter-intuitive – intuitively, an individual with exactly the same beliefs as I have but entirely different sensory experiences would be irrational.

Probably the best response to this objection is for the coherentist simply to include sensory experiences, memories, and other appearance states (mental states by virtue of which something appears to one to be the case) among the items that one's beliefs must cohere with. Thus, justified beliefs will be justified by virtue of their coherence with one's beliefs-cum-appearances, rather than by virtue merely of their coherence with one's beliefs.<sup>12</sup>

### 2.3 The Alternate-Coherent-Systems Objection

There are many coherent systems of propositions that we do not thereby have any reason to accept. A good novel may express a highly coherent set of propositions, but this is no reason to regard those propositions as by and large true. This shows that coherence by itself does not provide justification for belief.

Probably the best response for the coherentist is to grant that the coherence of a set of propositions is not sufficient for its having any degree of justification, that coherence produces justification for belief only given certain further conditions. One condition that might be imposed is that a system of information should contain a significant number of *observations* (these might be interpreted either as a special kind of belief or as a distinct type of mental state).<sup>13</sup> On this view, the statements in a coherent novel fail to gain justification, owing to the absence of any observations among them.

### 2.4 Problems with Defining Coherence

What is it for a belief system to be coherent? The weakest sort of coherence one might expect the coherentist to recognize is that of mere logical consistency – a coherent belief

system should not contradict itself. The strongest sort of coherence one might expect the coherentist to recognize is that of mutual entailment – a belief system is coherent to the extent that every belief in it is entailed by the rest of the system.

Unfortunately, it seems that the weakest notion of coherence is *too strong* for the coherentists' purposes, because it is possible to be justified in believing an inconsistent set of propositions. The lottery scenario provides an alleged instance: it seems that one is justified in believing, of each ticket in a large lottery, that it will lose, yet one is also justified in believing that some ticket or other will win.<sup>14</sup> However, some would deny that one is justified in believing of each ticket that it will lose; perhaps one is only justified in believing of each ticket that it will *probably* lose. Perhaps a more compelling example of justified inconsistent beliefs is provided by my present belief that at least one of my beliefs is false. The evidence for this, I should say, is overwhelming – I have often discovered false beliefs of mine in the past; to the best of my knowledge, every other person in the world has at least some false beliefs; and I have no good reason to suppose myself dramatically less error prone than the rest of the world. But, as long as I believe that I have at least one false belief, my belief system as a whole is incoherent (it is logically impossible that all my beliefs should be true). The coherence theory thus appears to imply, wrongly, that I am unjustified in believing that I have at least one false belief.<sup>15</sup>

At the same time, it seems that the strongest notion of coherence is *too weak*, because it is too easy, starting from any set of propositions, to add further propositions to the set in such a way that every member of the set is entailed by the rest of the set. For instance, suppose I believe two propositions, *P* and *Q*. I may add to my belief system the belief that *P if and only if Q*. Of the set of propositions, {*P*, *Q*, (*P* ↔ *Q*)}, any two entail the third. Since this is true regardless of what *P* and *Q* are, it seems all too easy to generate coherent systems.<sup>16</sup> A coherentist might respond to this worry by proposing that it is only coherence among *independently*

*acquired* beliefs – for instance, beliefs caused by distinct perceptual experiences – that generates justification. If I know that my belief in  $(P \leftrightarrow Q)$  was acquired solely by my inferring it from  $P$  and  $Q$ , then its coherence with  $P$  and  $Q$  does not serve to enhance its justification.

### 2.5 The Circularity Objection

Perhaps the central intuitive objection to the coherence theory is that the theory is committed to endorsing circular reasoning. Suppose I announce that I believe it will snow tomorrow. You ask me why I believe this. I reply: “I think it will snow tomorrow, because it will snow tomorrow.” It seems clear that I have not thereby articulated a possible legitimate justification for my belief. Nor is the situation apparently improved if I make the circle larger and more complex. If I claim that my belief that  $A$  is justified because it is supported by  $B$  and  $C$ , which I also believe; that  $B$  is justified because it is supported by  $A$  and  $C$ ; and that  $C$  is justified because it is supported by  $A$  and  $B$ ; then I have merely engaged in a slightly more complex form of circular reasoning. It seems that the coherence theory of justification involves the same error – the theory merely makes the circle larger and more complicated, so that it perhaps includes all or most of the elements of one’s belief system.

Sometimes a distinction is drawn between *linear* coherentism, which explicitly endorses circular reasoning, and *holistic* coherentism, which allegedly avoids such circularity. The holistic coherentist holds that justification applies in the first instance, not to individual beliefs, but to a whole *system* of beliefs.<sup>17</sup> But this hardly seems to make any difference – suppose that I modify my earlier account as follows: I announce that the *system* consisting of  $A$ ,  $B$ , and  $C$  is justified because  $B$  and  $C$  support  $A$ ,  $A$  and  $C$  support  $B$ , and  $A$  and  $B$  support  $C$ . This hardly disguises the circularity. If we found it implausible to maintain that individual beliefs can be justified by circular reasoning, how plausible is it to claim that a *system* of beliefs can be justified by virtue of the very sort of logical relations among its members that such circular reasoning exhibits?

This argument might be thought to beg the question, because the premise that circular reasoning is unacceptable is too close to a simple denial of the coherentist’s central thesis. The coherentist cannot, however, escape the force of the intuition merely by taking up a thesis that directly contradicts it. Nearly everyone – even the coherentist, in all likelihood – shares the sense that there is something wrong with simple, clear cases of circular reasoning. The coherentist therefore has the burden to explain why an exercise in circular reasoning should become acceptable merely by dint of the size of the circle.

## 3. FOUNDATIONALISM

Foundationalists believe that some beliefs are non-inferentially justified, or “foundational”, and that all other justified beliefs depend for their justification on these foundational beliefs. Most foundationalists accept beliefs about simple necessary truths (such as that  $1 + 1 = 2$  or that nothing can be entirely blue and entirely red at the same time) as well as beliefs about one’s own present conscious mental states (such as the belief that one is presently in pain) as foundational.<sup>18</sup> Some also accept perceptual beliefs, such as the belief that there is a red object in front of one (when one sees such an object), as foundational.<sup>19</sup>

In assessing the merits of foundationalism, it is important to avoid saddling the view with needlessly strong assumptions. Foundational beliefs, by definition, are beliefs that have some degree of justification not dependent on reasons. Foundational beliefs need not be infallible, absolutely certain, or unrevisable; the degree of non-inferential justification they have may be relatively modest.<sup>20</sup> Foundational beliefs also are not precluded from being supported by reasons; the thesis that a belief does not *require* reasons in order to be to some degree justified does not entail that the belief *cannot have* reasons that further support it.

### 3.1 Arguments for Foundationalism

The most popular argument for foundationalism is the regress argument. It maintains

that foundationalism is the only plausible response to the regress problem discussed in section 1. This argument premises that we have justified beliefs, justification cannot be produced by circular reasoning, and we do not have an infinite series of reasons for our justified beliefs; it follows that, when we have a justified belief, the chain of reasons for it terminates. Assuming that an unjustified belief cannot serve as a reason for another belief, the chain of reasons must terminate in a non-inferentially justified belief.<sup>21</sup>

A second argument for foundationalism consists in an appeal to apparent examples of non-inferentially justified beliefs. Suppose I go to the doctor, complaining that I have developed arthritis. The doctor says, "Why do you think you have arthritis?" This is a sensible question. Suppose I respond, "Because I have a pain in my wrist." Now suppose the doctor asks, "Why do you think you're in pain?" This question is just bizarre. When one has a pain, one is acquainted with it; one does not seek *reasons for thinking* that one is in pain. The same might be said of other present, conscious mental states, and the same might perhaps be said of simple and obvious necessary truths.

Critics have tried to identify ways in which even these beliefs are based on reasons. The most common suggestion seems to be that an introspective belief that one is in pain rests on the belief that one's introspection is reliable, or that one is good at recognizing pains.<sup>22</sup> An initial problem with this suggestion is that the alleged reason does not by itself provide any support for the claim that it is alleged to be a reason for – that is, the following is not at all a cogent inference:

- P1: My introspection is reliable. Therefore,  
C: I am in pain.

To have a cogent inference, one must add a further premise, as in the following:

- P1: My introspection is reliable.  
P2: I have an introspective belief that I am in pain. Therefore,  
C: I am in pain.

But this sort of inference does not provide a plausible model of how a typical introspective belief such as C is justified. For C to be justified in this way, P2 would have to be justified. But P2 is another introspective belief. If we apply the same model to explain how P2 is justified, we must say that it is justified by the following inference:

- P1: My introspection is reliable.  
P3: I have an introspective belief that I have an introspective belief that I am in pain. Therefore,  
P2: I have an introspective belief that I am in pain.

But P3 is another introspective belief, so we can ask how it is justified. This leads us down the path of an infinite regress. And this regress is particularly vicious, because at each stage the second premise in the argument becomes more complex. Very shortly it will become too complex for a normal human being to grasp, after which point the premise will be both false and unjustified. Better, then, to have rested at the start with the non-inferential belief that one is in pain.

### 3.2 The Arbitrariness Objection

Some object that putatively foundational beliefs are "arbitrary".<sup>23</sup> It is unclear what this objection amounts to. If "arbitrary" just means "unjustified", then the objection begs the question – it simply asserts that the beliefs foundationalists take to be justified are not justified. If an "arbitrary belief" just means a belief for which one has no reason, then the objection again begs the question – it simply reminds us that foundationalists endorse the notion of non-inferential justification, and presupposes that this is problematic.

A more sophisticated worry is that foundationalism conflicts with the principle of the supervenience of epistemic justification. This is the principle that, whenever a belief is justified, there is some other feature of it (possibly relational and/or conjunctive) in virtue of which it is justified. For simplicity, let us suppose that there is a single feature, F, that makes all non-inferentially justified beliefs justified, a feature that distinguishes these

beliefs from merely arbitrary beliefs. One may ask whether a belief's having this feature renders it likely to be true. If not, then it is obscure how this feature can render beliefs justified. If so, then it appears that one *does* have a reason for any given foundational belief: namely, that the belief has F. Therefore, the allegedly foundational belief is not foundational after all.<sup>24</sup>

This objection appears to rest on a misunderstanding discussed earlier, that of attributing to foundationalists the thesis that non-inferentially justified beliefs are *incapable* of being supported by reasons. Rather, non-inferentially justified beliefs do not *need* reasons in order to have some degree of justification; this does not preclude our finding reasons for them.

Nevertheless, one may feel that, although the existence of a reason supporting each allegedly foundational proposition does not show that these propositions are not foundational, it defeats much of the *point* of foundationalism. Foundational propositions are posited largely to avoid the infinite regress of reasons. If each putatively foundational proposition has reasons supporting it, then it may seem that we get an infinite regress anyway, so foundationalism hasn't bought us anything.<sup>25</sup>

There are two replies to this last worry. One is that what foundationalism buys us, in part, is an account of why individuals who are less than ideally self-aware and philosophically sophisticated nevertheless have justified beliefs. On the foundationalist view, one may have a non-inferentially justified belief, provided one's belief has feature F. It is not necessary that one also *believe* (even dispositionally) that one's belief has F, nor that one believe that beliefs with F are likely to be true. This is a desirable feature of the theory, since we usually do not in fact reflect on our beliefs and their epistemic status, and there is no one who is *always* aware of his own beliefs. Even if, on some occasion, I both believe that P and believe that my belief that P has feature F, it is unlikely that I will also believe that my belief that my belief that P has feature F has feature F, and so on *ad infinitum*.

Second, even when we restrict our attention to ideally self-aware and philosophically sophisticated thinkers, the availability of a reason for each putatively foundational belief does not necessarily render the positing of non-inferential justification otiose. Suppose that S believes that P; that his belief has feature F, where this is a feature that does not require the belief to be based on reasons; and that beliefs with F are thereby likely to be true. If S is also self-aware and philosophically sophisticated, then S has available a reason for believing P: namely, that the belief that P has F. Now, there are at least two ways in which S's belief that P might be justified: (a) it might be non-inferentially justified merely in virtue of having F, or (b) it might be inferentially justified in virtue of S's having the mentioned reason for believing P. The anti-foundationalist claims that there is no good reason to posit justification (a), since justification (b) is available in any case and suffices to account for the belief's justification. It might therefore be thought that Occam's Razor dictates rejection of the foundationalist posit. However, notice that the real difference between (a) and (b) is the difference between a belief's merely having a certain feature, and the subject's recognizing that feature. So the dispute between the foundationalist and the anti-foundationalist here is over whether a belief's having F renders it justified, or whether one must also add the subject's recognition of, or belief about, its having F. Put this way, it is not at all clear that the anti-foundationalist has the simpler theory.

Furthermore, there is no peculiarly problematic feature of foundationalism to be found here – an exactly parallel dispute could arise about any theory of justification. Suppose a theory of justification claims that some belief is justified by virtue of its having feature J. This may be any feature whatsoever, including features that require the belief to belong to a coherent system, that require the belief to be supported by an infinite series of reasons, and so on. One could always propose that, instead of being justified by virtue of having J, the belief is justified (for ideally self-conscious, philosophically sophisticated thinkers) by virtue of one's *recognition* that it has J. It is



unclear why this move should be any less plausible when applied to other theories of justification than it is when applied to foundationalism. But presumably we should not conclude that no theory of justification is well motivated. It therefore does not seem that the objection shows foundationalism to be ill-motivated.

### 3.3 The Bootstrapping Problem

Suppose I have a method of belief-formation that produces foundational beliefs. Suppose for the sake of argument that this method is ordinary observation – that is, when I believe what perceptually appears to me to be the case, my belief is non-inferentially justified. By hypothesis, since perceptual beliefs are non-inferentially justified, I need not have any beliefs about the reliability of my perceptual faculties in order to have justified perceptual beliefs. Now suppose it occurs to me to wonder whether sense perception is reliable. I decide to check: I go out and form some perceptual beliefs; each time, I check (perceptually) to see whether the belief is true, and I find that it is. For instance, upon seeing a red fire hydrant, I think to myself:

- P1: That is a red fire hydrant.
- P2: I have a perceptual belief that that is a red fire hydrant.
- C1: Therefore, I have a true perceptual belief in this case.

P1 is foundational (known by observation), and P2 is justified by introspection. I repeat this sort of inference many times, collecting many examples of cases in which my perceptual belief was true. I infer inductively that my perceptual beliefs are reliable.

This is known as a bootstrapping inference – I have used a faculty or belief-forming method in the effort to verify its own reliability. Intuitively, it seems that there is something wrong with this way of proceeding; it seems that the bootstrapping inference involves an undesirable form of circularity. Yet, if foundationalism is true, it is unclear why the proposed reasoning should not be legitimate. Both P1 and P2 are justified; they entail C1;

C1 seems to constitute inductive evidence for the reliability of sense perception; and the inference is non-circular since, as we have supposed, my belief in the red fire hydrant does not depend upon any prior beliefs about the reliability of sense perception – indeed, it depends on no prior beliefs for its justification at all, which makes it difficult to see how it can fail to be a legitimate starting point for reasoning.

This example motivates the thought that, to account for the intuitive circularity of the bootstrapping argument for the reliability of perception, we should suppose that perceptual beliefs do after all depend for their justification on the belief that perception is reliable; therefore, perceptual beliefs are not foundational. Since the same sort of argument could be made with respect to any cognitive faculty or belief-forming method, we should conclude that no beliefs are foundational.<sup>26</sup>

This is a very interesting problem to which I do not know the answer – I do not know whether bootstrapping arguments are legitimate, or why they are illegitimate if they are. But notice that the problem does not only confront foundationalists. Though it is meant to apply to non-inferentially justified beliefs, the problem can be generalized so as to apply also to inferentially justified beliefs. For example, assume that *modus ponens* is a cogent form of inference. One can construct a bootstrapping argument for the conditional reliability of *modus ponens* (i.e., for the conclusion that inferences following *modus ponens* tend to produce true conclusions given true premises). One can simply form a number of beliefs using *modus ponens*, observe introspectively that one used *modus ponens* to form them, and conclude that one has a number of instances of *modus ponens* leading to true beliefs, hence, inductive evidence that *modus ponens* is conditionally reliable. Intuitively, this is just as problematic as using perception to verify the reliability of perception, and the bootstrapping inference here is at least not explicitly circular – nowhere does one rely on the premise “*modus ponens* is conditionally reliable” (inferences of the form *modus ponens* do not generally require “*modus ponens* is conditionally reliable” as premises).

One can generalize the problem to apply to nearly any theory of justification. Suppose an epistemological theory holds that a belief's having feature J is sufficient for its being justified. Provided that J is sometimes detectable, one could construct a bootstrapping inference for the claim that beliefs with J are generally likely to be true: simply form a number of beliefs with J, realize that they have J, and conclude that their truth constitutes inductive evidence for the generalization that beliefs with J tend to be true.<sup>27</sup> For example, suppose a coherentist holds that a belief's cohering with the rest of one's belief system is sufficient for it to be justified. He might then think to himself:

P3: Dogs have four legs.

P4: The belief that dogs have four legs coheres with the rest of my belief system.

C2: So coherence correlates with truth in this case.

where P3 is justified solely because it coheres with the rest of his belief system. After rehearsing a number of similar cases, the coherentist concludes that he has inductive evidence that coherence is a reliable indicator of truth. Intuitively, this bootstrapping inference is problematic in precisely the same way as the bootstrapping inference for the reliability of sense perception. And the coherentist will have just as much difficulty explaining what is wrong with the former inference as the foundationalist has in explaining what is wrong with the latter.<sup>28</sup> The coherentist bootstrapping inference is at least not explicitly circular – by hypothesis, P3 is justified solely by virtue of the fact that it coheres with the rest of one's belief system; it is not inferentially justified on the basis of the belief that coherence is a reliable indicator of truth. It thus appears that the bootstrapping problem cannot motivate a preference for coherentism over foundationalism.

### 3.4 Which Beliefs Are Foundational?

Among foundationalists, there are several views as to which beliefs have non-inferential

justification. First, there is the traditional view that a belief is non-inferentially justified if and only if it is infallible, either in the sense that one's holding the belief entails that it is true, or in the sense that one's holding the belief for the sort of reason one holds it entails that it is true.<sup>29</sup> For instance, my belief that I exist is infallible in the former sense; my belief that I am in pain, based on my introspective awareness of pain, is infallible in the latter sense. Second, there is the view that a belief is non-inferentially justified if the believer is *acquainted with*, or directly aware of, a fact that makes it true. Typically, facts about one's own mental states and perhaps about simple abstract objects are held to be among the possible objects of acquaintance.<sup>30</sup> Third, one may hold that one has non-inferential justification for a belief if the belief is the product of a reliable belief-forming mechanism that does not require the input of any other beliefs.<sup>31</sup> For example, provided that one's senses are in fact reliable, one can acquire non-inferentially justified beliefs about the external world by simply believing what perceptually appears to one to be the case. Fourth, there is a view according to which, whenever one believes something, the mere fact that one believes it provides at least some degree of *prima facie* justification for it.<sup>32</sup> Finally, some hold the view that whenever something *appears* to be the case (where this is distinct from one's believing it), not as a result of reasoning, one has at least some degree of *prima facie* justification for believing it.<sup>33</sup>

This last view, which we may dub "Phenomenal Conservatism", has important advantages over the others. The most interesting of these advantages is that it may be the only theory that allows most of our actual beliefs to be justified. The argument for this relies on two premises. First, that for a belief to be justified, the factor in virtue of which the believer has justification for the relevant proposition must itself be *the reason why* the believer accepts that proposition.<sup>34</sup> For instance, suppose that Jon has a great deal of scientific evidence for the theory of evolution; however, none of this evidence moves him. The actual reason why he believes in evolution is due to a recent reading of tea leaves.

In that case, though Jon *has justification* for the theory of evolution, his actual belief is unjustified.

Second, the reason why we believe all or most of the things that we in fact believe is that they seem to us to be the case. Sometimes, indeed, we believe a thing because we have an argument for it; but an argument will only move us to endorse a conclusion to the extent that the argument seems to us to be sound, thereby making it seem to us that the conclusion is true. Likewise, sometimes we have evidence that appearances are deceiving, as when we become aware of a sensory illusion; but even in such a case we accept that evidence only because it seems to us to be true – in short, because we accept certain other appearances that we find more compelling.

If these two premises are correct, it follows that all or most of our beliefs are unjustified, unless something's seeming to us to be the case gives us some justification for believing it.<sup>35</sup> Once we accept the principle of Phenomenal Conservatism, we can account for all the central intuitive examples of justified beliefs, such as perceptual beliefs, introspective beliefs, memory beliefs, and intuitive beliefs.

#### 4. CONCLUSION

Foundationalism is the best theory of the structure of epistemic justification. It is supported by intuitive examples of non-inferentially justified beliefs, such as one's belief that one is in pain when one is, and the belief that  $1 + 1 = 2$ . Some object that beliefs not supported by reasons must be "arbitrary", but this objection appears simply to beg the question. Some critics point out that, provided that foundational beliefs have some feature, F, that distinguishes them from merely arbitrary beliefs, and provided that this feature renders beliefs that have it likely to be true, a sufficiently self-aware and sophisticated thinker always has a reason supporting any foundational belief – namely, the fact that the belief has F. This is not incompatible with the belief's being foundational, since foundational beliefs are defined as beliefs that *do not need* reasons, rather than as beliefs that

*cannot have* reasons supporting them. Nor does this render the posited non-inferential justification otiose; the thesis that F confers justification on a belief is no less simple than the thesis that an individual's awareness of F is required for justification, and the latter thesis is analogous to a claim that could be posed with equal plausibility as an alternative to any theory of justification. In any case, the non-inferential justification posited by the foundationalist is required to explain why less than ideally self-aware and sophisticated thinkers have justified beliefs.

Critics have charged foundationalism with implausibly licensing bootstrapping inferences, in which one employs a belief-forming method to verify the method's own reliability. It is difficult to say what, if anything, is wrong with bootstrapping inferences; however, the problem is not unique to foundationalism, as bootstrapping inferences can be constructed equally for nearly any alleged source of justified belief, including types of inferences.

The coherence theory, by contrast, faces two major difficulties. The first is that the probabilistic argument typically advanced in support of it fails to show how coherence alone can provide justification for any belief. This is because the truth of a collection of claims explains their coherence only on the assumption that each claim would have been more likely to be made if it were true than if it were false; hence, each claim must have some individual credibility. Furthermore, we must have prior knowledge of our own mental states (particularly our beliefs) and of their coherence, before coherence can be a source of justification.

The second problem is that the coherence theory appears simply to endorse a complicated form of circular reasoning, and there is no clear reason why large, complex exercises in circular reasoning should be any more acceptable than simple and obvious ones. The approach of holistic coherentism, which claims that a whole system of beliefs is justified by virtue of its coherence, makes little advance on this problem, since the coherence of a system seems to consist mainly in the individual beliefs' standing in the very support relations



to each other that circular reasoning relies on; therefore, if circular reasoning does not provide justification for the individual beliefs, it is implausible that coherence provides justification for the system.<sup>36</sup>

## NOTES

- 1 Sometimes the problem is phrased in terms of *knowledge*, but I shall continue to focus on justified belief.
- 2 If a belief has more than one reason for it, more than one of these structures can be instantiated. The series of reasons behind P might, for example, have a circular branch, an infinite branch, and a terminating branch. Or different beliefs may have different structures of reasons. For simplicity, I shall neglect these possibilities.
- 3 Chisholm, 1977; Alston, 1989, ch. 1; Fumerton, 2005; Huemer, 2001, ch. 5.
- 4 BonJour, 1985; Davidson, 1990; Lehrer, 1974, chs 7–8; Elgin, 2005; Kvanvig, 2003.
- 5 Klein, 1999.
- 6 Sextus Empiricus, 1964, pp. 72–5; Unger, 1975; Oakley, 1976.
- 7 BonJour, 1985, pp. 95–9; Lehrer, 1974, pp. 163–5; Thagard, 2000, pp. 43, 53. Each of these three factors may be thought of as contributing to overall coherence, or one or more of them (most likely the consistency condition) may be thought of as *necessary* for coherence.
- 8 BonJour, 1985, pp. 147–8. The example is from Elgin, 2005, p. 157.
- 9 Huemer (1997) contains a proof of this result for a simple model; Olsson (2005, pp. 68–9, 218) provides a more general proof. See also Olsson (2005, ch. 7) and Bovens and Hartmann (2003, ch. 1) for arguments that there can be no measure of coherence such that coherence is in general truth-conducive *ceteris paribus*. However, all these theorems depend on the assumption that the witnesses' reports are probabilistically independent, both given the truth of their story and given the falsity of their story. I now believe that a coherentist can and should reject this assumption (see my forthcoming b).
- 10 Van Cleve (2005, p. 173) presses this argument. BonJour acknowledges that the coherentist must presuppose that we can correctly identify our own beliefs (1985, pp. 146–7) and that the coherentist cannot justify this assumption (1985, pp. 105–6).
- 11 Lehrer, 1974, pp. 187–8.
- 12 See Kvanvig, 1995. BonJour (1985, pp. 117–24, 139–40) provides a different response.
- 13 BonJour, 1985, pp. 141–4; Kvanvig, 1995.
- 14 Foley, 1979. Lehrer (1974, pp. 192–7) responds to the lottery problem but remains vulnerable (pp. 202–3) to the example following in the text.
- 15 Lycan (1996, p. 10) responds to this objection by denying that consistency is necessary for coherence.
- 16 Fumerton, 1995, pp. 145–7.
- 17 Elgin, 2005, p. 156; Kvanvig, 2003, sect. 1; Thagard, 2000, pp. 76–7.
- 18 Chisholm, 1977, pp. 20–3, 27–30, 40–6; Audi, 1993, pp. 307, 310. See also note 19 below.
- 19 Reid, 1983, pp. 272–3; Huemer 2000; 2001, ch. 5; Audi, 1993, p. 308; Pollock and Cruz, 1999, pp. 84–8, 191–239.
- 20 Alston, 1989, ch. 2; Audi, 1983. Thagard (2000, pp. 4, 71–2) mistakenly takes foundationalism to be committed to indubitable, unrevisable foundations.
- 21 Alston, 1989, pp. 26–7. The argument traces back to Aristotle (*Posterior Analytics*, 72b17–25; *Metaphysics*, 1006a5–12).
- 22 Oakley, 1976, pp. 222–4; Cohen 2002.
- 23 Klein, 1999, p. 297. I address this objection in detail in my 2003.
- 24 BonJour (1985, pp. 30–2) poses a more sophisticated version of the objection.
- 25 This is my understanding of Klein's (2004) argument.
- 26 Cohen, 2002.
- 27 The only exceptions would be cases in which one cannot independently know that a belief has J, or (if this is possible) in which a belief's having J entails that it is partly based on the belief that J is a reliable indicator of truth.
- 28 Cohen (2002, pp. 323–5) discusses this problem.
- 29 This view is inspired by Descartes (1984, pp. 16–17). See Lehrer (1974, pp. 78–100) and Fumerton (2001, pp. 9–12) for refutation of the view.
- 30 Russell, 1997, ch. 5; Fumerton, 2001, pp. 12–18; BonJour, 2001, pp. 24–31.
- 31 Goldman, 1992, pp. 117–18; Sosa, 1991, pp. 189–90.
- 32 Foley (1983) criticizes this view convincingly.
- 33 Lycan, 1996, p. 5; Huemer, 2001, ch. 5.
- 34 I leave the notion of "the reason why" one holds a belief unanalyzed, except to note (i) that it presumably applies to a factor that is at least the most salient causal factor in producing the agent's belief, and (ii) that the notion does not

- imply that the belief in question is inferential (the reason why one believes that P need not be that one believes something else that supports P).
- 35 I elaborate this argument in my (forthcoming a).
- 36 I wish to thank Jonathan Kvanvig for his helpful comments on this paper.

## BIBLIOGRAPHY

- Alston, W.: *Epistemic Justification* (Ithaca, NY: Cornell University Press, 1989).
- Aristotle: *The Basic Works of Aristotle*, ed. R. McKeon (New York: Random House, 1944).
- Audi, R.: "Foundationalism, Epistemic Dependence, and Defeasibility," *Synthese* 55 (1983), 119–39.
- Audi, R.: *The Structure of Justification* (Cambridge: Cambridge University Press, 1993).
- BonJour, L.: *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1985).
- BonJour, L.: "Toward a Defense of Empirical Foundationalism," in *Resurrecting Old-Fashioned Foundationalism*, ed. M. DePaul (Lanham, MD: Rowman and Littlefield, 2001), 21–38.
- Bovens, L. and Hartmann, S.: *Bayesian Epistemology* (Oxford: Clarendon Press, 2003).
- Chisholm, R.: *Theory of Knowledge*, 2nd edn (Englewood Cliffs, NJ: Prentice-Hall, 1977).
- Cohen, S.: "Basic Knowledge and the Problem of Easy Knowledge," *Philosophy and Phenomenological Research* 65 (2002), 309–29.
- Davidson, D.: "A Coherence Theory of Truth and Knowledge," in *Reading Rorty*, ed. A. Malachowski (Cambridge, MA: Blackwell, 1990), 120–38.
- Descartes, R.: *Meditations on First Philosophy*, in *The Philosophical Writings of Descartes*, ed. J. Cottingham, R. Stoothoff and D. Murdoch, vol. 2 (Cambridge: Cambridge University Press, 1984).
- Elgin, C.: "Non-foundationalist Epistemology: Holism, Coherence, and Tenability," in *Contemporary Debates in Epistemology*, ed. M. Steup and E. Sosa (Malden, MA: Blackwell, 2005), 156–67.
- Foley, R.: "Justified Inconsistent Beliefs," *American Philosophical Quarterly* 16 (1979), 247–57.
- Foley, R.: "Epistemic Conservatism," *Philosophical Studies* 43 (1983), 165–82.
- Fumerton, R.: *Metaepistemology and Skepticism* (Lanham, MD: Rowman and Littlefield, 1995).
- Fumerton, R.: "Classical Foundationalism," in *Resurrecting Old-fashioned Foundationalism*, ed. M. DePaul (Lanham, MD: Rowman and Littlefield, 2001), 3–20.
- Fumerton, R.: "Foundationalist Theories of Epistemic Justification," in *Stanford Encyclopedia of Philosophy*, ed. E. Zalta, Summer 2005 edition: <http://plato.stanford.edu/archives/sum2005/entries/justep-foundational/>.
- Goldman, A.: *Liasons: Philosophy Meets the Cognitive and Social Sciences* (Cambridge, MA: MIT Press, 1992).
- Huemer, M.: "Probability and Coherence Justification," *Southern Journal of Philosophy* 35 (1997), 463–72.
- Huemer, M.: "Direct Realism and the Brain-in-a-vat Argument," *Philosophy and Phenomenological Research* 61 (2000), 397–413.
- Huemer, M.: *Skepticism and the Veil of Perception* (Lanham, MD: Rowman and Littlefield, 2001).
- Huemer, M.: "Arbitrary Foundations?," *Philosophical Forum* 34 (2003), 141–52.
- Huemer, M.: "Compassionate Phenomenal Conservatism," *Philosophy and Phenomenological Research* (forthcoming [a]).
- Huemer, M.: "Weak Bayesian Coherentism," *Synthese* (forthcoming [b]).
- Klein, P.: "Human Knowledge and the Infinite Regress of Reasons," *Philosophical Perspectives* 13 (1999), 297–325.
- Klein, P.: "What IS Wrong with Foundationalism Is That It Cannot Solve the Epistemic Regress Problem," *Philosophy and Phenomenological Research* 68 (2004), 166–71.
- Kvanvig, J.: "Coherentists' Distractions," *Philosophical Topics* 23 (1995), 257–75.
- Kvanvig, J.: "Coherentist Theories of Epistemic Justification," in *Stanford Encyclopedia of Philosophy*, ed. E. Zalta, Winter 2003 edition: <http://plato.stanford.edu/archives/win2003/entries/justep-coherence/>.
- Lehrer, K.: *Knowledge* (Oxford: Clarendon Press, 1974).
- Lycan, W.: "Plantinga and Coherentisms," in *Warrant in Contemporary Epistemology*, ed. J. Kvanvig (Lanham, MD: Rowman and Littlefield, 1996), 3–23.
- Oakley, I. T.: "An Argument for Scepticism Concerning Justified Beliefs," *American Philosophical Quarterly* 13 (1976), 221–8.
- Olsson, E.: *Against Coherence* (Oxford: Clarendon Press, 2005).
- Pollock, J. and Joseph, C.: *Contemporary Theories of Knowledge*, 2nd edn (Lanham, MD: Rowman and Littlefield, 1999).
- Reid, T.: *Inquiry and Essays*, ed. R. Beanblossom and K. Lehrer (Indianapolis, IN: Hackett, 1983).
- Russell, B.: *The Problems of Philosophy* (New York: Oxford University Press, 1997).

- Sextus Empiricus: *Scepticism, Man, and God*, ed. P. Hallie and trans. S. Etheridge (Middletown, CT: Wesleyan University Press, 1964).
- Sosa, E.: "The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge," in *Knowledge in Perspective* (Cambridge: Cambridge University Press, 1991), 165–91.
- Thagard, P.: *Coherence in Thought and Action* (Cambridge, MA: MIT Press, 2000).
- Unger, P.: *Ignorance: A Case for Scepticism* (Oxford: Clarendon Press, 1975).
- Van Cleve, J.: "Why Coherence Is Not Enough: A Defense of Moderate Foundationalism," in *Contemporary Debates in Epistemology*, ed. M. Steup and E. Sosa (Malden, MA: Blackwell, 2005), 168–80.

MICHAEL HUEMER

## Recent Work on the Internalism–Externalism Controversy

Though the internalism–externalism issue has remained a main focus of epistemological discussion during the past dozen years or so, it is far from clear that any real progress has been made towards a resolution. While some new arguments have been adduced (most extensively by Alvin Goldman: see below), both the definition of the main positions and the ultimate significance of the dispute seem now, if anything, less clear than they previously appeared to be. As happens with rather alarming frequency in philosophy, an issue that seemed initially clear-cut and well defined seems sometimes to be simply dissolving under closer scrutiny.

In the limited space available here, I shall focus on three central topics, at the cost of saying nothing about a good deal of other interesting work.<sup>1</sup> I shall begin (§1) with a discussion of the main distinction between internalism and externalism, focusing on the two most prominent accounts of internalism. Next (§2), I shall discuss some of the relevant arguments, focusing mainly on Goldman's recent claim to have demolished internalism. I shall then turn (§3) to a discussion of the larger significance of the issue, trying to determine to what extent the two main views are genuine rivals, between which a choice must be made.

### §1. WHAT IS INTERNALISM?

Here I shall assume that *internalism* is the view that is mainly in need of clarification, with externalism being best-defined as simply the denial of internalism. The first question that needs to be addressed concerns the main application of this term: it seems clear that it applies to accounts of some fundamental epistemic concept, but to which one primarily? Here the main apparent alternatives are the concept of knowledge itself, the concept of warrant, and the concept of epistemic justification.

There is no doubt that the label "internalist" (or the contrasting label "externalist") can be applied to accounts of knowledge; but it seems clear that, if they are to mark an interesting distinction here, this must be in virtue of their bearing on some narrower element or ingredient of knowledge, presumably either warrant or justification – the point being that the other plausible ingredients of knowledge, namely belief, truth, and some anti-Gettier condition, seem each to fall clearly on either the internalist (belief) or externalist (the other two) sides of the divide. Michael Bergmann has suggested that the idea of internalism should be taken to pertain primarily to *warrant*, understood (in the technical sense introduced by Plantinga<sup>2</sup>) as whatever has to be added to true belief to yield knowledge;<sup>3</sup> but in addition to thus including the pretty obviously external anti-Gettier condition, this notion of warrant is to my mind a highly artificial one, corresponding to nothing in ordinary thought, thereby making some of the main intuitive arguments difficult or impossible to assess. For these reasons, I shall initially follow most of the literature in taking the idea of internalism and the resulting distinction to apply primarily to accounts of *epistemic justification*, assumed for the moment to be one of the requirements for knowledge (along with belief, truth, and an anti-Gettier condition) – though I shall later (in §3) discuss some reasons for doubting how satisfactory this conception of the issue really is.

What, then, does it mean to say of an account of epistemic justification that it is

internalist in character? What exactly is supposed to be *internal* to what? In a well-regarded introductory textbook, Matthias Steup offers the following account:

What makes an account of justification internalist is that it imposes a certain condition on those factors that determine whether a belief is justified. . . . The condition requires [such factors] to be *internal to the subject's mind* or, to put it differently, *accessible on reflection*.<sup>4</sup>

Here we have what are clearly the two main alternatives in recent discussions for an account of internalism,<sup>5</sup> alternatives that seem in fact to be clearly distinct.<sup>6</sup> According to the first, for which Conee and Feldman (perhaps its main proponents) have adopted the label “mentalism”, an account of justification is internalist just in case it appeals only to things that are “internal to the person's mental life”, to “occurrent and dispositional mental states, events, and conditions”.<sup>7</sup> According to the second, which is most standardly referred to as “access internalism”, an internalist account is one according to which the justificatory “factors” – all of them, on the most standard versions – must be “accessible” to the person, in a sense that is in need of at least some further clarification. Here it is worth noting that the basis for the mentalist conception of internalism appears, somewhat surprisingly, to be primarily metaphysical in character; while, in contrast, the basis for the access internalist's conception is recognizably epistemological.<sup>8</sup>

Which of these two views best captures the central intuitive idea of internalism? In thinking about this question, it will be helpful to have a sample *externalist* view of epistemic justification available to serve as a dialectical foil, for it is surely one good test of an account of internalism that it can give a perspicuous account of what is supposed to be lacking in externalist views. The obvious choice for this purpose, and indeed the only specific externalist view that will be discussed here, is *reliabilism*: the view that a belief is epistemically justified if it results from a cognitive process that is (sufficiently) reliable in producing true beliefs.<sup>9</sup> I shall argue that

mentalism does an unsatisfactory job of capturing the intended contrast with such an externalist view, and that for this reason access internalism is to be preferred.<sup>10</sup>

There are two sides to this reason. The first is that there are possible examples of mental states and processes that do not seem to yield the sort of internalist justification that would contrast in the right way with reliabilism. Consider, for example, the fairly widely accepted idea of *unconscious* mental states, states which could presumably be combined into unconscious mental processes. There is no apparent reason why such unconscious states or processes could not in some way make likely the truth of a particular belief, nor why they might not causally contribute to such a belief being held. This would be a kind of reliable cognitive process, one that differs from those (such as the perceptual processes more commonly cited by the reliabilist) only in being entirely “internal” to the mind. But why should this metaphysical fact have any epistemological significance, given that in both cases the reason for the belief that is available in the situation is not one of which the person in question will normally have any awareness? This is a standard internalist complaint about reliabilism, but it seems to apply just as well to the “internalist” (according to mentalism) justification just described. A different sort of example, about which the same point can be made, is a justification that appeals to some property of a person's mental state – perhaps the coherence of his entire system of beliefs – that is too complex and multi-faceted for the person to ever be capable of reflectively apprehending that it obtains. Again, while a belief that results from the presence of such a property might be reliably caused, it is hard to see why a justification that appeals to this sort of entirely mental situation should count as internalist in a way that ordinary reliabilist justifications do not: once again there is a metaphysical difference, but one that seems to have no very obvious epistemological significance.

The other side of the coin is the possibility of elements that are *external* to the mind but none the less capable of being directly apprehended in a way that allows them to play a

justificatory role, one that contrasts in what intuitively seems like just the right way with a typical reliabilist justification. Direct realist theories of perception hold that external material objects are capable of being directly apprehended in a way that allows their perceived presence to justify the corresponding beliefs. And many rationalist accounts of *a priori* justification hold that entities such as universals, mathematical entities, and logical connections can be directly or intuitively apprehended and thereby play a role in the justification of *a priori* claims. In both of these cases there are also mental states involved, but the supposed justification seems essentially to involve the relation of such states to these non-mental entities, so that the non-mental entities play an indispensable role.<sup>11</sup> Of course, both of these views have been seriously criticized, and my own view is that the first one is mistaken. But surely neither of them is so obviously wrong as to be reasonably dismissed out of hand. And the point is then that the justification in each case, though it appeals in part to *non-mental* entities, seems to contrast with reliabilist justification in exactly the right, epistemologically significant way: the basis for the justification is one that the person is able (i) to reflectively grasp and (ii) to critically evaluate, and it is these features, I submit, that make it plausible to classify it as internalist in character.<sup>12</sup>

Thus, I suggest, the right account of internalism is after all access internalism: what mainly matters is not the metaphysical status of a justificatory element or factor, but instead its *availability* to the person as a *reason* (or the basis of a reason) for accepting a particular belief, a reason that he is thereby in a position to critically evaluate – for it is the absence of these features that make a typical reliabilist justification seem unsatisfactory, according to recognizably internalist intuitions. This is to say that the relevant sort of internality is *being internal to a person's first-person cognitive perspective*, for which being internal to his mind in a metaphysical sense is clearly not sufficient and arguably not necessary.

This way of understanding internalism has two further corollaries. One, which we have been in effect assuming all along, but which

is sometimes denied,<sup>13</sup> is that only a view according to which *all* of the elements required to yield a cogent reason for the belief in question are appropriately accessible can satisfy the fundamental internalist intuition. Only if this is so does the person genuinely *have* a reason, as opposed to merely part of a reason, for the belief, one that could seem reflectively to justify the belief and that is available for critical assessment. And an externalist view will then be one that allows some element that is essential to the cogency of a justificatory reason to be outside of, external to, the person's cognitive perspective. It is the fact that this is clearly possible for reliabilism that makes it an externalist view.

The other corollary, which has been even more widely challenged, is that the “accessibility” of the justificatory factors or elements must be understood in a very strong way. Part of the point has already been made in the insistence of Steup and others on *reflective* accessibility: that I have access in a sense to a reference work or to a perceptible situation that could yield a reason does not make those reasons internally accessible in the right way, since until I do whatever is required to avail myself of them I do not actually *have* the reason in question. But equally (and here is another problem with mentalism) the fact that a basis for a reason is lurking somehow in my perceptual experience or system of beliefs or other mental contents in an entirely unnoticed way also does not seem to give me an appropriately internalist reason for the corresponding belief (again the unrecognized coherence of my entire belief system is a good example): until I do whatever is needed to isolate the ingredients of such a reason and put them together in an appropriate way, I again do not actually *have* the reason in question, and obviously am also not in a position to evaluate it critically. I do not want to insist here that all of this must be done in a fully overt, explicit way, though from a strictly epistemological standpoint this is obviously optimum. But some sort of tacit or implicit awareness of what is involved seems required if I am to have an actual reason rather than a merely potential one. And a merely potential reason does not differ from a reliabilist justification



in the right way: any individual reliabilist justification is one that the person at least in principle might become explicitly aware of, and merely the fact that such an awareness is easier to arrive at in one case than the other does not seem in itself to mark an epistemological difference in the justification of the belief in the situation where such an awareness has not yet been attained.

## §2. RECENT ARGUMENTS

Turning to arguments for and against this conception of internalism and the correlative conception of externalism, I want first to take brief note of some arguments that have become standard on each side (focusing again entirely on reliabilist versions of externalism), but about which little needs to be said in a review of recent developments. On the internalist side, we have: (1) the appeal to Cartesian demon cases and the like to show that reliability is not *necessary* for justification; (2) the appeal to clairvoyance examples and the like to show that reliability is not *sufficient* for justification; and (3) the objection to reliabilism that derives from the problem of choosing for a specific cognitive process the general specification relevant to determining its reliability (“the generality problem”). On the externalist side, we have: (1) the argument that various unsophisticated epistemic subjects have justified beliefs and knowledge, even though they do not satisfy the requirements of any plausible internalist account; and (2) the argument that internalism inevitably leads to skepticism.<sup>14</sup> Apart from Goldman’s more specific arguments, considered below, some of which amount to more specific elaborations of externalist argument (2), there has been relatively little that is new in the ongoing discussion of these arguments in the period here under review. One point worth making, however, is that there continues to be nothing even approximating a solution to the generality problem, strongly suggesting to my mind that there simply is no solution to be had – and hence that the main externalist view cannot in the end even be clearly formulated.<sup>15</sup>

Here I shall focus almost entirely on what is surely the most substantial addition to the argumentative landscape in this area in the relevant period, namely Alvin Goldman’s very extended and, in his view, thoroughly devastating critique of internalism in his paper “Internalism Exposed”.<sup>16</sup> In this paper, Goldman argues that both internalism and the main arguments in its behalf are, when carefully examined, “rife with problems”. He concludes: “I see no hope for internalism; it does not survive the glare of the spotlight” (p. 293).

In fact, there are many aspects of Goldman’s discussion that are seriously problematic. Though he begins with what seems to be an access internalist formulation of internalism, his discussion often seems to presuppose mentalism instead, with no apparent realization of the differences between these views. His conception of the basic rationale for internalism in terms of what he calls the “guidance-deontological conception of justification”, while it undeniably reflects things that some internalists have said, seems to me to fail to capture the central internalist intuition. And the specific versions of internalism that he formulates along the way do not correspond very well to any views that real internalists have ever held. All this seems to me to give a somewhat strawman-nish character to a good part of his discussion.<sup>17</sup> In the present discussion, however, I shall be primarily concerned with some of the more specific problems that Goldman raises for internalism. I do not think that any of these arguments supports Goldman’s conclusion that internalism is fundamentally untenable. But they do raise issues which an internalist view needs to address, issues which can in fact contribute in important ways to refining and clarifying the internalist position.

Consider first what Goldman calls “the problem of stored beliefs”:

... At any given time, the vast majority of one’s beliefs are stored in memory, rather than occurrent or active. ... Furthermore, for almost any one of these beliefs, one’s conscious state at the time includes nothing that justifies it. No perceptual experience, no conscious

memory event, and no premises consciously entertained at the selected moment will be justificationaly sufficient for such a belief. According to strong internalism, then, none of these beliefs is justified at that moment. (p. 278) (Here “strong internalism” is the view that only facts about the agent’s conscious states at a particular time can justify his beliefs at that time.)

In fact, the main problem Goldman appears to be concerned with here is not really limited to stored beliefs as such, but has to do rather with whether there can be an adequate basis consciously in mind *at a moment* for the justification of almost any belief one might choose, whether stored or not. But, while some internalists may perhaps at times have foolishly accepted the limitation to what is available at a moment (however long that is!), there is, I submit, nothing about the basic rationale of internalism that in any way requires such a quixotic view. What is directly available from within the first-person epistemic perspective does not cease to be available or somehow become external in character just because it has to be collected and reviewed and collated over time. To be sure, a justifying reason or argument whose elements have been assembled over time still has to be in some way grasped as a unified whole, using the aid of memory and perhaps written records. But there is no reason why the grasping in question has to be *momentary*. Thus, this first objection has no serious force against reasonable forms of internalism.<sup>18</sup>

The second objection I want to discuss is what Goldman calls “the problem of forgotten evidence”. This is concerned with cases where the person in question has simply forgotten the evidential basis upon which a belief was originally accepted, but still retains the belief itself. Thus, in his example, Sally has read an article about the health benefits of eating broccoli in the “Science” section of the *New York Times* (which we may assume to be a highly reliable source) and formed the corresponding belief. She still retains the belief, but has forgotten how she acquired it. Thus, she seems to have no properly internalist justification available, but, according to Goldman, her belief is still justified (p. 280).<sup>19</sup>

In fact, whether Sally has an internalist reason for her belief, and how strong that reason is, will depend on further details of the case, some of them fairly subtle. One question is whether Sally has good reason to think that she is generally careful about the sources on the basis of which she accepts beliefs – or perhaps just beliefs held with the degree of assurance with which she holds this one. Another question is whether Sally *believes* that she acquired the belief from a reliable source, even if she can’t remember which one, and whether she has reason to think that both her judgments of the reliability of sources and her memories thereof are themselves reliable. A third question is whether Sally can recall various sorts of details that amplify and reinforce the belief in question: details about the specific ways in which broccoli is conducive to good health and about how it leads to these effects, even if the memories of these details are also beliefs whose original justificatory basis she does not recall. Sally might well have good internalist reasons to think that beliefs for which she can remember details of these sorts are more likely to have been derived from a reliable source and perhaps also more likely to be accurately remembered. On the basis of some or all of these answers, Sally might have the resources for an internalist reason for her main belief, whose strength would obviously vary with the details. Admittedly, these justifications are unlikely to be as strong as the one she had at the time she read the original article, but this seems like exactly the right result and in no way implausible.

A third alleged objection arises from “the problem of the doxastic decision interval”. Goldman argues that the supposed varieties of internalism that allow appeal only to conscious mental states (whether only occurrent ones or also those stored in memory) must be expanded to allow access to logical or probabilistic relations, and proposes that this be done by allowing some of the “formal properties of mental states, that is, logical and mathematical properties of their contents” to count as part of the basis for internalist justifications or reasons, while insisting that these should be restricted to the ones that

“are knowable by the agent at the time of doxastic decision” by employing a range of “computational relations or algorithms” (pp. 282–3). And the issue is then how long a period of time should be allowed for such computations, an issue that is aggravated, according to Goldman, by the worry that the agent’s mental states might change during the allowed interval in such a way as to affect the justification of the proposition in question (pp. 283–4).

But this way of formulating the supposed issue seems to me misconceived in a number of ways. As we have already seen, there is nothing about the basic rationale for internalism that limits the relevant justificatory factors or elements to mental states. If logical properties or relations are directly accessible via *a priori* reasoning, then they can also play an internalist justificatory role. Moreover, though this is a more debatable issue, I see no reason for limiting the properties and relations in question to those that are *formal* in any interesting sense – or, still less, to those which are known via “computational procedures or algorithms”.<sup>20</sup>

More importantly, the idea of a fixed “decision interval”, within which the eligible logical or probabilistic properties must be determined, while still excluding the possibility of significant mental change, is artificial in the extreme. Obviously changes of various sorts in one’s mental states can affect issues of internalist justification, but this has, as far as I can see, no very essential connection to the issue of which *a priori* knowable properties and relations can be appealed to. The bearing of such mental changes on justification is simply an independent issue, and there is no reason that I can see why the two issues must be dealt with together in the way that Goldman suggests.<sup>21</sup>

A fourth, somewhat related problem is “the availability problem”. Though Goldman poses this issue in relation to the supposed set of computational operations whose results are eligible for inclusion in internalist justifications or reasons (p. 285), I think, for reasons already sufficiently indicated, that it is better posed as simply the question of which *a priori* discernible properties or relations

are thus eligible. What is supposed to create the problem is the large disparities that exist between people with respect to their capacity, as a result of both training and basic intellectual ability, to arrive at such *a priori* insights. And the question is then which such insights are internalistically admissible.

My response to this supposed problem is very simply that I am unable to see why there is any problem here at all. It seems obvious that the answer should be that a particular *a priori* insight can play a role in the internalist justifications or reasons of a given person just in case that person is able to grasp or apprehend it. Logical relations that are too complicated or subtle for me to grasp cannot contribute to *my* justification for my beliefs, but if you can grasp them they may perfectly well play a role in *your* justification. As Goldman says, this means that “two people in precisely the same evidential state (in terms of perceptual situation, background beliefs, and so on) might have different epistemic entitlements” (p. 286), that is, might differ in which beliefs are justified for them. But it is hard to see why anyone should find this at all surprising or objectionable.

The fifth and last of the issues that I want to consider here has to do with the accessibility to ordinary or naïve epistemic agents of the epistemic principles that underlie putative internalist reasons. Goldman argues, though he does not put the point in quite this way, that such principles should be viewed as essential parts of the reasons in question, so that they would have to be accessible, presumably on an *a priori* basis, from the first-person epistemic perspective if such reasons are to be internalistically acceptable. I am inclined to agree. And his further argument on this basis is that most or all ordinary epistemic agents and indeed at least most epistemologists fail to be capable of formulating and recognizing such principles. As he has formulated the basic internalist claim, this means that such principles are not eligible as components of internalist reasons or justifications for anyone, leading to “wholesale skepticism” (pp. 287–8).

This is the most serious of the problems that Goldman raises and the hardest to deal



with in a brief space. But we can see right away that something is wrong with his formulation of the issue if we ask why an epistemic principle that is genuinely self-evident from an *a priori* standpoint for one epistemic agent should be regarded as ineligible to contribute to a reason or justification for that agent's beliefs just because it happens to be too complicated or subtle to be discernible by other, less sophisticated epistemic agents. Surely the right thing to say here is that the epistemic principles that contribute to a given agent's reasons or justifications must be available to whatever extent is necessary to that agent. This will no doubt mean that the accessibility of such reasons will vary from one such agent to another, but I am unable to see that there is anything implausible about such a result.

The second thing to say about this issue is that availability can be a matter of degree, with not all degrees of availability requiring the capacity for explicit formulation. Here the situation is quite parallel to that which pertains to logical principles. Ordinary, naïve reasoners are very unlikely to be able to formulate explicitly a principle like *modus ponens*, but may none the less be able, after it is formulated and explained to them, to recognize it as the principle that they were implicitly following. And something analogous may well be true for various epistemic principles, though the degree to which this is a plausible claim to make will vary widely from case to case. Both this point and the previous one will mean that the degree to which various beliefs are justified from an internalist standpoint is likely to vary widely from person to person, but once the issue of who gets to wave the knowledge flag is set aside as uninteresting, as I believe that it should be, it is hard to see that there is anything alarming or even particularly surprising in such a result.

The final thing that I want to say, on this occasion, about this last issue is that, while it may be relevant to the assessment of particular internalist views, it does not constitute any real objection to internalism itself. If an internalist arrives at an otherwise plausible account of the justification of a particular sort of belief, but one that relies on an

epistemic principle that is not plausibly one that ordinary people are even implicitly aware of, then that account of justification will not have any relevance to the issue of whether the belief or beliefs in question are justified for them – though it might still be of great interest as an account of how the belief or beliefs in question might be justified for those who come to appreciate the principle in question. I do not think that Goldman is right that all internalist accounts of the justification of the main sorts of beliefs that common sense regards as justified or reasonable will turn out to have this status; but, if they did, that would simply be a philosophical result to be respected like any other. If such a result seems implausible, as I believe that it does, this is, I submit, because we all believe that we do have good reasons for our beliefs about the world, not because we believe them to be justified in some unspecified way that need not involve the possession of such reasons.

### §3. IS THERE A GENUINE ISSUE?

Having said this much in explication of the internalist position and defense of it against Goldman's objections, I want to step back a bit and ask whether the issue between internalism and externalism is really as clearly defined as it is still very commonly taken to be. One reason for doubting whether this is so is the seemingly intractable character of the dispute, unusual even for a philosophical issue, in which the opposing parties not only cannot agree on a solution but often have a very hard time agreeing on even what counts as a good reason on one side or the other.

Another reason is the way in which the issue is most standardly formulated: as an issue concerning the correct account or analysis of the supposed property of *epistemic justification*. But what property is this exactly? The term is obviously a technical one, not clearly present in common sense, and it is worth asking whether there is any way of specifying its supposed meaning that is neutral enough between the two opposing sides to provide a clear target for the dispute. Here the most standard answer is the one

already briefly alluded to earlier. Epistemic justification is supposed to be one of the requirements for *knowledge*: the one that needs to be added to belief, truth, and the satisfaction of an anti-Gettier condition.

But this way of specifying the target concept will work only if there is a clear and univocal concept of knowledge itself, something that seems, to me anyway, more and more doubtful. Some epistemologists have in fact suggested that there are at least two rather different concepts of knowledge, one at least predominantly externalist in character and one predominantly internalist in character,<sup>22</sup> a suggestion that seems to me plausible, as far as it goes, but quite possibly still too confining. And if anything like this is the case, then an appeal to the concept of knowledge obviously will not serve to pick out the unique concept of epistemic justification about which internalists and externalists are supposedly disagreeing.

Indeed, given the technical character of the term “epistemic justification”, might it not be that there are simply different and incommensurable concepts of epistemic justification, one (or more?) of them internalist and one (or more?) of them externalist in character – leaving it unclear in what way these are competitors between which a choice has to be made? Even a card-carrying internalist like myself can recognize that there are important and clearly epistemological issues for which a predominantly externalist approach seems entirely appropriate and perhaps even preferable. Most of these issues fall within the confines of what Philip Kitcher has aptly labeled “the meliorative epistemological project”,<sup>23</sup> that is, the general project of assessing and improving the reliability of human cognitive efforts in a broadly empirical way.<sup>24</sup> Nor does it seem to me that any reasonable externalist should be unwilling to recognize the value and importance of the kinds of epistemic considerations that the internalist is mainly concerned with.

The most explicit and developed version of the general sort of view that I am suggesting here is that advocated by Alston in his recent book *Beyond “Justification”*.<sup>25</sup> While Alston does not, of course, deny that epistemologists

could stipulate various senses for the term “justification”, thereby creating concepts of justification that might perhaps be useful for various purposes, he denies outright that (apart from such stipulations) there is any “objective status or property of beliefs picked out by ‘justified’ ” (p. 27). And the implication, since internalists and externalists clearly are not disputing the content of merely stipulated concepts of justification, is that there is in fact no genuine issue in dispute between them at all. Alston’s alternative is what he calls an “epistemic desiderata” approach to epistemology, in which there are many different epistemic qualities that a belief can have, including at least approximately those reflected in both internalist and externalist views. All of these “desiderata” are epistemically valuable; and, while there are interesting connections among them, there is no real point in arguing for an exclusive or even primary focus on some of them as opposed to others. If Alston is right, the internalism–externalism issue, on behalf of which so much ink has been spilled, evaporates entirely.

Though there is no space here to consider the details of his argument, my own belief is that Alston is in fact very largely right: first, in his claim that there is simply no univocal, non-stipulative conception of epistemic justification to be the object of the dispute; and, second, in his suggestion that there are many different epistemic values, between which there is no need to make a choice. In particular, there are many different kinds of epistemological issues, including many that are naturally approached from a third-person, externalist standpoint, together with some that are more naturally viewed as internalist issues. What is most unfortunate from this perspective is the tendency of each side of the internalist–externalist dispute to claim exclusive possession of the field of epistemology.<sup>26</sup> Echoing Alston (and, of course, Chairman Mao), why not let many epistemological flowers bloom?

Having been reconciliatory to that extent, however, I want to insist that there is a clear way in which an internalist approach, in addition to being intellectually legitimate on

its own, has a fundamental kind of priority for epistemology as a whole. (This is why I said only that Alston was “very largely” right.) No matter how much work may be done in delineating externalist conceptions of knowledge or justification or reliability and in investigating how these apply to various kinds of beliefs or areas of investigation, there is a way in which all such results are merely hypothetical and insecure as long as they cannot be arrived at from the resources available within a first-person epistemic perspective. If, for example, an epistemologist claims that a certain belief or set of beliefs, whether his own or someone else’s, has been arrived at in a reliable way, but says this on the basis of cognitive processes of his own whose reliability is for him merely an external fact to which he has no first-person access, then his proper conclusion is merely that the belief or beliefs originally in question are reliably arrived at (and perhaps thereby are justified or constitute knowledge in externalist senses) *if* the epistemologist’s own cognitive processes are reliable in the way that he believes them to be. The only apparent way to arrive at a result that is not ultimately hypothetical in this way is for the reliability of at least some processes to be establishable on the basis of what the epistemologist can know directly or immediately from his first-person epistemic perspective. (Here the epistemic regress problem lurks.)

And this is why internalism is indispensable to epistemology as a whole, even if it is not the best way to approach every epistemological issue. Though there are many other legitimate questions and issues, only an internalist approach will ultimately do when, now paraphrasing Bishop Butler, “I sit down in a cool hour” and ask whether I ultimately have any good reasons for thinking that my beliefs are true – or indeed any good reasons for thinking that they are reliably arrived at.

# NOTES

- 1 I shall mention in particular Hilary Kornblith’s naturalistically oriented defense of externalism in his 2002. For critical discussion, see BonJour, 2006.

- 2 See Plantinga, 1993, p. 3.
- 3 Bergmann, 1997.
- 4 Steup, 1996, p. 84.
- 5 More or less the same accounts appear, e.g., in Fumerton, 1995, pp. 60–6 (though Fumerton also mentions two other versions of internalism); and in Conee and Feldman, 2004, p. 55.
- 6 Steup attempts to reconcile them by saying that “the expression ‘accessible on reflection’ may be taken as an elaboration on how the expression ‘internal to the mind’ is to be understood” (Steup, 1996, p. 85), but this seems clearly unsatisfactory, for reasons that will emerge later on in my discussion.
- 7 Conee and Feldman, “Internalism Defended,” in Conee and Feldman, 2004, pp. 55–6. In their preferred formulation, the internalist thesis is that epistemic justification supervenes on items of these kinds.
- 8 Though I shall focus almost entirely on these two conceptions of internalism, they are not the only ones to be found in the literature. It is reasonably clear, however, that these two are the most prominent and generally accepted.
- 9 The founder and main proponent of reliabilism is, of course, Alvin Goldman. See *inter alia* Goldman, 1986.
- 10 Conee and Feldman’s initial reason for adopting mentalism as the account of internalism is that “it closely parallels the counterpart distinctions [between internalism and externalism] in the philosophy of mind and ethics” (Conee and Feldman, 2004, p. 57): the “internalist” views in these other areas are supervenience theses, and Conee and Feldman’s preferred formulation of mentalism, as noted above, holds that justificatory status is supervenient on mental states. But this reason obviously has no real tendency to show that mentalism thus understood is the best way of capturing the epistemological distinction we are concerned with: that there is a significant parallel between epistemology and these other areas is obviously not something that can be assumed prior to giving an epistemologically adequate account of internalism.
- 11 I am indebted here to a valuable discussion in Fumerton, 1995, pp. 60–2.
- 12 Another example would be G. E. Moore’s view that sense-data, though objects of direct experience and so seemingly capable of contributing to justification, are themselves non-mental in character – a claim that is surprisingly plausible on a least some conceptions of sense-data and of what counts as mental. See Moore, 1922, pp. 1–30.

- 13 E.g. by Bergmann, 1997.
- 14 For more detailed discussion of these arguments, see Bonjour, 2002.
- 15 For an excellent discussion of this problem, see Conee and Feldman, "The Generality Problem for Reliabilism," in Conee and Feldman, 2004, pp. 135–65. By far the best-developed and most extensive recent attempt to solve the problem is William Alston's, in chapter 6 of his 2005. Contrary to Alston's claim, I am not convinced that he has succeeded in solving the problem, but there is very little doubt that his discussion is the place from which any future attempts in this direction should begin.
- 16 Goldman, 1999. Bracketed references in this section will be to the pages of this article.
- 17 For some elaboration of these points, see Bonjour, 2001. This paper also contains a more complete discussion of Goldman's arguments than is possible here, and I have drawn on it for some of the material of the present discussion. (For another, rather different internalist take on some of Goldman's arguments, see Earl Conee and Richard Feldman, "Internalism Defended," in Conee and Feldman 2004, pp. 53–82).
- 18 There is, to be sure, a problem in the vicinity for *some* internalist views. This is reflected in what Goldman refers to as "the problem of concurrent retrieval", which is, as he suggests, particularly a problem for the justifying reasons or arguments characteristic of holistic coherentism: it is doubtful that a reason that depends on the coherence of one's overall system of beliefs or even any very large subsystem thereof is capable of being adequately assembled and grasped even with the aid of memory and written records.
- 19 And, if true, this case constitutes an instance of knowledge. I see no reason for an internalist to deny that there is a sense or use of "knowledge" in which Sally's belief counts as knowledge (together with, very likely, other senses in which it does not and still others in which the result is, so far, uncertain). Sometimes epistemologists seem to speak of *knowledge* as though it were a flag that one is entitled to wave or an emblem that one can wear, and I have no objection to having Sally wave the knowledge flag if she cares to (or to having someone else wave it on her behalf).
- 20 People in fact rarely make use of procedures of this sort for deciding logical as opposed to mathematical issues. Certainly the picture that Goldman evokes at one point of a would-be coherentist attempting to establish the consistency of his or her set of beliefs by using a very large truth table is not a plausible view of *anyone's* actual cognitive operations.
- 21 There still might seem to be an issue as to how long the person can take to arrive at *a priori* insights of various sorts or indeed at the resulting justificatory arguments, but it is hard to see that it amounts to very much. The simple thing for an internalist to say is that justification can only result from those logical or probabilistic connections that have actually been discerned – though it must be added that such discernment can itself be more or less precise and explicit, with the strength of the resulting reason varying accordingly. If one person takes longer than others and thereby arrives at a complicated but still cogent reason, then, as long as that reason is adequately grasped in the end, there is no apparent reason why an internalist should exclude it on the basis of the time required.
- 22 For versions of such a view, see Mackie, 1976, pp. 217–20, and Sosa, 1991, p. 240 (and elsewhere in this same volume).
- 23 In Kitcher, 1992, pp. 64–5.
- 24 Goldman, in the article discussed above, describes in some detail one investigation of this general kind: a psychological investigation of the phenomenological differences (though he does not use this label) between ostensible memories of perceptions that genuinely reflect previous perceptions and those that are merely a product of imagination, with the former obviously being more reliable with respect to the truth of the resulting claims about the things that were allegedly perceived. Genuine memories of previous perception turn out, it is reported, to be richer in information concerning perceived properties like color and sound, richer in contextual information about the time and place in question, and more detailed, whereas the spurious memories resulting from imagination tend to be impoverished in these respects, but to contain much more information about the subjective cognitive operations of the person in question (pp. 290–1). Here we have an investigation from an essentially external, third-person perspective, one that yields results that might clearly be valuable, for example, in the assessment of the reliability of witnesses in various contexts and that a person who was familiar with them might also apply in assessing his or her own ostensible memories. And it *might* perhaps be useful,

though hardly essential, to formulate these results by saying that memory beliefs that satisfy the criteria for being genuine are justified in an externalist or, more specifically, a reliabilist sense.

- 25 Alston, 2005. Bracketed references in this section are to the pages of this book.
- 26 A tendency that seems to me (but perhaps I am biased here) to have been manifested more strongly by those of the externalist persuasion (along with those of the closely allied naturalist persuasion).

#### BIBLIOGRAPHY

- Alston, W.: *Beyond "Justification": Dimensions of Epistemic Evaluation* (Ithaca, NY: Cornell University Press, 2005).
- Bergmann, M.: "Internalism, Externalism and the No-defeater Condition," *Synthese* 110 (1997), 399–417.
- BonJour, L.: "The Indispensability of Internalism," *Philosophical Topics* 29 (2001), 47–65.
- BonJour, L.: "Internalism and Externalism," in *The Oxford Handbook of Epistemology*, ed. P. Moser (Oxford: Oxford University Press, 2002), 234–63.
- BonJour, L.: "Kornblith on Knowledge and Epistemology," *Philosophical Studies* 127 (2006), 317–35.
- Conee, E. and Feldman, R.: *Evidentialism* (Oxford: Oxford University Press, 2004).
- Fumerton, R.: *Metaepistemology and Skepticism* (Lanham, MD: Rowman and Littlefield, 1995).
- Goldman, A.: *Epistemology and Cognition* (Cambridge, MA: Harvard University Press, 1986).
- Goldman, A.: "Internalism Exposed," *The Journal of Philosophy* 96 (1999), 271–93.
- Kitcher, P.: "The Naturalists Return," *Philosophical Review* 101 (1992).
- Kornblith, H.: *Knowledge and Its Place in Nature* (Oxford: Oxford University Press, 2002).
- Mackie, J. L.: *Problems from Locke* (Oxford: Oxford University Press, 1976).
- Moore, G. E.: *Philosophical Studies* (London: Routledge and Kegan Paul, 1922).
- Plantinga, A.: *Warrant: The Current Debate* (Oxford: Oxford University Press, 1993).
- Sosa, E.: *Knowledge in Perspective* (Cambridge: Cambridge University Press, 1991).
- Steup, M.: *An Introduction to Contemporary Epistemology* (Upper Saddle River, NJ: Prentice Hall, 1996).

LAURENCE BONJOUR

## A Priori Knowledge

The contemporary discussion of a priori knowledge is shaped by Immanuel Kant (1965). Central to his discussion is a conceptual framework that involves three distinctions: the *epistemic* distinction between a priori and a posteriori (or empirical) knowledge (see A PRIORI/A POSTERIORI in Part III), the *metaphysical* distinction between necessary and contingent propositions (see NECESSARY/CONTINGENT in Part III), and the *semantic* distinction between analytic and synthetic propositions (see ANALYTICITY in Part III). Within this framework, Kant addresses four primary questions: What is a priori knowledge? Is there a priori knowledge? What is the relationship between the a priori and the necessary? Is there synthetic a priori knowledge? The contemporary discussion proceeds largely within Kant's framework, addressing the questions he raises and taking his answers as a starting point.

#### THE KANTIAN BACKGROUND

Kant (1965, p. 43) characterizes a priori knowledge as "absolutely independent of all experience" and contrasts it with empirical knowledge, which has its "sources" in experience. Kant's characterization cannot be taken literally since he allows that there are some respects in which a priori knowledge can depend on experience. For example, experience may be necessary to acquire the concepts involved in some propositions known a priori. Although Kant is not explicit on this point, it is generally accepted that by "source" of knowledge Kant has in mind source of justification. Hence, according to Kant,

- (APK) S knows a priori that p if and only if S's belief that p is justified a priori and the other conditions on knowledge are satisfied; and
- (APJ) S's belief that p is justified a priori if and only if S's justification for the belief that p does not depend on experience.



Kant, however, does not articulate the requirements of justification that is independent of experience.

Kant's defense of the existence of a priori knowledge rests on the contention that necessity and strict universality are criteria of the a priori. Although he maintains that the two criteria are inseparable, Kant's (1965, p. 43) most famous argument for the existence of a priori knowledge invokes the first: "if we have a proposition which in being thought is thought as *necessary*, it is an *a priori* judgment". Since Kant regards it as evident that mathematical propositions are necessary and that we know some mathematical propositions, the existence of a priori knowledge is quickly established.

Kant's contention that necessity is a criterion of the a priori commits him to the following principle regarding the relationship between a priori knowledge and necessary truth:

- (K1) All knowledge of necessary propositions is a priori.

Although he is not explicit on this matter, he also appears to endorse the converse of (K1):

- (K2) All propositions known a priori are necessary.

The conjunction of (K1) and (K2) does not entail that the categories of a priori knowledge and necessary truth are coextensive, which is a view frequently attributed to Kant, since it does not entail that all necessary truths are known.

Kant introduces the distinction between analytic and synthetic propositions with respect to propositions of the form "All A are B", and marks it in two different ways:

- (A1) An analytic (synthetic) proposition is one in which the predicate is *covertly contained* in (lies outside) the subject.  
 (A2) An analytic (synthetic) proposition adds nothing (a predicate) to the subject that has not been *thought* in it.

Kant goes on to endorse two principles linking a priori knowledge to the semantic concepts:

- (K3) All knowledge of analytic propositions is a priori.  
 (K4) Some propositions known a priori are synthetic.

Kant takes (K3) to be obvious, but maintains that (K4) requires independent support. Here he (1965, p. 53) argues that " $7 + 5 = 12$ " is synthetic since "the concept of 12 is by no means already thought in merely thinking this union of 7 and 5", and that "The straight line between two points is the shortest" is synthetic since "my concept of *straight* contains nothing of quantity, but only of quality". The first argument appeals to (A2), while the second appeals to (A1).

#### THE CURRENT DEBATE

Much of the recent work on a priori knowledge centers around Kant's primary contentions. Analyses of the concept of a priori justification divide into two primary camps. Some, such as Panayot Butchvarov (1970) and Laurence Bonjour (1998), regard (APJ) as insufficiently informative and favor a positive characterization of a priori justification:

- (APJ\*) S's belief that p is justified a priori if and only if S's belief that p is justified by  $\phi$

where " $\phi$ " designates some specific source of justification. They diverge, however, in their descriptions of the source. Others endorse (APJ), but there is controversy over the respect in which justification must be independent of experience. Albert Casullo (2003) maintains that S's justification for the belief that p does not depend on experience just in case S's belief that p is justified by some non-experiential source. Philip Kitcher (1983) argues that S's justification for the belief that p does not depend on experience just in case S's belief that p is justified by some non-experiential source and cannot be defeated

by experience (*see* A PRIORI/A POSTERIORI in Part III).

The major divide in contemporary epistemology is between those who endorse and those who reject the existence of a priori knowledge. Defenses of the a priori frequently exploit the general pattern exemplified by Kant's argument. They identify some feature of propositions that we purportedly know and contend that propositions having that feature cannot be known on the basis of experience. R. M. Chisholm (1977) and Jerrold Katz (1995), following Kant, appeal to necessity. A. J. Ayer (1952) appeals to irrefutability by experience. Bertrand Russell (1971) appeals to certainty. Detractors of the a priori fall into three camps. Some, such as Hilary Putnam (1983) and Kitcher (1983), provide an analysis of the concept of a priori knowledge and argue that alleged examples of a priori knowledge fail to satisfy the analysis. Others, such as John Stuart Mill (1973) and W. V. Quine (1963), argue that propositions alleged to be knowable only a priori can be known empirically. Those in the third camp, such as Paul Benacerraf (1973), raise concerns about the alleged source of a priori knowledge.

Renewed interest in the relationship between a priori knowledge and necessary truth is due largely to the work of Saul Kripke (1971, 1980). Kripke maintains that, since the a priori/a posteriori distinction is epistemic and the necessary/contingent distinction is metaphysical, it cannot be assumed without further argument that they are related in some non-trivial fashion. Moreover, he argues against (K1) and (K2) by providing examples of necessary propositions alleged to be known a posteriori and contingent propositions alleged to be known a priori.

The Kantian position that has received most attention is (K4). There are three different reactions. The first is to accept (K4), but to reject some of Kant's examples of synthetic a priori knowledge. Gottlob Frege (1974), for example, challenges only Kant's claim that the truths of arithmetic are synthetic. The second is to reject (K4). Ayer (1952), for example, argues that all a priori knowledge is of analytic propositions. A third, more radical reaction comes from

Quine (1963), who challenges the cogency of the analytic/synthetic distinction.

Kant's most fundamental claim is that there exists a priori knowledge. His most controversial claim is that some a priori knowledge is synthetic. The goals of the remainder of this essay are to review the case for and against the existence of a priori knowledge and the controversy surrounding synthetic a priori knowledge. For further discussion of the concept of a priori knowledge, *see* A PRIORI/A POSTERIORI in Part III. For further discussion of the relationship between a priori knowledge and necessary truth, *see* NECESSITY, MODAL KNOWLEDGE in Part III.

#### THE EXISTENCE OF A PRIORI KNOWLEDGE

Kant's defense of the existence of a priori knowledge begins with the contention that all mathematical propositions are necessary. But, according to (K1), one cannot know a necessary proposition on the basis of experience. Hence, Kant concludes, all knowledge of mathematical propositions is a priori. Let us grant that all mathematical propositions are necessary and consider the contention that one cannot know a necessary proposition on the basis of experience. The phrase "know a necessary proposition" masks a crucial distinction between knowledge of the truth value of a proposition as opposed to knowledge of its general modal status:

- (T) S knows the *truth value* of p just in case S knows that p is true or S knows that p is false.
- (G) S knows the *general modal status* of p just in case S knows that p is a necessary proposition or S knows that p is a contingent proposition.

Hence, we must distinguish two readings of (K1):

- (K1T) All knowledge of the truth value of necessary propositions is a priori.
- (K1G) All knowledge of the general modal status of necessary propositions is a priori.

Kant (1965, p. 43) offers the following defense of (K1):

(K0) "Experience teaches us that a thing is so and so, but not that it cannot be otherwise."

(K0) supports (K1G) but not (K1T) since it allows that experience can provide knowledge that a thing is so and so. Hence, Kant's argument does not establish that knowledge of the truth value of mathematical propositions is a priori – i.e., it does not establish that knowledge that  $7 + 5 = 12$  is a priori.

A proponent of the a priori might maintain that, although Kant's argument fails to establish that knowledge of the truth value of mathematical propositions is a priori, it does establish that knowledge of their general modal status is a priori – i.e., it does establish that knowledge that " $7 + 5 = 12$ " is necessary is a priori. This reading of Kant's argument, however, rests on the premise that experience provides information about only the actual world. Although this premise receives some support from the observation that one cannot "peer" into other possible worlds, it conflicts with the fact that a good deal of our scientific knowledge goes beyond what is true of only the actual world. And yet we are not at all tempted to think that such knowledge is a priori. Consequently, if a posteriori knowledge of some non-actual worlds is possible, it remains to be shown why a posteriori knowledge of all non-actual worlds is not possible.

A second feature of mathematical propositions that is invoked to show that they are knowable only a priori is their alleged immunity from empirical disconfirmation. Ayer (1952) maintains that if experiential evidence justifies us in believing that mathematical propositions are true, then it must be possible for experiential evidence to justify us in believing that they are false. But, he contends, we would not regard any experiences as justifying us in believing that a mathematical proposition is false. Therefore, experiential evidence does not justify us in believing that mathematical propositions are true. To support his second premise, Ayer

invites us to consider a situation in which we count what we take to be five pairs of objects and arrive at the result that there are only nine objects. He contends that in such a situation we would not maintain that the experiential evidence disconfirms the proposition " $2 \times 5 = 10$ ", but we would explain away the disconfirming evidence as merely apparent by invoking whatever empirical hypothesis fits best with the facts of the situation.

Ayer's support for his second premise is open to immediate objection since it is a standard feature of scientific practice to explain away isolated cases of apparent disconfirming instances to well-established generalizations by invoking some auxiliary hypotheses. But such generalizations are not immune to empirical disconfirmation. Hence, more needs to be said to substantiate the claim that mathematical propositions are immune from empirical disconfirmation. Ayer must show that the type of empirical evidence that would disconfirm a well-established scientific principle would not disconfirm a mathematical proposition. What would be required to disconfirm the former is a large number of apparent disconfirming instances together with independent tests of the auxiliary hypotheses that are invoked to explain them away as merely apparent. If the disconfirming instances are numerous and the auxiliary hypotheses are not supported by independent test, then the principle would be rejected. Suppose now that we have a large number of apparent disconfirming instances for a mathematical principle and that the auxiliary hypotheses introduced to explain them away as merely apparent are not supported by independent test. In such a situation, it appears unreasonable to dismiss the experiential disconfirming evidence as merely apparent since one has independent evidence that it is genuine.

A third feature of mathematical propositions that is invoked to show that they are knowable only a priori is their alleged certainty. Russell (1971) argues that if mathematical propositions are justified on the basis of experiential evidence, then they are justified by inductive generalization from observed cases. The conclusions of inductive



generalizations, however, are not certain. Since some mathematical propositions are certain, they are knowable only a priori. Proponents of this argument are faced with the task of specifying the sense in which mathematical propositions are certain. One response is to invoke the deductive character of mathematical proof; such proofs guarantee the truth of their conclusions. There are two problems with this response. First, the conclusion of a mathematical proof is known with certainty only if its premises are known with certainty. But the deductive character of a mathematical proof does not provide an account of the sense in which its premises are known with certainty. Second, proponents of the a priori typically maintain that it is only the most basic mathematical propositions and their obvious consequences that are known with certainty. But the most basic mathematical propositions are not known on the basis of proof. A second response maintains that epistemically basic propositions are certain in the sense that mistakes regarding them is not possible. On this account, S's belief that *p* is certain just in case, necessarily, if S believes that *p* then it is true that *p*. This sense of certainty is trivially satisfied by any necessary truth that S believes. It follows that if someone believes a mathematical truth on the basis of experience, then that person's belief is certain. Consequently, this sense of certainty does not underwrite the claim that if a proposition is certain, then it is knowable only a priori. A third response is to invoke the degree of support that a proposition enjoys. Since a proposition that has the highest degree of support is one that is not open to future disconfirmation, *p* is certain for S just in case there is no possible situation in which S would be justified in believing that not-*p*. This proposal faces the objection that mathematical propositions are not immune to empirical disconfirmation.

Some theorists argue against the existence of a priori knowledge on conceptual grounds. Putnam (1983) and Kitcher (1983) maintain that if S's belief that *p* is justified a priori, then S's belief that *p* is not rationally revisable in light of experiential evidence. But, they contend, the propositions traditionally alleged

to be knowable a priori, such as mathematical propositions, are rationally revisable in light of experiential evidence. Therefore, knowledge of such propositions is not a priori. The initial premise of the argument, however, is problematic on two counts. First, there are grounds for denying that it captures Kant's conception of a priori justification since his arguments in support of a priori knowledge do not address the issue of whether experience can defeat one's justification for believing mathematical propositions. They focus exclusively on the source of such justification. Second, it settles by fiat a substantive philosophical question since it rules out the possibility that mathematical propositions are justified both a priori and by experience (see A PRIORI/A POSTERIORI in Part III).

A second strategy for arguing against the existence of a priori knowledge is to offer an empiricist account of knowledge of the propositions alleged to be knowable only a priori. Mill (1973) offers an inductive empiricist account of knowledge of mathematical propositions. According to inductive empiricism, epistemically basic mathematical propositions are directly justified by observation and inductive generalization. Non-basic mathematical propositions are indirectly justified by virtue of their logical and explanatory relationships to the basic mathematical propositions. Mill focuses on the basic propositions of arithmetic and geometry, and argues that they are justified inductively on the basis of observed cases. Mill's position is open to numerous objections but, for our purposes, let us assume that he provides a plausible inductive empiricist account of knowledge of mathematical propositions. Casullo (2005) argues that it does not follow that such propositions are not known a priori. To draw such a conclusion is to overlook the possibility of *epistemic overdetermination*: the possibility that mathematical propositions are (or can be) justified by both experiential and non-experiential sources.

Quine (1963) offers a holistic empiricist account of mathematical knowledge. According to holistic empiricism, individual mathematical propositions are not directly confirmed or disconfirmed by experience.

Such propositions are components of scientific theories, and only entire scientific theories are confirmed or disconfirmed by experience via their observational consequences. Once again, a successful holistic empiricist account of mathematical knowledge does not entail that such propositions are not known a priori because of the possibility of epistemic overdetermination. We shall consider other aspects of Quine's attack on the a priori when we turn to (K4).

The final strategy for arguing against the existence of a priori knowledge is to maintain that such knowledge violates a plausible constraint on an adequate theory of knowledge. A recurrent concern of critics of the a priori is that such knowledge appears mysterious. If there is a priori knowledge, then presumably it has its source in some human cognitive processes. But, the critics allege, proponents of the a priori say very little about the nature of these processes or the manner in which they produce a priori knowledge. One common response of apriorists is to contend that the source of a priori knowledge is the process of "intuition" or "intuitive apprehension", and that the process produces such knowledge by providing cognitive access to abstract entities in a manner analogous to that in which visual perception provides cognitive access to physical objects. Benacerraf (1973) draws attention to one significant problem with this contention. Perception is a process that involves causal interaction between perceivers and the objects of perception. Abstract entities, however, are incapable of standing in causal relations. Given this disanalogy, some alternative explanation of how the process produces a priori knowledge is necessary.

#### SYNTHETIC A PRIORI KNOWLEDGE

Kant's most enduring contribution to the controversy surrounding a priori knowledge is (K4). One question that immediately arises is why the existence of synthetic a priori knowledge is significant. Kant regards it as significant because it sets the stage for his primary theoretical undertaking, which is to answer the question: How is synthetic a

priori knowledge possible? Kant's question, however, is puzzling in one respect. Having established that there is a priori knowledge, he is in a position to pose the question: How is a priori knowledge possible? The fact that he deems it necessary to draw the analytic/synthetic distinction and to defend (K4) indicates that Kant does not think that a priori knowledge *in general* is problematic. In particular, he views analytic a priori knowledge as unproblematic.

If synthetic a priori knowledge is epistemically problematic but analytic a priori knowledge is not, then they must differ in some way. What, according to Kant, is the difference? Kant maintains that knowledge of analytic propositions requires only possession of the relevant concepts and the principle of contradiction. Synthetic a priori knowledge, however, requires more. For example, in order to know that  $7 + 5 = 12$ , Kant (1965, p. 53) maintains: "We have to go outside these concepts, and call in the aid of the intuition which corresponds to one of them." Synthetic a priori knowledge raises special epistemic problems because of its alleged source in intuition.

The significance of (K4) is rooted in the assumption that the source of synthetic a priori knowledge is different from the source of analytic a priori knowledge. Kant, however, does not defend this assumption. Although he maintains that knowledge of analytic propositions requires only knowledge of the principle of contradiction and the content of concepts, he does not explicitly address the source of such knowledge. Since he does not explicitly address the source of analytic a priori knowledge, Kant has no basis for claiming that the source of such knowledge is different from the source of synthetic a priori knowledge, let alone that the latter is epistemically more problematic than the former. Consequently, the epistemic significance of (K4) is presupposed rather than established.

Frege endorses (K4), but contends that the truths of arithmetic are analytic. His defense of this contention requires a modification of Kant's conceptual framework. In the case of mathematical propositions, Frege (1974, p. 4<sup>e</sup>) explicates the analytic/synthetic distinction in terms of features of the proof of the

proposition in question: "If, in carrying out this process [of following the proof of a proposition], we come only on general logical laws and on definitions, then the truth is an analytic one." Frege's notion of proof, however, is silent with respect to the issue of how one knows the primitive general laws, definitions, and logical principles employed in such proofs.

Frege's conception of analyticity is broader than Kant's. It does not restrict analytic propositions to those in which the predicate is contained in the subject. Instead, any proposition whose proof consists solely of general logical laws and definitions qualifies as analytic. Armed with this broader conception of analyticity, Frege's project is to demonstrate

(F1) All arithmetic truths are analytic.

This project faces a number of formidable technical obstacles (see MATHEMATICAL KNOWLEDGE in Part III). Let us assume that they can be overcome in order to assess the epistemic implications of Frege's project. A successful demonstration of (F1) has no significant epistemic implications. A demonstration that all arithmetic truths can be proved from general logical laws and definitions, taken by itself, tells us little about knowledge of those truths since it is silent about knowledge of general logical laws and definitions. (F1) is compatible with the epistemic claim that the truths of arithmetic are knowable only via intuition.

One might suggest that, although (F1) fails to establish that arithmetic knowledge is not grounded in intuition, it does have a significant consequence regarding such knowledge. (F1) establishes that if knowledge of logic and definitions does not have its source in intuition, then knowledge of arithmetic does not have its source in intuition. This result is significant since it establishes that there is a uniform explanation of knowledge of logic, definitions, and arithmetic.

The claim that (F1) establishes that there is a uniform explanation of knowledge of logic, definitions, and arithmetic rests on an unsubstantiated assumption: the only route to arithmetic knowledge is through proof from general logical laws and definitions.

This assumption has an unwelcome consequence. It entails a wide-ranging scepticism with respect to the elementary truths of arithmetic. If arithmetic knowledge is possible only through proof from general logical laws and definitions, then very few, if any, have such knowledge.

Kant took for granted that most literate adults know a priori that  $7 + 5 = 12$ , and set out to provide an account of such knowledge. If most literate adults have such knowledge, then there must be a route to it other than the type of proof envisioned by Frege. Therefore, Frege fails to show that there is a uniform explanation of the typical literate adult's knowledge of logic, definitions, and arithmetic. There are two possible explanations of the typical literate adult's knowledge of arithmetic: either its source is the same as the source of knowledge of logic and definitions, or it is different. If it is the same, then Frege's program for establishing (F1) is unnecessary. If it is different, then we are still faced with the problem of explaining how such knowledge is possible. If Kant is right about the scope of a priori arithmetic knowledge, then Frege fails to provide an explanation of such knowledge.

Ayer (1952, p. 73) rejects (K4) on the grounds that either we must "accept it as a mysterious inexplicable fact" that there is synthetic a priori knowledge, or we must "accept the Kantian explanation which . . . only pushes the mystery a stage further back". There are two approaches to rejecting (K4). Radical empiricists, such as Mill, maintain that alleged examples of synthetic a priori knowledge are not a priori. Logical empiricists, such as Ayer, maintain that they are not synthetic; they endorse

(LE) All propositions known a priori are analytic.

According to Ayer, (LE) provides an explanation of a priori knowledge that is free of the mystery that plagues Kant's account.

Ayer (1952, p. 78), like Frege, rejects Kant's account of the analytic/synthetic distinction, and offers an alternative: "a proposition is analytic when its validity depends solely on the definitions of the symbols it

contains, and synthetic when its validity is determined by the facts of experience". Ayer's conception of analyticity is broader than both Kant's and Frege's since it does not restrict analytic propositions either to propositions whose predicate is contained in the subject or to propositions whose proof consists of general logical laws and definitions. Any proposition that is true in virtue of the definitions of the symbols it contains qualifies as analytic.

Ayer's (1952, pp. 78–9) most explicit defense of (LE) is presented in the context of discussing logical truths:

[T]he proposition "Either some ants are parasitic or none are" is an analytic proposition. For one need not resort to observation to discover that there either are or are not ants which are parasitic. If one knows what is the function of the words "either," "or," and "not," then one can see that any proposition of the form "Either  $p$  is true or  $p$  is not true" is valid, independently of experience.

His argument can be stated as follows:

- (AJ1) One need not resort to observation to discover that there either are or are not ants which are parasitic.
- (AJ2) Therefore, the proposition "Either some ants are parasitic or none are" is an analytic proposition.

(AJ1) is an epistemic premise: it asserts that the proposition in question can be known a priori. (AJ2), however, is a semantic conclusion: it asserts that the proposition in question is analytic. The validity of the argument depends on the following intermediate premise, which links the epistemic premise and the semantic conclusion:

- (AJ3) All propositions knowable a priori are analytic.

Hence, Ayer's defense of (LE) is circular.

Suppose we grant (AJ2). Does (AJ2) provide an explanation of a priori knowledge of logical truths? If we return to the previously cited passage, Ayer offers the following premise in support of (AJ1):

- (AJ0) If one knows what is the function of the words "either", "or", and "not", then one can see that any proposition of the form "Either  $p$  is true or  $p$  is not true" is valid, independently of experience.

Ayer explains a priori knowledge of logical truths in terms of an ability to see that they are true independently of experience. The literal sense of "see" involves a causal relation between the cognizer and the object perceived. But, presumably, the sense invoked by Ayer to explain knowledge of logical truths does not. Therefore, his explanation appeals to a metaphorical sense of "see" that is not further explained. Like Kant's appeal to intuition, that explanation "only pushes the mystery a stage further back".

Quine (1963) rejects the cogency of the analytic/synthetic distinction and, *a fortiori*, the cogency of (K4). Although he is not explicit on this point, his arguments are widely viewed as impugning the existence of a priori knowledge. Quine's attack is directed at a variant of Frege's conception of analyticity: a statement is analytic if it can be turned into a logical truth by replacing synonyms with synonyms. His primary target is the notion of *synonymy* and his leading contentions can be summarised as follows. First, synonymy cannot be explained in terms of definition, interchangeability *salve veritate*, or semantic rules. Second, the verification theory of meaning does provide an account of statement synonymy; but the theory presupposes radical reductionism, which is a failed program. A vestige of that program survives in the view that individual statements admit of confirmation or disconfirmation. Quine objects to this vestige since it lends credence to the idea that there are statements confirmed no matter what, which he (1963, p. 43) rejects on the grounds that "no statement is immune to revision".

There are two strands to Quine's argument. The first challenges the cogency of semantic concepts such as synonymy. The second challenges the remaining vestige of reductionism. Neither contention, however, is explicitly directed at a priori knowledge. Hence, if Quine's argument does present a

challenge to the existence of a priori knowledge, then some additional premise is necessary that connects one of its explicit targets to the a priori.

One standard reading of Quine's argument is that his goal is to undermine the central tenet of logical empiricism, (LE), by showing that the analytic/synthetic distinction is not cogent. Suppose we grant that (LE) is indeed Quine's target and that his arguments establish that the analytic/synthetic distinction is not cogent. It does not follow that there is no a priori knowledge. As we saw earlier, proponents of (LE), such as Ayer, do not base their case for the existence of a priori knowledge on the premise that we have some knowledge of analytic truths. Instead, they first offer arguments that do not invoke the concept of analyticity to show that some of our knowledge is a priori, and then they offer independent arguments to show that the propositions known a priori are analytic. Hence, the cogency of Quine's argument establishes only that their thesis about the nature of the propositions known a priori is incoherent. But from this it does not follow that either their claim that there is a priori knowledge or their supporting arguments for that claim are incoherent.

One might attempt to close the gap in the argument by maintaining that (LE) is a conceptual claim. If the concept of a priori knowledge involves the concept of analytic truth and the latter concept is incoherent, then the former is also incoherent. There are two ways in which the concept of a priori knowledge might involve the semantic concept of analytic truth: explicitly or implicitly. As we saw earlier, neither Kant's conception of a priori knowledge, (APK), nor his conception of a priori justification, (APJ), explicitly involves the concept of analytic truth. The only plausible case for maintaining that the concept of a priori knowledge implicitly involves that concept is based on two premises. The first alleges that the concept of a priori knowledge involves the concept of necessary truth. The second contends that the concept of necessary truth is analyzable in terms of the concept of analytic truth. Both premises are problematic since (APK) does

not involve the concept of necessary truth, and there is no available analysis of the concept of necessary truth in terms of the concept of analytic truth.

Putnam (1983) and Kitcher (1983) offer an alternative reading of Quine's argument, which alleges that it involves an implicit conceptual claim: If S's belief that *p* is justified a priori, then S's belief that *p* is not rationally revisable in light of experiential evidence. Since Quine maintains that no statement is immune to revision in light of recalcitrant experience, it follows that there is no a priori knowledge. We argued in the previous section, however, that the conceptual claim is open to objection. Hence, we are still left without a clear route from Quine's explicit contentions in "Two Dogmas" to the conclusion that there is no a priori knowledge.

There remains a final challenge to the existence of a priori knowledge that draws its inspiration from "Two Dogmas". One goal of a theory of knowledge is to articulate the sources and extent of human knowledge; a second goal is to explain how those sources generate the knowledge in question. Therefore, if a theory of knowledge endorses a category of knowledge but cannot explain how that knowledge is possible, then the theory is unacceptable. The only available non-mysterious explanation of how a priori knowledge is possible involves the analytic/synthetic distinction, which Quine has shown to be incoherent. Therefore, a theory of knowledge that endorses the a priori is unacceptable.

This challenge rests on the contention that the analytic/synthetic distinction offers the only available non-mysterious explanation of how a priori knowledge is possible. As we saw earlier, however, neither Kant's nor Frege's nor Ayer's conception of analyticity offers much in terms of an explanation of a priori knowledge. Nevertheless, the problem posed by the final challenge remains. A theory of knowledge endorsing the a priori must offer some explanation of how such knowledge is possible. Once we recognize that the concept of analyticity offers little in terms of explanatory power, we are in a position to recognize that the explanatory problem goes beyond



the coherence of the analytic/synthetic distinction. The more general problem that must be addressed is how a priori knowledge is possible.

#### THE POSSIBILITY OF A PRIORI KNOWLEDGE

In order to advance the debate over the existence of a priori knowledge, one must be clear about its parameters. Apriorists maintain that there are certain uncontroversial domains of human knowledge, such as logic and mathematics, that are known a priori. Radical empiricists do not deny that we have the knowledge in question. Their contention is that such knowledge is empirical. Radical empiricists, on the other hand, place an exclusive premium on empirical knowledge and, in particular, on scientific knowledge. Apriorists do not deny that we have empirical and scientific knowledge. Their contention is that such knowledge involves elements, such as logical and mathematical principles, that are not justified by experience. Hence, the primary debate between apriorists and radical empiricists is not over the extent of human knowledge. There are broad domains, including logic, mathematics, and the sciences, where both agree that we have knowledge. Their primary disagreement is over the source of that knowledge. They offer two different accounts of the source of that knowledge, and each claims that its theory provides the better explanation of the knowledge in question.

Two points emerge regarding the debate between apriorists and radical empiricists. First, in order for apriorists to advance their case against radical empiricism, they must offer supporting evidence for the claim that there are non-experiential sources of knowledge, and they must show that those sources explain knowledge of the truths in the relevant domains. Second, in advancing their case against radical empiricism, apriorists can appeal to any domain of knowledge where both camps agree that we have the knowledge in question. In particular, since both camps

agree that we have empirical and scientific knowledge, apriorists can appeal to such knowledge in making their case for a priori knowledge.

Apriorists, such as Bonjour, (1998), Christopher Peacocke (2004), and Sosa (2000), typically offer philosophical supporting evidence, a mix of phenomenological and a priori considerations, for the existence of a priori knowledge. Casullo (2003) argues that a more promising strategy is to supplement the philosophical considerations with empirical evidence. This strategy consists of two related projects. The first, the Articulation Project, is philosophical in character and involves providing a more precise characterization of alleged non-experiential sources of knowledge and the range of beliefs that they justify:

Provide (a) a generally accepted description, at least at the phenomenological level, of the cognitive states that non-inferentially justify beliefs a priori; (b) the type of beliefs they justify; and (c) the conditions under which they justify the beliefs in question.

The second, the Empirical Project, provides empirical evidence that underwrites the claim that the non-experiential sources generate knowledge of the propositions in question and explains how they do so:

Provide (a) evidence that the cognitive states identified at the phenomenological level are associated with processes of a single type or relevantly similar types; (b) evidence that the associated processes play a role in producing or sustaining the beliefs they are alleged to justify; (c) evidence that the associated processes are truth-conducive; and (d) an explanation of how the associated processes produce the beliefs they are alleged to justify.

Two considerations support the dual strategy. First, a case for the a priori that is based on empirical considerations is one that radical empiricists must acknowledge by their own lights. Second, by relying solely on a priori considerations, apriorists place themselves in a needlessly handicapped position when defending their primary contention.



## BIBLIOGRAPHY

- Ayer, A. J.: *Language, Truth and Logic* (New York: Dover, 1952).
- Benacerraf, P.: "Mathematical Truth," *The Journal of Philosophy* 70 (1973), 661–79.
- BonJour, L.: *In Defense of Pure Reason* (Cambridge: Cambridge University Press, 1998).
- Butchvarov, P.: *The Concept of Knowledge* (Evanston, IL: Northwestern University Press, 1970).
- Casullo, A.: *A Priori Justification* (New York: Oxford University Press, 2003).
- Casullo, A.: "Epistemic Overdetermination and A Priori Justification," *Philosophical Perspectives* 19 (2005), 41–58.
- Chisholm, R. M.: *Theory of Knowledge*, 2nd edn (Englewood Cliffs, NJ: Prentice-Hall, 1977).
- Frege, G.: *The Foundations of Arithmetic*, 2nd edn revised, trans. J. L. Austin (Evanston, IL: Northwestern University Press, 1974).
- Kant, I.: *Critique of Pure Reason*, trans. Norman Kemp Smith (New York: St Martin's Press, 1965).
- Katz, J.: "What Mathematical Knowledge Could Be," *Mind* 104 (1995), 491–522.
- Kitcher, P.: *The Nature of Mathematical Knowledge* (New York: Oxford University Press, 1983).
- Kripke, S.: "Identity and Necessity," in *Identity and Individuation*, ed. M. K. Munitz (New York: New York University Press, 1971).
- Kripke, S.: *Naming and Necessity* (Cambridge, MA: Harvard University Press, 1980).
- Mill, J. S.: *A System of Logic*, ed. J. M. Robson (Toronto: University of Toronto Press, 1973).
- Peacocke, C.: *The Realm of Reason* (Oxford: Oxford University Press, 2004).
- Putnam, H.: "Two dogmas' revisited," in *Realism and Reason: Philosophical Papers*, vol. 3 (Cambridge: Cambridge University Press, 1983).
- Quine, W. V.: "Two Dogmas of Empiricism," in *From a Logical Point of View*, 2nd edn revised (New York: Harper and Row, 1963).
- Russell, B.: *The Problems of Philosophy* (Oxford: Oxford University Press, 1971).
- Sosa, E.: "Modal and Other A Priori Epistemology: How Can We Know What Is Possible and What Impossible?," *Southern Journal of Philosophy* 38 (2000), Supplement, 1–16.

## FURTHER READING

- Boghossian, P. and Peacocke, C., eds: *New Essays on the A Priori* (Oxford: Oxford University Press, 2000).

Casullo, A., ed.: *A Priori Knowledge* (Aldershot: Ashgate, 1999).

DePaul, M. and Ramsey, W., eds: *Rethinking Intuition* (Lanham, MD: Rowman and Littlefield, 1998).

ALBERT CASULLO

## The Common Sense Tradition

## I

Among the most prominent representatives of the common sense tradition in philosophy are the eighteenth-century philosopher Thomas Reid and the twentieth-century philosophers G. E. Moore and Roderick Chisholm. In describing the common sense tradition, Chisholm writes, "It is characteristic of 'commonsensism,' as an alternative philosophical tradition, to assume that we do know, pretty much, those things we think we know, and having identified this knowledge, to trace it back to its sources and formulate criteria that will set it off from things we do not know."<sup>1</sup> We may say that the common sense tradition holds that we do know many of the things we ordinarily take ourselves to know, and our philosophical views should be adequate to the fact that we do know them. It holds that we may use much of what we take ourselves to know as data for developing and assessing philosophical theories and principles.

We find such a stance in the works of Thomas Reid. He took Hume's empiricism to imply that we have no knowledge of the material world, no knowledge of the future, the past, or other minds, and no knowledge of ourselves as continuing subjects of experience. Reid took Hume to show that the wages of empiricism are a rather thoroughgoing skepticism. Reid writes:

A traveller of good judgment may mistake his way, and be unawares led into a wrong track; and while the road is fair before him, he may go on without suspicion and be followed by others but, when it ends in a coal pit, it requires no great judgment to know he hath gone wrong, nor perhaps to find out what misled him.<sup>2</sup>

Reid held that, since we do know many of the things that Hume's empiricism would rule out, so much the worse for empiricism. We find a similar theme throughout Moore's writings. In his 1919 essay, "Some Judgments of Perception", Moore begins by considering some views about perception and the possibility of perceptual knowledge. He rejects certain philosophical views because they imply that we cannot know various facts about the external world:

But it seems to me a sufficient refutation of such views as these, simply to point to cases in which we do know such things. This, after all, you know, really is a finger: there is no doubt about it: I know it, and you all know it. And I think we may safely challenge any philosopher to bring forward any argument in favour either of the proposition that we do not know it, or of the proposition that it is not true, which does not at some point rest upon some premiss which is, beyond comparison, less certain, than is the proposition which it is designed to attack.<sup>3</sup>

Chisholm takes a similar stance: "We reject the sceptical view according to which there is no reason to believe the premises of an inductive argument ever confer evidence upon the conclusion. If the skeptical view were true, then we would know next to nothing about the world around us."<sup>4</sup> Chisholm holds that, since we do know a lot about the world around us, so much the worse for skepticism about induction.

Chisholm says that it is characteristic of the common sense tradition to assume that we do in fact know pretty much what we ordinarily think we know. But what does this include? In his "A Defense of Common Sense", Moore lists a variety of propositions that he takes himself to know. Among these are the propositions that he was much smaller when he was born, that the earth existed for many years before he was born, that since he was born he has been in contact with or not far from the surface of the earth, and that there are other people who have had a variety of different experiences. He also claims to know that other people have known similar things. So among the things Moore takes himself to know are propositions about the past, external

objects, other people's mental states, and *epistemic* propositions to the effect that many people know those same, or similar, things. Reid and Chisholm would agree with Moore's assessment.

But *how* do we know such things? What, for example, makes one's belief that there are other people epistemically justified? Those, of course, are important epistemological questions. Indeed, there are significant differences among Reid, Moore, and Chisholm concerning the nature of our knowledge of the external world.<sup>5</sup> But, while the search for criteria that explain what makes justified beliefs justified or what makes a belief an instance of knowledge is a central epistemological undertaking, philosophers in the common sense tradition hold that one need not know such criteria in order to have knowledge and justified beliefs, and one need not know such criteria in order to identify instances of knowledge and justified belief. Thus, I can know that I *do* know there are other people without knowing *how* I know it and certainly without having a well-developed theory about how knowledge of other minds is possible. Moore writes, "We are all, I think, in this strange position that we do know many things, with regard to which, we know further that we must have had evidence for them, and yet we do not know how we know them, i.e. we do not know what the evidence was."<sup>6</sup> The view that we can pick out some instances of knowledge and justified belief independently of knowing general criteria of knowledge and justification is characteristic of *particularism* in epistemology. Reid, Moore, and Chisholm, are epistemological particularists.<sup>7</sup>

Let us consider some views to which the common sense tradition is *not* committed. First, the common sense tradition is not committed to the view that everything that might be properly called a "common sense belief" is true or even reasonable. For example, in his "A Defense of Common Sense", Moore writes:

The phrases "Common Sense view of the world" or "Common Sense beliefs" (as used by philosophers) are, of course, extraordinarily vague; and for all I know, there may be many

propositions which may be properly called features in “the Common Sense view of the world” or “Common Sense beliefs”, which are not true, and which deserve to be mentioned with the contempt with which some philosophers speak of “Common Sense beliefs”. But to speak with contempt of “Common Sense beliefs” which I have mentioned is quite certainly the height of absurdity. And there are, of course, enormous numbers of other features in “the Common Sense view of the world” which, if these are true, are quite certainly true too: e.g. that there have lived on the surface of the earth not only human beings, but also many different species of plants and animals.<sup>8</sup>

Moore’s defense of common sense beliefs is clearly limited. Suppose we think of a “common sense proposition” as one that is deeply and widely held. In that case, Moore holds that some common sense propositions, e.g. that there are other people and that they think and feel, are known. But Moore is not committed to the view that *all* such propositions are known. Again, the same could be said of the views of Reid and Chisholm.

Second, the common sense tradition is not committed to the view that there is a *sui generis* “faculty of common sense” or that we know various common sense proposition on the basis of such a faculty. Such a view is simply not to be found in Moore or Chisholm. Reid, of course, does refer to a faculty of common sense. It is not clear, however, that Reid considers it to be a *sui generis* faculty. Indeed, he suggests that “common sense” is “only another name for one branch or degree of reason”.<sup>9</sup>

Third, while the common sense tradition holds that some common sense propositions are known, it is *not* committed to the view that they are known *in virtue of* their being common sense propositions or *in virtue of* their being deeply and widely held. It is not committed to the view that being deeply and widely held confers, or is a source of, any positive epistemic status on a proposition. Furthermore, the common sense tradition is not committed to, and in fact rejects, the view that we know common sense propositions, *p*, on the basis of inferring *p* from:

(1) Whatever is a common sense proposition

is true, and (2) *p* is a common sense proposition. Our knowledge that there are other people, for example, does not depend on such an inference. (Reid does seem to hold that *knowing* that a proposition is widely and deeply held by almost everyone is *some* reason for believing it. But even for Reid this reason is defeasible. Such a view is not to be found in Moore or Chisholm.)

It would be a mistake to view the proponents of the common sense tradition as arguing:

- (1) If *p* is a common sense proposition (one that is widely and deeply accepted), then *p* is true or reasonable.
- (2) Theory T denies that *p* is true.
- (3) Therefore, Theory T is false or unreasonable.

“Why”, one might ask, “should we reject a metaphysical, epistemological, or scientific theory simply because it conflicts with some proposition that is widely and deeply believed? Indeed, why assume that Premise (1) is true?” I see no reason why we should regard Moore, Reid, or Chisholm as committed to this sort of argument. What they *do* hold is that there are some common sense propositions that we know and that we should reject philosophical theories because they imply that we do not know those things. We may take them typically to be arguing:

- (1) Proposition *p* is a common sense proposition that I and many others know.
- (2) Theory T implies that no one knows *p*.
- (3) Therefore, Theory T is false.

But this does not seem to be an unreasonable stance. Rejecting a theory because it conflicts with something one knows is not an unreasonable procedure. So consider, for example, the proposition, *p*: that there are other people. I and many others know this proposition. Some forms of skepticism imply that no one knows *p*. They imply that it is false that some people know that there are other people. We may take the common sense philosopher to hold that, since he and many others know that *p*, it follows that those forms of skepticism that deny that anyone knows *p* are mistaken.<sup>10</sup>

Philosophers in the common sense tradition hold that they can pick out instances of knowledge and justified belief and use these as data in formulating and assessing philosophical theories. Many of the beliefs they take as data are common sense beliefs. But *why* take various common sense beliefs as data? Commonsensists have sometimes offered different answers. For example, Chisholm says that in investigating the theory of knowledge from a philosophical point of view, we assume that what we know is pretty much that which, on reflection, we think we know. He goes on to say, "This might seem like the wrong place to start. But where else *could* we start?"<sup>11</sup> As a defense of the common sense tradition this does not seem compelling. Surely there are alternatives one *could* adopt. One could, for example, follow a Cartesian model and confine one's data to what is certain and indubitable, or one could confine one's data to propositions about one's own mental states. One cannot plausibly defend the common sense tradition on the ground that it is the only approach available.

A second answer can be found in the following striking passage from Reid's *Inquiry*:

To what purpose is it for philosophy to decide against common sense in this or any other matter? The belief in a material world is older, and of more authority, than any of the principles of philosophy. It declines the tribunal of reason, and laughs at the artillery of the logician. It retains its sovereign authority in spite of all the edicts of philosophy, and reason itself must stoop to its orders. Even those philosophers who have disowned the authority of our notions of an external world, confess that they find themselves under a necessity of submitting to their power.

Methinks, therefore, it were better to make a virtue of necessity; and, since we cannot get rid of the vulgar notion and belief of an external world, to reconcile our reason to it as well as we can; for, if Reason should stomach and fret ever so much at this yoke, she cannot throw it off; if she will not be the servant of Common Sense, she must be her slave.<sup>12</sup>

In this passage, Reid refers to the "authority" of our belief in the material world. Indeed, he

tells us that it has *more* authority than any of the principles of philosophy. The concept of "authority" is a normative notion, and Reid is making a normative claim about such beliefs. He is not merely claiming that the belief in the material world is more firmly held, but that it is also *worthy* of being more firmly held, that it is more *reasonable* to believe. Still, Reid also points out that the belief in the material world is irresistible. He counsels us to "make a virtue of necessity"; since we can't get rid of certain common sense beliefs, such as a belief in the existence of a material world, we should make our philosophical views fit these deeply held beliefs. Indeed, Reid often points to the difficulty of giving up certain sorts of beliefs and to the futility of skeptical arguments. In explaining why he does not ignore the testimony of his senses in the face of skeptical arguments, Reid says, "because it is not in my power; why then should I make a vain attempt. . . . My belief is carried along by perception, as irresistibly as my body by the earth."<sup>13</sup> Here again Reid seems to be defending various common sense beliefs on the basis of their irresistibility.

The appeal to irresistibility, however, does not seem to be a satisfactory way of defending the common sense tradition. First, why would the irresistibility of such beliefs matter for assessing philosophical theories? The fact that a belief is irresistible might be an excuse for holding it, but it is not a good reason for holding it and it would not be a good reason for rejecting other beliefs that conflict with it. Consider a bad habit that one cannot give up, such as smoking or overeating. Surely one should not abandon one's belief that such habits are unworthy of being indulged simply because one cannot give them up. It would be unreasonable to change one's beliefs about the value of one's habits simply because of a conflict between one's beliefs and one's deeply rooted habit. Similarly, would it not be unreasonable to change one's philosophical views simply because of a conflict between them and one's deeply rooted and irresistible beliefs?

Second, perhaps many common sense beliefs are irresistible. Perhaps this is true of beliefs such as there are bodies and there are



other people, but, as we have seen, the common sense philosopher also takes as data various epistemic propositions. Such epistemic propositions include: I know that I have a body, and almost everyone knows that there are other people. Is it really clear that these epistemic propositions cannot be given up? Could one not hold, for example, "Yes, I have a body. I cannot give up that belief. But it isn't something I *know*. Maybe it's more likely than not, but I don't know it." It is not clear, therefore, that *all* the propositions that various common sense propositions would take as data are irresistible. If we ask, then, what makes it reasonable to take various common sense beliefs as data, irresistibility cannot be the explanation.

Suppose we reject both Chisholm's suggestion that we should take various common sense beliefs as data because we have no place else to start and Reid's view that we should take various common sense beliefs as data because they are irresistible. Why, then, should we take various common sense beliefs as data? Certainly, one answer is that these beliefs *are* instances of knowledge and reasonable belief. The common sense philosopher might thus point to the positive epistemic status of the beliefs he takes as data. Reid, Moore, and Chisholm all hold that they *know* various common sense propositions and that almost everyone *knows* them. Thus, Moore says he *knows* that this is a finger and that this proposition is *more reasonable* for him than the premises of skeptical arguments which imply that it is false. Similarly, Chisholm writes in defense of his theory of knowledge and his commonsensism that "it corresponds with what we do know".<sup>14</sup> The common sense philosopher may point to the positive epistemic character of various beliefs, to their being epistemically justified or instances of knowledge, as that which makes it reasonable for him to take them as data.

### III

The common sense tradition has faced a variety of objections. Sometimes it is objected

that common sense beliefs can be false and unreasonable. Consider the following criticism:

Intuitive and common-sense judgments can be false, as a little reflection illustrates. Such judgments, furthermore, seem not always to be supported by the best available evidence. Consider, for instance, how various judgments of "common sense" are at odds with our best available evidence from the sciences or even from ordinary perception.<sup>15</sup>

Perhaps *some* common sense beliefs are epistemically shoddy, but there is no reason to think that is true of the particular common sense beliefs to which Moore, Chisholm, and Reid appeal. Philosophers in the common sense tradition need not endorse everything that might be properly called a common sense belief. As we have seen, Moore does *not* claim that all common sense beliefs are true or justified. Even though there are good reasons for rejecting *some* common sense beliefs because they are "not supported by the best available evidence" or are at odds with "our best available evidence from the sciences or even from ordinary perception", there is no reason whatsoever to reject Moore's claims that I know this is a finger or I know that there are other people on such grounds. Those particular claims don't suffer from such defects.

As we have seen, philosophers in the common sense tradition hold that they can pick out particular instances of knowledge and justification. Critics of commonsensism sometimes raise questions about the epistemic status of our particular epistemic judgments:

It is often left unclear what the epistemic status of the relevant preanalytic epistemic data is supposed to be. Such data, we hear, are accessed by "intuitions" or by "common sense". We thus have some epistemologists talking as follows: "Intuitively (or commonsensically), justification resides in a particular case like *this*, and does not reside in a case like *that*." A statement of this sort aims to guide our formation of a notion of justification or at least a general explanatory principle concerning justification. A simple question arises: is such a statement self-justifying, with no need of independent epistemic support? If so, what notion of

self-justification can sanction the deliverances of intuition or common sense, but exclude spontaneous judgments no better, epistemically, than mere prejudice or guesswork?<sup>16</sup>

This passage suggests that the common sense tradition holds that epistemic facts are accessed *via* "intuition" or "common sense". If this means that the common sense tradition holds that we know epistemic facts on the basis of a faculty of common sense or intuition, as we know about the past on the basis of memory, then there is no good reason to accept this claim. There is no good reason to think that Moore or Chisholm held such a view. Though one might be able to make a stronger case that this is Reid's view, it is not clear that this is his view, either. Moreover, the commonsensist need not hold that particular epistemic judgments such as "I know this is a finger" are self-justifying. He might hold that particular epistemic judgments are justified in virtue of their cohering with one's other beliefs, or because they are the product of a reliable cognitive faculty or intellectual virtue, or because they are appropriately related to what is given in experience. There are many candidates for sources of justification besides self-justification.

Of course, what makes particular epistemic beliefs justified is an important problem for epistemology. But, as we have seen, Moore distinguishes knowing *that* some proposition is true from knowing *how* we know it or what *makes* it an instance of knowledge. Moore also holds that one can have the former sort of knowledge without the latter. Such a view seems plausible. Consider our knowledge of evaluative properties such as the beauty of a piece of music or the wrongness of an action. One might know that a particular piece of music is beautiful or that an act is wrong without knowing what makes it so. I might know, for example, that it would be wrong for me to kill my secretary now without knowing what makes it so. I might not know whether it is wrong because it fails to maximize utility or fails to treat her as an end or for some other reason. In any case, the common sense philosopher may agree that it is unclear what justifies our particular epistemic beliefs,

without denying that they are justified. Even if our particular epistemic judgments are problematic in the sense that we do not have a satisfactory explanation of what makes them justified, we need not deny that we know they are justified.

Still, many philosophers would find this response unsatisfactory. Suppose we accept the plausible view that epistemic properties such as *being justified* and *being an instance of knowledge* are evaluative properties and that, like other evaluative properties, *supervene* on descriptive properties. According to this view, just as the rightness of an action supervenes on some of its descriptive features, e.g. maximizing the balance of pleasure over pain or not treating any human as a mere means, so, too, a belief has the property of being justified or being an instance of knowledge in virtue of some of its descriptive properties. But what must one know about a belief in order to know that it has such an epistemically evaluative property? Consider the following suggestion by William Alston:

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 has *Q* is justified.)<sup>17</sup>

Alston suggests that in order to be justified in believing that some particular belief is justified one must be justified in believing some general epistemic principle. He supports this view by appealing to the supervenient character of epistemic properties.

Reflection on the preceding passage suggests the following principle:

- S: One is justified in believing that *x* has *F* (where *F* is an evaluative property) only if one is justified in believing both  
(a) that *x* has some non-evaluative



property Q, and (b) a general principle to the effect that whatever has Q has F.

According to S, one is justified in a particular attribution of an evaluative property only if one is justified in believing some general evaluative principle. If S is true, then the common sense philosopher is justified in believing his particular epistemic claims only if he is justified in believing the relevant general principle. But, the objection goes, since they are not justified in believing the relevant general principle, they are not justified in believing their particular epistemic claims. If S is true, then the particularism of common sense philosophers such as Reid, Moore, and Chisholm is mistaken.<sup>18</sup>

But is S true? Many critics would say “no”. First, we should be careful not to confuse claims about exemplification with claims about application. In assuming that evaluative properties are supervenient, we assume that a thing has an evaluative property in virtue of its *exemplifying* some descriptive properties. But from this it does not follow that the *application* or attribution of evaluative properties must be based on a justified belief that a thing has those properties or justified belief in a general evaluative principle. So S does not follow from the mere fact that evaluative properties are supervenient.

Second, it seems plausible that there are justified attributions of supervenient properties that do not depend on justified belief in general principles. Many philosophers hold, for example, that mental properties supervene on physical properties. They hold that being in pain or thinking about Boston supervenes on various physical properties, e.g. physical properties of one’s brain. But does our knowledge that we are in pain depend on our knowing some other propositions such as (a) I have F, and (b) whatever has F is in pain? Surely I can know that I am in pain without knowing on what physical properties pain supervenes. I can know that I am in pain without knowing some general principle about the connection between being in a certain physical state and being in pain. If this is right, then justified attributions of supervenient properties need not meet the requirements

we find in S. The supervenient character of epistemic properties would seem to pose no problem for the commonsensist’s view.

Sometimes the epistemic claims of the commonsensist are rejected on the basis of skeptical arguments. Typically, such arguments hold that the beliefs of the commonsensist fail to meet some condition that the skeptic holds is necessary for knowledge. Consider, for example, the following passages by Keith Lehrer:

Thus, before skepticism can be rejected as unjustified, some argument must be given to show that the infamous hypotheses employed by the skeptics are incorrect and the beliefs of common sense have the truth on their side. If this is not done, then the beliefs of common sense are not completely justified because conflicting skeptical hypotheses have not been shown to be unjustified. From this premiss it follows in a single step that we do not know these beliefs to be true because they are not completely justified.<sup>19</sup>

To meet the agnolological challenge of skepticism, we must provide some argument to show that the skeptical hypothesis is false and the beliefs of common sense are correct. And this leads to a second equally inescapable conclusion. The challenge cannot be met.<sup>20</sup>

Lehrer suggests that knowledge that *p* requires two things. First, one must *show* that *p* is true; one must give some argument that *p* is true. Second, one must *show*, and thus give an argument, that skeptical hypotheses incompatible with *p* are false. In order to know that *p* one must show that *p* is true and one must be able to “exclude” any hypothesis one knows to be incompatible with *p*. So we have the following “principle of exclusion”:

PE: One knows that *p* only if one can show that *p* is true and one can show to be false any proposition, *q*, one knows to be incompatible with *p*.

Now, consider the proposition that I have hands (*h*) and the skeptical hypothesis that I am a handless brain in a vat being caused to have experience as if I had hands (*v*). Let us suppose further that I know that *v* is

incompatible with *h*. I know that if *v* is true, then *h* is false. So, if *v* is true, then I do not know *h*. According to PE, I know that *h* only if I can show that *h* is true and show that *v* is false. But, says Lehrer, since we cannot do either, we do not know that *h* is true. If such an argument is sound, then a great many of the epistemic claims of the common sense philosopher are false.

In response, however, the common sense philosopher might ask why should we accept PE? First, why would a belief that *p* amount to knowledge only if one could *show* that it is correct? Such a view rules out the possibility of immediate knowledge. Again, why must one show that *p*? Would it not be enough that one had evidence for *p*? Why, in addition to having evidence for *p*, or being justified in believing that *p* would one also need to show that *p* is true? Second, why think that knowledge requires that one *show* that the skeptical alternatives are false? Why would it not be sufficient that one know that they are?

As we have seen, Moore holds that it is more reasonable for him to believe the epistemic claim, I know this is a finger, than one or more of the premises in skeptical arguments that deny that claim. So now consider the following propositions:

*h*: I know I have hands.

PE: One knows that *p* only if one can show that *p* is true and one can show to be false any proposition, *q*, one knows to be incompatible with that *p*.

Which of these propositions is it more reasonable for us to believe? Is it not clearly *h*? Again, while it would be desirable to have an explanation for the greater reasonableness of various common sense beliefs, we need not have such an explanation in order for them to *be* more reasonable or for us to *know* that they are. Still, it is open to the common sense philosopher to appeal to a familiar array of epistemic views in order to explain the greater reasonableness of his common sense epistemic claims over the competing philosophical principles. He might hold, for example, that we are more reliable about the former, that our belief in the former issues from a more

virtuous intellectual faculty, or that the former cohere better with the rest of our beliefs. He might point to such factors or to a combination of them to explain why his common sense epistemic claims are more reasonable.

Consider, for example, which coheres better with one's total body of beliefs: the proposition that I know I have hands or some philosophical principle such as PE or S. One's total body of beliefs will include such things as I have hands, almost all human beings have hands, I am very good at identifying hands and other human appendages, almost everyone knows that he and others have hands, I see that I have hands, perception under the present conditions is highly reliable, I and others know a great many things about the world around us, etc., etc. If we wonder which coheres better with our total body of beliefs, can there be any serious doubt that it would be the Moorean epistemic claims rather than the premises of the skeptical argument and the philosophical principles which underlie them?

Of course, it might be held that reasonable belief requires more than coherence. Perhaps it will be urged that the reliability of the source of belief is also relevant to the reasonableness of belief. But, even if this is so, is there any good reason to think that the common sense knowledge claims would fare worse in that respect than the philosophical principles? Indeed, the common sense philosopher might plausibly hold that we are far more reliable about whether we know humble epistemic propositions such as "I know this is a finger" than abstract philosophical principles such as PE or S. In any event, the reasonableness of one's believing a proposition would seem to depend, at least in part, on the extent that the proposition cohered with one's total body of beliefs. So, even if one held that coherence was not the *only* relevant factor in assessing a belief's reasonableness, it seems that it is *a* factor and one that the common sense epistemic claims enjoy to a much greater extent than the premises of the skeptical argument and the philosophical principles which underlie them.

Finally, let us consider one further objection raised by Laurence Bonjour:

[T]o accept commonsense convictions as Moore and other particularists do, does appear to rule out illegitimately even the possibility that skepticism might in fact be true, that common sense might be mistaken. And, equally important, if this solution is taken at face value, it would have the effect of stifling or short-circuiting epistemological inquiry at least as effectively as would simply acquiescing to skepticism.<sup>21</sup>

Does common sense particularism stifle epistemological inquiry? That depends on what one takes to be involved in epistemological inquiry. It does not seem to stifle the search for criteria of knowledge and justification. On the contrary, common sense philosophers such as Chisholm take the search for such criteria to be among the principle aims of epistemology. Even if we assume that we do know various things about the world around us, we might wonder how we know them. We still might seek philosophical answers to that question.

Furthermore, even if we are common sense particularists we need not ignore skeptical arguments or consider them idle. One could be a particularist and take skeptical arguments seriously, attempting to identify what plausible, yet mistaken, assumptions yield skeptical conclusions.

BonJour also objects that common sense particularism appears to rule out illegitimately the possibility that skepticism is true. But here we should be careful. Does it rule out the possibility that skepticism is true? Not if the relevant sense of possibility is that of "logical possibility". Common sense philosophers such as Chisholm, Moore, and Reid grant that it is logically possible that skepticism is true. They grant that it is logically possible that we know next to nothing about the world around us. They grant the logical possibility of deception by evil demons or dream experiences. Still, they would deny that skepticism is "epistemically possible" in the sense that, given what we know, it is reasonable to believe that skepticism is true. They would hold that, since we do know much about the external world, it is not reasonable to accept skepticism about the external world. But they would deny that this is to rule out skepticism *illegitimately*.<sup>22</sup>

## NOTES

- 1 Chisholm, 1982, p. 113.
- 2 Reid, 1983, p. 11.
- 3 G. E. Moore, "Some Judgements of Perception" (in Moore, 1960), p. 228.
- 4 Chisholm, 1973b, p. 232.
- 5 They disagree, for example, about the role of sense data or "ideas" in perception and about whether our knowledge of the material world is immediate. Indeed, their own views seem to change over time. In some places, for example, Moore holds that we have some immediate knowledge of the external world and in other places he suggests that such knowledge is inferential. Cf. Moore, "Hume's Philosophy Examined" (in Moore, 1953), p. 142, and "Four Forms of Scepticism" (in Moore, 1959), p. 226.
- 6 G. E. Moore, "A Defence of Common Sense" (in Moore, 1959), p. 45.
- 7 Chisholm discusses particularism in his 1973b. For more on the topic, see Sosa, 1980; and Lemos, 2004, ch. 6.
- 8 G. E. Moore, "A Defence of Common Sense" (in Moore, 1959), p. 45.
- 9 Reid, 1969, p. 567.
- 10 In the passage from his 1960 "Some Judgments of Perception" cited above, Moore rejects certain skeptical views because he takes them to imply that it is false that he and others know that this is a finger. Moore rejects these skeptical views because they imply, incorrectly, that he and others do not know that this is a finger.
- 11 Chisholm, 1977, p. 16.
- 12 Reid, 1983, p. 4.
- 13 *ibid.*, p. 85.
- 14 Chisholm, 1977, p. 121.
- 15 Moser, 1998, p. 364.
- 16 *ibid.*, p. 363.
- 17 Alston, 1976, p. 170; cf. Alston, 1980, p. 579.
- 18 The view that knowledge of particular ethical or evaluative facts depends on knowing general criterion is widely held. Mill writes in the first chapter of *Utilitarianism*: "A test of right and wrong must be the means, one would think, of ascertaining what is right or wrong, and not a consequence of having already ascertained it." Mill, 1979, p. 2; cf. also Hare, 1952, p. 111.
- 19 Lehrer, 1978, pp. 358–9.
- 20 *ibid.*, p. 361.
- 21 BonJour, 2002, p. 265.
- 22 For sympathetic and more detailed treatments of the common sense tradition's response to skepticism, see: Sosa, 1999, pp. 141–53; Lycan, 2001, p. 41; Lemos, 2004 and 2006.

## BIBLIOGRAPHY

- Alston, W.: "Two Types of Foundationalism," *The Journal of Philosophy* 73 (1976).
- Alston, W.: "Some Remarks on Chisholm's Epistemology," *Nous* 14 (1980).
- BonJour, L.: *Epistemology* (Lanham, MD: Rowman and Littlefield, 2002).
- Chisholm, R. M.: *The Foundations of Knowing* (Minneapolis, MN: University of Minnesota Press, 1982).
- Chisholm, R. M.: *Theory of Knowledge*, 2nd edn (Englewood Cliffs, NJ: Prentice Hall, 1977).
- Chisholm, R. M.: "On the Nature of Empirical Evidence," in *Empirical Knowledge*, ed. Roderick M. Chisholm and Robert J. Swartz (Englewood Cliffs, NJ: Prentice-Hall, 1973).
- Chisholm, R. M.: *The Problem of the Criterion* (Marquette, WI: Marquette University Press, 1973b).
- Hare, R. M.: *The Language of Morals* (Oxford: Oxford University Press, 1952).
- Lehrer, K.: "Why Not Scepticism?," in *Essays on Knowledge and Justification*, ed. G. Pappas and M. Swain (Ithaca, NY: Cornell University Press, 1978).
- Lemos, N.: *Common Sense* (New York: Cambridge University Press, 2004).
- Lemos, N.: "Moore and Skepticism," in *The Oxford Handbook of Skepticism*, ed. J. Greco (New York: Oxford University Press, 2006).
- Lycan, W.: "Moore against the New Skeptics," *Philosophical Studies* 103 (2001).
- Mill, J. S.: *Utilitarianism* (Indianapolis, IN: Hackett, 1979).
- Moore, G. E.: "Some Judgments of Perception," in *Philosophical Studies* (London: Routledge and Kegan Paul, 1960).
- Moore, G. E.: "Hume's Philosophy Examined," in *Some Main Problems of Philosophy* (New York: Macmillan, 1953).
- Moore, G. E.: "Four Forms of Skepticism," in *Philosophical Papers* (London: George Allen and Unwin, 1959).
- Moser, P.: "Epistemological Fission," *The Monist* 81 (1998).
- Reid, T.: *Essays on the Intellectual Powers of Man* (Cambridge, MA: MIT Press, 1969).
- Reid, T.: *Inquiry and Essays*, ed. Ronald E. Beanblossom and Keith Lehrer (Indianapolis, IN: Hackett, 1983).
- Sosa, E.: "The Foundations of Foundationalism," *Nous* 14 (1980), 547–65.
- Sosa, E.: "How to Defeat Opposition to Moore," in *Philosophical Perspectives* 13, ed. J. Tomberlin (Cambridge, MA: Blackwell, 1999), 141–53.

NOAH LEMOS

## The Power of Perception

Philosophy often begins with wonder about the ordinary, and nothing is more ordinary than perceptual belief. It is as familiar as everyday experience. We look at a gray coffee mug and believe what we see: it's a mug, it's gray, it's empty, and so on. Beliefs such as these tend to enjoy some form of epistemic merit: typically they are epistemically justified, warranted, and instances of knowledge.<sup>1</sup> Each such belief gains its status from originating in perception. Indeed, none of the coffee mug beliefs just mentioned would have any epistemic merit at all if they had their origin not in perception, but instead in wishful thinking, imagining, or pure guesswork. Here, then, is our question:

How does perceptual experience give external world beliefs epistemic merit, when these other ways of forming the same beliefs do not?

In this essay, I shall discuss various attempts to answer this question. My discussion will be based on the assumption that perceptual beliefs are often *basic*. A perceptual belief is basic if its epistemic merit does not derive from any supporting beliefs but is gained from perceptual experience alone. Thus, more precisely, our question is going to be:

How can a belief acquire epistemic merit solely by virtue of being supported by a perceptual experience?

Clearly, then, my approach to the issue is foundationalist. However, coherentism will not be dogmatically ignored. To the contrary, as I shall explain in the first section, my benchmark test for a successful theory of perception will be whether the theory can give a satisfactory rebuttal to each of several key arguments in defense of coherentism.

## EXPECTATIONS FOR A SUCCESSFUL THEORY

There are three seemingly sound arguments which aim to undermine foundationalism and support coherentism. Each of them

purports to show that perceptual experiences cannot, without the help of supporting beliefs, justify perceptual beliefs. For a theory of basic perceptual support to be successful, it must explain why these arguments fail. This is the first expectation a successful theory of perceptual support must meet. The second expectation is that a theory of perceptual support explain the limits of basic perceptual support. Let us briefly consider each of these desiderata.

The first anti-foundationalist argument is the *dilemma argument*. It says that there are only two ways for any mental state to support a belief epistemically:

- (i) A mental state, *having epistemic merit itself*, can *transfer* its epistemic merit on a belief by virtue of having a propositional content that logically supports the content of the belief.

Experiences, however, are not bearers of epistemic merit, for they are not the sorts of things that can be justified or warranted. Experiences cannot, therefore, support belief in this first way, even if they have, as some epistemologists claim, propositional content.

- (ii) A mental state can support a belief indirectly via epistemically merited beliefs about the mental state's occurrence. If we know that the occurrence of a particular mental state reliably indicates the truth of a particular proposition, then our knowledge that the mental state has occurred supports our belief in the proposition.

However, if an experience supports a belief in this way, its support isn't basic. Our beliefs about the experience – that it has occurred and that its occurrence is a reliable sign of the proposition's truth – provide the epistemic support, not the experience itself.

The argument concludes that, since there is no third way for a mental state to support a belief epistemically, there is no such thing as basic perceptual support.<sup>2</sup>

The *background information argument* begins with the premise that there is a significant difference between perceiving an *F* and

perceiving *that* a particular thing is an *F*. For example, it is one thing to perceive a coffee-mug on a table, another to perceive *that* the object on the table is a coffee-mug. Basic perceptual support requires the latter. Our mere experience of a mug, when we have no idea of what we are experiencing, does not epistemically support the belief that there's a mug. The next premise asserts the necessity of background information: To perceive that some object *a* is an *F*, we need the background belief that an object looking the way object *a* is looking in our experience typically is an *F*. We don't see that the object on the table is a mug unless we believe that an object looking the way the object on the table does tends to be a mug. As a result, no perceptual experience provides perceptual support for a corresponding perceptual belief without the requisite background beliefs. Since such background beliefs are always a necessary part of our perceptual evidence, it follows that our perceptions never provide basic support.

The *reliability argument* is founded on the *reliability principle*, which asserts that the beliefs we form through a cognitive faculty have positive epistemic status only if they are based on a belief in the faculty's reliability, where that belief enjoys a positive epistemic status in turn. Thus, no perceptual belief has positive epistemic status without being based on a further epistemically privileged belief that our perception is reliable. There is, therefore, no such thing as basic perceptual support.<sup>3</sup>

Each of these three arguments enjoys a good deal of initial plausibility. But, if they are sound, our perceptual beliefs never acquire epistemic merit through *basic* support. Consequently, a successful theory of basic perceptual support must explain where these arguments fail.

Let us turn to the second expectation. A successful theory of perception must explain the limits of basic perceptual support. Consider the following observation of Thomas Reid's:

Not only men, but children, idiots and brutes, acquire by habit many perceptions which they had not originally. Almost every employment in life, hath perceptions of this kind that are peculiar to it. The shepherd knows every sheep



of his flock, as we do our acquaintance, and can pick them out of another flock, one by one.<sup>4</sup>

The shepherd knows his sheep the way a Supreme Court justice knows pornography. He just sees which sheep is which, employing no conscious or unconscious beliefs about their characteristics. How far can such basic perceptual support extend?<sup>5</sup> If the shepherd can just see that a particular sheep is Maggie, can the mountain guide just see that a certain toehold is weak, can we just see that our friends are sad, or that certain behaviors are rude or even immoral? In general terms, the question is this: What kind of propositions can be justified or even known via basic perceptual support? A successful theory must provide a satisfactory answer to this question.<sup>6</sup>

Let's now examine the two main theories of perceptual support: representational realism and direct realism.<sup>7</sup>

#### REPRESENTATIONAL REALISM

Representational realism, as an account of basic perceptual knowledge, makes three claims. First, when we have basic empirical knowledge, the content of our knowledge is immediately given to us in experience. Second, only facts about mental entities or states, such as sense-data or adverbial features of our experience (e.g. that we are visually appeared to in a certain way), are immediately given to us.<sup>8</sup> Third, our perceptual knowledge of the external world ultimately derives from our basic knowledge of our mental states.<sup>9</sup>

Representational realism can be motivated in two ways. First, its advocates may employ the *argument from illusion*, the conclusion of which asserts that the immediate object of our awareness in perception is always a mental state or entity such as a sense-datum or a way in which an object appears in one's experience. Arguably, if this conclusion is correct, perceptual supports ultimately derives from sense-data or ways in which objects appear to us in our experiences.<sup>10</sup> Second, if the content of our perceptual knowledge is a mental entity or state of which we are

immediately aware, then representational realism gives us a reply to the three arguments introduced above. Since the experience that a belief about the experience describes is *immediately present* to us, this experience can support the belief about it without:

- (i) having any propositional content generating a need for the experience itself to be justified;
- (ii) the aid of background beliefs about how the objects we are perceiving appear to us in the experience the belief is about;
- (iii) our knowing that the perceptual faculties involved in the experience are reliable.

The dilemma argument is rebutted by (i), the background information argument by (ii), and the reliability argument by (iii).

Representational realism is fraught with problems. One way to appreciate them is to consider how they arise within a carefully developed statement of the view, of which Laurence Bonjour has recently provided an excellent example.<sup>11</sup> Bonjour takes knowledge to require justification, where having justification for a belief amounts to possessing good reasons for thinking that it is true. He also assumes that "for a mental state to be conscious is for it to have an intrinsic, constituent feature, one that is part of its own internal character and that depends not at all on any further reflective or apperceptive state", and that "the basic content of sensory experience is essentially non-conceptual in character".<sup>12</sup> Our basic perceptual knowledge is of conceptual descriptions of the non-conceptual content of our experiences. The experiences themselves are our good reasons for our beliefs.<sup>13</sup> This is because we are "in a good, an ideal, position to judge directly whether the conceptual description is accurate as far as it goes, and if so, to be thereby justified in accepting the belief".<sup>14</sup> With regard to our perceptual beliefs about the external world, we do not occupy such an ideal position, for we lack "direct or immediate cognitive access to the physical world . . . thus preventing the perceptual awareness of physical objects from constituting in itself a basis of foundational justification".<sup>15</sup>



To see how Bonjour's theory fails to meet our expectations, let's start with the dilemma argument. Bonjour rejects the argument's two alternatives: An experience doesn't support a belief through a relation between their respective propositional contents or through our knowledge that its occurrence is a reliable sign of the belief's truth. He offers a third alternative: An experience is a good reason for a belief about it because we are in an "ideal position" to judge whether the belief correctly describes the experience. It would appear, then, that Bonjour's theory offers us a reply to the dilemma argument. Yet what is "ideal" about our position, and how does this ideal status enable an experience to support a belief independently of any supporting beliefs? Bonjour never says. He correctly rejects the view some early representational realists (e.g. Descartes) have defended, according to which our position vis-à-vis our experiences is ideal by virtue of being infallible about them.<sup>16</sup> This view must be rejected because mistakes about the content of our experiences seem possible. Nor could our position's being ideal be explained by saying that our position vis-à-vis our experiences allows us to form non-inferential beliefs about them. The problem with this suggestion is that many of our perceptual external-world beliefs of epistemic merit are also non-inferential, even though our position vis-à-vis external objects is supposed to be non-ideal. Bonjour's view may simply be that our position is ideal because it allows for our beliefs about our experiences to be justified without receiving support from other beliefs. But the appeal to our being ideally positioned vis-à-vis our experiences is intended to explain how non-inferential justification is possible. Obviously, then, it would not be illuminating to explain our position's being ideal in terms of non-inferential justification. So the problem posed by the dilemma argument remains unresolved.

Bonjour's theory also fails to explain where the background information argument goes wrong. Consider Roderick Chisholm's case of the speckled hen.<sup>17</sup> Suppose you observe a hen with three speckles. Focusing on your perceptual experience, you believe you are

experiencing a visual image of a hen with three speckles. Contrast this with a case in which you observe a hen with dozens of speckles. Focusing on your experience, you believe you are experiencing a visual image of a hen with forty-eight speckles. Whereas your belief in the first case is justified, your belief in the second is not. Appealing to the premises of the background information argument, we can easily account for the difference between the two cases. In the first case, you are in possession of the background information that being appeared to three-speckles-like is a feature of your experience that you can reliably detect. In the second case, you are in possession of the background information that being appeared to forty-eight-speckles-like is *not* a feature of your experience that you can reliably detect.

Bonjour treats the case in terms of four necessary and jointly sufficient conditions for basic empirical justification. Where *E* is an experience and *F* is one of its features, we have basic perceptual justification for believing that *E* is *F* if and only if

- (1) we are aware of *E*,
- (2) we are attending to the fact that *E* is *F*,
- (3) we believe that *E* is *F*, and
- (4) we apprehend the agreement between the fact that *E* is *F* and the conceptual description given by our belief.

You meet all four conditions when your experience involves a three-speckled image. You only meet the first three when your experience involves a forty-eight speckled image, for in this second case you don't apprehend agreement between what your experience is actually like and the content of our belief about this experience.<sup>18</sup>

What is it, however, to apprehend (as in the first case) or to not apprehend (as in the second case) agreement between one's experience and one's belief about it? Bonjour does not say. Apprehending such agreement presumably does not consist in having the kind of justified background beliefs to which the background information argument appeals. Nor can Bonjour appeal to *de facto* reliable abilities of discrimination, for such an appeal

would contradict his internalist view that belief-forming processes need not be reliable to supply good reasons for our beliefs.

Finally, consider the challenge presented by the reliability argument. In response to this argument, representational realists could consistently reject the reliability principle for each of our faculties. This move would be problematic, however, for it would partially undercut the restriction – essential to representational realism – of basic empirical knowledge to knowledge of one's own mental states. If neither one's immediate awareness of one's own mental states nor one's mediate perception of external facts is subject to the reliability principle, why restrict a subject's basic knowledge to the former? BonJour opts for a double standard. Perceptual beliefs about the external world *are* subject to the reliability principle. Beliefs about the content of one's own experiences based on one's immediate awareness of them are *not*. They need only meet the lesser requirement that one do not have any reason to believe that one's immediate awareness is unreliable. In their case, "as long as there is no special reason for thinking that a mistake is likely to have occurred, the fact that such a belief seems via direct comparison to accurately characterize the conscious experience that it purports to describe apparently provides an entirely adequate basis for thinking that the description is correct and hence an adequate basis for justification".<sup>19</sup>

Yet why does the fact that the belief is formed through a "direct comparison" of the non-conceptual content of experience and the conceptual content of the belief make it an exception to the reliability principle? If the directness of the comparison does not ensure some particularly tight truth connection such as infallibility or incorrigibility, why does it make a difference? Even if a particular belief is formed by an infallible faculty, it doesn't follow that this belief is not subject to the reliability principle, according to which a belief is justified only if it has its origin in a faculty we know to be reliable. Suppose our memory is only 80 percent reliable and some other faculty is 100 percent reliable. This difference in reliability hardly explains why our

memory, but not the other faculty, is subject to the requirement that we must know a faculty to be reliable before it can provide us with knowledge.<sup>20</sup>

BonJour's theory also comes up short relative to our second expectation that a good theory explain the limits of basic support. On his account, our experiences provide basic support for our beliefs about them so long as we are ideally positioned to determine their accuracy. Without knowing what it is for our position to be ideal, however, we cannot determine just when or why this is the case. Our position clearly is not ideal with regard to all beliefs about our experiences. When we see the coffee-mug, our experience does not, according to representational realism, give basic support to our belief that the experience was caused by a coffee-mug. We might say that our position is ideal with regard to beliefs describing aspects of our experience that are "immediately given" to us, but which aspects are those?<sup>21</sup>

There is a further problem with representational realism in general that merits attention. Representational realists must deny that our ordinary perceptual beliefs about the external world have epistemic merit, as we clearly don't base them on beliefs about our experiences. We see and believe that there's a gray coffee-mug before us without forming such beliefs as that we are appeared to grayly or that we have a mug-type sense-datum, and without inferring the mug's existence from such beliefs. Representational realists, however, do not embrace skepticism as a consequence of their view. Rather, they generally present their theory as offering an alternative basis for our external-world beliefs. They hold that, if we were to employ this alternative basis, it would give our beliefs about the external world epistemic merit.<sup>22</sup> The problem here is two-fold. First, according to some critics, the alternative basis is inadequate: there is no legitimate way to move via inference from knowledge of our own mental states to knowledge of the external world.<sup>23</sup> Second, and more fundamental, typically we gain knowledge of the external world without employing the representational realists' alternative strategy. For example, we can see and know, without employing any inferences

from beliefs about our experiences, that there is a gray mug on the table.

#### DIRECT REALISM

Direct realism asserts that basic perceptual support can confer epistemic merit upon not only beliefs about our own mental states but also upon beliefs about the external world.<sup>24</sup> It thus affirms our actual knowledge of the external world without appealing to an awareness of one's own mental states beyond our ordinary perceptual experiences. Unlike representational realism, direct realism does not maintain an *ad hoc* exception to the reliability principle, affirming it for external-world beliefs but not for beliefs about the content of our experience. Rather, direct realism rejects the principle equally for beliefs about one's own experiences and beliefs about the external world, and thus allows its adherents to rebut the reliability argument in a straightforward fashion, namely by insisting that for a belief to gain epistemic support from a faculty we do not need any reason to believe that this faculty is reliable. (Direct realists would agree, however, that if a faculty is to generate epistemic merit we must not be in possession of a good reason to think it unreliable.)<sup>25</sup>

Direct realism faces an objection based on counter-examples involving hallucination. Compare two cases: In the first, a subject *S* forms a true external world belief *B*, based on a perceptual experience *E*. Call this the *veridical case*. In the second case, *S* forms a belief with the same content, *B\**, on the basis of a hallucinatory experience *E\**. Call this the *hallucination case*. Suppose further *E* and *E\** are subjectively indistinguishable. The objection proceeds as follows:

- (1) In the hallucination case, *B\** is not based on direct awareness of external reality. (This follows from the nature of hallucination.)
- (2) If *B\** is not based on direct awareness of external reality, then *E\** does not give basic support to *B\**. (This is based on the assumption that basic support requires direct awareness of external reality.)

- (3) If *E\** does not give basic support to *B\** in the hallucination case, then *E* does not give basic support to *B* in the veridical case. (This is claimed to follow from the assumption that *E* and *E\** are subjectively indistinguishable.)
- (4) *E* does not give basic support to *B* in the veridical case.

Since an analogous argument can be constructed for any case of alleged basic support, the critic of direct realism will conclude that there are no cases of basic perceptual support at all: direct realism is false.<sup>26</sup>

Direct realists can meet this objection by rejecting either (2) or (3).<sup>27</sup> In either case, they must develop an account of basic perceptual support that sustains their reply. The focus of the debate thus shifts to the question of whether the proposed account succeeds in answering our initial question. Direct realists have developed answers along both internalist and externalist lines. No version has been fully successful.

#### EXTERNALIST DIRECT REALISM

According to some forms of direct realism, the epistemic merit of our perceptual beliefs stems from external features attaching to:

- (i) perceptual experiences themselves, or
- (ii) perceptual belief-forming processes of which the experiences are part, or
- (iii) the faculty of perception.

External features are such that we cannot detect them by reflection alone. Standard candidates for external features include the following conditions:

- (i) that our experience be an adequate indication of how things are (things would not seem to us as they do if our belief were not true);
- (ii) that our perceptual process be truth-reliable (the process must be an instance of a relevant type that generally produces true beliefs rather than false ones);

- (iii) that the reliable function of our perceptual faculty be in accord with a design plan aimed at the production of true beliefs.<sup>28</sup>

Note that these conditions can obtain without our meeting the requirements proposed in the arguments against basic perceptual support. Thus, externalist versions of direct realism allow for a ready response to the dilemma and background information arguments. The dilemma argument mistakenly assumes that either our experience itself or our belief in its occurrence must constitute a reason in support of our perceptual belief. Such reasons are not required. A belief's being caused under the right external conditions is sufficient for its enjoying basic perceptual support. The background information argument mistakenly assumes that to know perceptually that there's a mug on the table our evidence must include the information that what we are experiencing is how mugs generally appear to us. Again, we needn't base our belief on such evidence. For our perceptual beliefs to acquire epistemic merit through basic support, it's sufficient that they be formed in the externally correct way.

The appeal to external conditions also explains the limit of basic perceptual support. The limit is set by our ability to form perceptual beliefs in a way that is non-inferential and has the right external features. When the shepherd learns to recognize his sheep visually, he develops a new way of forming his beliefs about their identities, one that does not employ an inference from other beliefs but meets the external conditions required for knowledge.

The externalist form of direct realism is none the less subject to serious criticisms.<sup>29</sup> Some critics claim that it overlooks an important aspect of basic perceptual support: if an experience is to justify a belief, the subjective character of the experience must be appropriately related to the belief. Suppose you believe on the basis of sight and taste that the liquid in the mug is water. You also believe that it is H<sub>2</sub>O. Your experience provides basic support for the first belief but not the second. None the less, both of your beliefs might meet the same external conditions. For

example, each might be formed by a reliable process, or each might be such that things would not seem to you as they do unless the belief were true. What makes the difference, and is overlooked by externalism, is the internal "fit" between the subjective character of your experience and the content of your beliefs. Your water belief fits the subjective character of your experience, but your H<sub>2</sub>O belief does not.<sup>30</sup>

This objection refutes some forms of externalist direct realism but not the approach in general. Externalists can acknowledge that basic perceptual support is defined by a mix of internal and external conditions, including a subjective fit between experience and belief. This particular aspect of basic perceptual support may even lend itself to an explanation in externalist terms after all. Perhaps your experience doesn't support the belief that the liquid is H<sub>2</sub>O because forming that belief on that basis fails to meet the external condition of being in accord with our design plan.

According to a second objection, reflecting the internalist point of view characteristic of traditional epistemology, externalist direct realism is based on a misconception of what knowledge requires. Knowledge requires justified belief. Thus, a successful account of how knowledge can be gained through perception should also explain how perception can generate justification. However, the necessary and sufficient conditions for our having justified beliefs are wholly internal: Experiences make beliefs justified for us only if we can determine the fact of their support by reflection alone.<sup>31</sup> Consequently, an exclusively externalist account of perceptual support cannot explain how justification can arise from perceptual experiences, and thus cannot completely explain how perceptual experiences can ground knowledge.

Externalists can respond in two ways. First, they can deny that perceptual knowledge, as traditionally understood, requires the kind of epistemic justification internalists have in mind.<sup>32</sup> Second, they can assert that they are offering a new, superior concept of knowledge, according to which the form of epistemic merit essential to knowledge involves solely external conditions. The argument for

abandoning the traditional conception of knowledge then needs to be made. Externalists variously claim that internalist conceptions of knowledge are flawed because they are:

- unable to account for our knowledge of the external world;
- over-intellectualized, requiring us to have beliefs we don't ordinarily form;
- unable to account for the perceptual knowledge of young children and animals, and unable to account for the metaphorical uses of "know".<sup>33</sup>

As we have observed, some of these criticisms have significant force against representational realism. Whether internalist forms of direct realism invariably succumb to them remains to be seen.

#### INTERNALIST DIRECT REALISM

How might a relation of basic support between experiences and external world beliefs be determined just by what is accessible to us in reflection? Since the variety of plausible responses includes more than we've space to consider, we'll examine a recent view that nicely exemplifies several of the strengths and weaknesses of internalist direct realism.<sup>34</sup>

When we have a perceptual experience, it seems to us as if a particular proposition is true. For example, when we perceive a mug on the table, it seems to us as if the proposition "There is a mug on the table" is true. Such perceptual seemings are reasons in support of our corresponding beliefs. Michael Huemer has developed this idea employing what he calls the *Principle of Phenomenal Conservatism*:

PC: If it seems to *S* as if *p*, then *S* thereby has at least *prima facie* justification for believing that *p*.<sup>35</sup>

Huemer offers PC as accounting for the justification of all foundational beliefs. The concept of seeming it employs is thus quite general, covering not only perception but also memory and *a priori* intuition. He takes the concept of justification in PC to be internalist, and he

explains *prima facie* justification as justification that is both foundational and defeasible. According to Huemer, PC is a necessary truth and, upon reflection, self-evident.<sup>36</sup>

Phenomenal conservatism offers the same reply to the reliability argument as other forms of direct realism. The view concedes, of course, that if on a particular occasion of acquiring a perceptual belief we subsequently gain good reason to suspect that our perceptual experience was on that occasion unreliable, then that reason might defeat our *prima facie* justification for the belief in question. However, phenomenal conservatism rejects the requirement that, to have basic, *prima facie* justified perceptual beliefs, we need to form a justified belief to the effect that our perceptual experiences are reliable. The dilemma and background information arguments are dispatched in an equally straightforward way. Contrary to the dilemma argument, our experiences, even though they aren't bearers of justificational status, support our external-world beliefs through logical relations between contents of the experiences and the contents of the beliefs. That things seem a certain way supports the belief that that's how they are, without there being any need for the seemings themselves to be justified. Contrary to the background information argument, our perception that something is an *F* epistemically supports our belief that there is an *F*, independently of our having reason to believe that our experience is an instance of how an *F* generally appears to us. If it seems to us as if something is an *F*, then we are *prima facie* justified in believing accordingly.

Phenomenal conservatism defines the limit of basic perceptual support as the limit of our ability to have experiences in which things perceptually seem a certain way to us without our making other beliefs the basis for our belief that that is how things are. When the shepherd initially looks at his sheep, none of them seems to be a particular one. As he becomes more familiar with each, he reaches the point where, without his relying on other beliefs, one just seems to be Maggie and another just seems to be Betsy. His beliefs about their identities are then basic *prima facie* justified



beliefs. How we acquire such perceptual seemings merits further investigation, but it's the seemings that make for basic perceptual support.

For all its promise, however, phenomenal conservatism doesn't tell us all we need to know. We begin with the question: How does the perceptual experience we have when we look at a coffee-mug on the table epistemically support our belief that there is a mug on the table? Phenomenal conservatism tells us the following: In our experience, it perceptually seems to us as if there is a mug on the table, and, necessarily, whenever it seems to us as if something is the case, we are, absent a defeater, justified in believing that that is how things are. Yet, if we ask *why* such a seeming renders the corresponding beliefs *prima facie* justified when desiring, imagining and guessing do not, phenomenal conservatism's answer is: "It just does." It's a brute, necessary, self-evident fact. Phenomenal conservatism thus gives up too soon. The recognition that this question needs an answer beyond "It just does" is, after all, a compelling motivation for externalist direct realism, which answers the question by an appeal to external features that distinguish perceptual seemings from other states.<sup>37</sup> Even though we reject externalism, we should not dodge the questions it attempts to answer.

Another defect in phenomenal conservatism is that its principle PC is false. As can easily be shown by way of examples, some seemings do not, owing to their origin in epistemically unsuitable processes, render the corresponding beliefs *prima facie* justified. Suppose that you have learned to identify a gold nugget on sight but I have no such knowledge. We both look at one. My desire to discover gold makes it seem to me as if it is gold; your learned identification skills make it seem that way to you. According to PC, the belief that the nugget is gold has *prima facie* justification for both of us. Yet, certainly, my wishful thinking does not gain my belief the same defeasible justification as your learned identification skills.<sup>38</sup>

To sum up, phenomenal conservatism contains two problems endemic to internalist direct realism. First, its explanation of basic

perceptual support is not sufficiently deep or revelatory. Second, it doesn't capture the difference between perceptual seemings from which epistemic merit does originate and other kinds of seemings – for example those originating in wishful thinking – from which it does not. It is tempting to blame the failure of phenomenal conservatism on its commitment to internalism. However, we have seen that the externalist alternative is fraught with problems of its own. How, then, within internalist direct realism, can we develop an adequate explanation of basic perceptual support?

#### A PRACTICAL THEORY OF PERCEPTION<sup>39</sup>

When we learn how to do something, we develop behaviors directed towards a defining purpose which unifies them and provides a basis for their evaluation. In learning how to reason, in particular, we develop ways of forming beliefs in response to other beliefs and experiences, and our defining purpose is to reason correctly, which amounts to forming beliefs on grounds that make them objectively likely.<sup>40</sup> Our reasoning behaviors are often unconscious but none the less accessible. We can generally determine the basis for our beliefs.

Since learning how to do something involves developing goal-directed behaviors, we may say, with caution, that knowing-how involves adopting norms that require particular behaviors in particular circumstances to serve a particular purpose. The point of caution is that to adopt a norm is not necessarily to formulate it consciously. Rather, what's sufficient for having adopted a norm is having acquired a disposition to engage in the purposeful behavior the norm describes. Our epistemic norms describe behaviors we internalize in learning to reason; they direct us to adopt specific beliefs in specific circumstances to serve the purpose of reasoning correctly.

Knowing how to do something has an internal dimension (adopting a set of behaviors directed towards a purpose) and an external one (having those behaviors actually support that purpose). We can have the



internal dimension without the external one. Suppose I am deceived by a demon who provides me with extensive bicycling experiences, although I never actually mount a bike. I just think I do. I learn to ride, just as I might through a virtual reality bicycling simulator, but I never *really* ride. I have the internal aspect of knowing how to bicycle, as I have adopted a set of behaviors directed towards the goal of riding correctly. I lack the external dimension. My behaviors do not result in my actually riding a bike. In a similar way, victims of a continuing demon-created hallucination might develop the internal dimension of knowing how to reason, and in particular how to form specific beliefs on the basis of distinctive experiences, but never actually form perceptual beliefs on a basis that makes their truth likely.

Our epistemic norms authorize that particular experiences support particular beliefs for us. We can explicate the concept of direct perceptual evidence as follows:

DE: *S*'s experience, *e*, directly evidences the belief that *p* for *S* if and only if (1) *e* is an instance of a phenomenological type, *t*, such that *S* has an epistemic norm that directs *S* to believe that *p* on the basis of *t*-type experiences in the absence of defeaters, and (2) *S* does not have an epistemic norm that directs *S* to believe that not-*p* on the basis of *t*-type experiences in the absence of defeaters.

An experience directly evidences, or epistemically supports, a belief for us just when we have internalized epistemic norms that univocally call for the formation of the belief on the basis of experiences of that phenomenological type.<sup>41</sup>

DE is not yet an adequate account of basic perceptual support. We need to attend to the fact that the learning process by which we develop our epistemic norms can be defective. Suppose that I am confused about what it is to reason correctly. I think it includes reasoning so as to gain spiritual peace. Feelings of spiritual peace reinforce certain belief-forming behaviors for me. Since I gain

such a feeling from believing that the angel Gabriel exists on the basis of experiencing a colorful sunset, I adopt, for the purpose of reasoning correctly, a norm directing me to believe upon experiencing such sunsets that Gabriel exists. According to DE, my colorful sunset experience directly evidences my belief in Gabriel for me. Within my system of epistemic norms, my belief in Gabriel has epistemic merit. None the less, viewed more broadly, from a perspective that takes into account the mistake in my development of my norms, my belief in Gabriel is defective. If I were not confused about the relation between correct reasoning and spiritual peace, I would not believe in Gabriel on the basis of colorful sunsets.

An account of basic perceptual support should honor this broader perspective. Let's say that a perceptual experience directly and properly evidences a belief for us just when the experience gives us direct and undefeated perceptual evidence for the belief and there are no flaws in the learning process by which we developed the relevant norms. Our learning was not, for example, guided by a misconception of what it is to reason correctly.

DPE: *S*'s experience, *e*, directly and properly evidences the belief that *p* for *S* if and only if (1) *S* bases the belief that *p* on an experience, *e*, which directly evidences *p* for *S*, (2) *S*'s belief that *p* on the basis of *e* is undefeated, and (3) the learning process by which *S* has adopted the norm that authorizes believing *p* on the basis of *e* is non-defective.

Beliefs have basic perceptual support for us when they are directly and properly evidenced by our experience. Our perception of the coffee-mug on the table directly and properly evidences our belief that there is a coffee-mug on the table. Our epistemic norms direct us to form the belief on the basis of the coffee-on-the-table-like experience in the absence of defeaters. Since we have no defeaters, and since we've learned our norms in a non-defective manner, our belief that there is a coffee-mug on the table has epistemic merit.

In DPE, then, we have in internalist terms specified how an experience can be a reason in support of an external world belief.<sup>42</sup>

Like other forms of direct realism, the practical account of basic perceptual support explains why the reliability argument is not sound. According to this argument, perceptual support requires beliefs attributing reliability to our perceptual experiences. DPE tells us that perceptual support does not require such beliefs. The practical account also shows where the dilemma and background information arguments go wrong. The dilemma argument overlooks an important form of epistemic support: experiences constitute reasons for our beliefs by virtue of our having properly learned to form those beliefs on the basis of those experiences. Contrary to the background information argument, our perception of something as *F* supports our belief that it is *F* independently of our believing with justification that what looks that way is *F*. It is enough that we have learned how to detect *F*s on the basis of that phenomenological sort of experience.<sup>43</sup>

The limit of basic perceptual support is the limit of our perceptual know-how. Whether experiences of a particular phenomenological sort directly and properly evidence beliefs of a particular propositional content for us is determined by whether a properly learned epistemic norm authorizes us to form beliefs with that content in response to experiences of that phenomenological type. The shepherd who knows his sheep has learned to identify each on the basis of a different phenomenological type of experience.<sup>44</sup>

## CONCLUSION

We have examined several strategies to accomplish the task of explaining the epistemic power of perception, and we have done so within a foundationalist framework. However, we have taken coherentism into account at least indirectly, since we articulated the expectation that a good theory of epistemic support must offer satisfactory rebuttals of three standard considerations on its behalf: the dilemma, background information and

reliability arguments. Externalist views were given a fair hearing, but we noted that they are open to serious objections. Among the internalist options, we have considered representational realism and phenomenal conservatism. Both of these theories have been found wanting because the former is fraught with difficulties while the latter tells us too little. Hope lies in the practical approach, which emphasizes our knowing how to identify external facts on the basis of experiences. The normative epistemic relation between our experiences and our beliefs is neither a brute necessary fact nor the result of the fact that the occurrence of the former happens to be associated with the truth of the latter. The epistemic merit of our basic perceptual beliefs stems from rules we've developed for ourselves, relative to the goal of forming beliefs on a basis that makes them likely to be true.<sup>45</sup>

## NOTES

- 1 Epistemically justified beliefs are ones it is reasonable or rational to believe. Epistemic warrant is whatever, when added to truth and belief, makes knowledge. Knowledge is true, epistemically warranted belief. I leave open whether epistemic justification is necessary for knowledge. I take justification, warrant and knowledge to be three of many forms of epistemic appraisal.
- 2 See Brewer (1999, p. 184) for a variation on this argument which uses the premise that the content of an experience must be endorsed in belief to provide epistemic support.
- 3 According to a further anti-foundationalist argument, if there were such a thing as basic perceptual support, it would be too easy for us to establish that our perception is reliable. I shall not consider this argument here as the main responses to it do not differentiate the traditional accounts of basic perceptual support from one another in new ways. For discussion, see Fumerton, 1995, ch. 6; Vogel, 2000; Cohen, 2002 and 2005; Bergmann, 2004; and Markie, 2005a.
- 4 Reid, p. 211.
- 5 Some philosophers claim that our development of an expert ability to "just see" that something is the case is insufficient for basic perceptual support. This appears to be Dretske's view (1993, p. 335). None the less, the

- question remains of how far basic perceptual support extends.
- 6 Another expectation deserves mention. Given that normative relations supervene on non-normative ones, a successful theory should explain basic perceptual support in non-normative terms.
  - 7 I shall not examine a third theory, phenomenalism, for lack of space. For discussion, see BonJour, 2001.
  - 8 For some background information on the concept of adverbial features of perceptual experience, see ADVERBIAL THEORY in Part III.
  - 9 See BonJour, 2001 and 2004; Dretske, 1993.
  - 10 For a discussion of the argument from illusion, see ARGUMENT FROM ILLUSION in Part III. See also Crane, 2005.
  - 11 BonJour, 2003.
  - 12 *ibid.*, p. 65 and p. 71.
  - 13 *ibid.*, p. 72.
  - 14 *ibid.*, p. 73.
  - 15 *ibid.*
  - 16 *ibid.*
  - 17 Chisholm, 1942. See the exchange between BonJour and Sosa in BonJour, 2003, esp. chs. 7, 10, and 11.
  - 18 BonJour, 2003, pp. 192–3.
  - 19 BonJour (2003, pp. 32–3) argues that it would be ad hoc to exempt perceptual beliefs about the external world from the reliability principle and subject them only to this lesser requirement. He thus continues the tradition of representational realism set by Descartes. According to this tradition, we need not know that our awareness of our mental states is reliable to know that things seem to us in a certain way, but we do need to know that we are not dreaming or being deceived to gain knowledge of the external world.
  - 20 Richard Fumerton (1995) develops a theory similar to BonJour's by appealing to a relation of direct acquaintance, which he explains as "a *sui generis* relation that holds between a self and a thing, property or fact" (p. 74). Fumerton's account is plagued by many of the same problems as BonJour's. See Sosa, 2003, pp. 129–40. For an additional criticism of both BonJour's and Fumerton's accounts, see Bergmann, 2006.
  - 21 See BonJour, 2003, ch. 5.
  - 22 BonJour admits that his strategy for gaining knowledge of the external world is not "in the minds of ordinary people when they make claims about the physical world (even though they are arguably aware in a less explicit and unified way of the relevant features of experience)" (2003, p. 96).
  - 23 See Goldman, 1976, esp. the final section; and Alston, 1993.
  - 24 Direct realism is often taken to also contain the metaphysical thesis that our perception of an external object does not involve our immediate awareness of a representation of that object, e.g. a sense datum. I shall concentrate on the epistemological thesis, which might be true even if the metaphysical thesis is false.
  - 25 Representational realists may, of course, object that direct realists are guilty of their own ad hoc move when they assert the following: whereas perceptual support does depend on our not having reason to think our perception is unreliable, it does not depend on our having reason to think that it is reliable. See BonJour, 2003, pp. 32–3. Direct realists, however, could rebut this criticism by pointing out the following: direct realism is consistent with the fact, revealed by a consideration of what we believe and why, that some of our perceptual beliefs have epistemic merit without being based on a belief in the reliability of our perception.
  - 26 For further discussion, see Fumerton, 1995, chs. 6 and 7; and Alston, 1999.
  - 27 Examples of direct realist views which deny premise (2) are considered below. For an example of a direct realist view that denies premise (3), see Brewer (1999) and the exchange between Fumerton (2001) and Brewer (2001).
  - 28 See Dretske, 1969; Goldman, 1986; and Plantinga, 1993b.
  - 29 In addition to the objections I consider here, some critics – Fumerton (1995, chs. 6 and 7) and Vogel (2000) – claim that externalist direct realism is open to the easy knowledge argument. See note 3. It should be noted, however, that this argument also challenges internalist theories.
  - 30 I've taken the example from Brewer (1999, pp. 102–4), who attributes it to Peacocke (1986, p. 154).
  - 31 Consider BonJour (2003, p. 174) and Alston (1999, p. 231) for similar points. BonJour (1985) claims that externalism abandons the traditional conception of knowledge.
  - 32 See Plantinga, 1993a.
  - 33 The standards for a successful analysis of knowledge are somewhat unclear and shifting. Goldman (1976, p. 791), for example, assumes that a successful theory will give us "a more accurate rendering of what the term 'know' actually means". Goldman (1979) later claims that a successful theory need not do so.

- 34 For further discussion of internalist direct realism, see BonJour, 2004; Markie, 2004 and 2005b.
- 35 Huemer, 2001, p. 99.
- 36 *ibid.*, pp. 99–103.
- 37 See Goldman, 1979.
- 38 If we embrace the background information argument, we can easily account for this case. Our perception that *a* is *F* defeasibly justifies the belief that *a* is *F* only if we are defeasibly justified in believing that this is how *F*'s appear to us. You meet this condition, but I don't.
- 39 The following view is developed in Markie, 2004 and 2006. For a similar approach, see Reynolds, 1991; and Pollock, 1986.
- 40 The goal of correct reasoning is not to form true beliefs rather than false ones. That is the goal of believing accurately, which is not the same as reasoning correctly. We may believe what is true through incorrect reasoning and what is false through correct reasoning.
- 41 This evidential support is defeasible. See Markie (2004) for an account of epistemic defeat when the evidence is an experience and the defeater is a proposition.
- 42 Is DPE truly an internalist account? Can we tell by reflection alone whether our epistemic norms are properly learned? I think so, though there's inadequate space to explore the issue here. To give an account of basic perceptual knowledge, we must add to DPE the condition that our epistemic norms are truth-reliable. DPE explains basic perceptual support in non-normative terms, assuming we can give a non-normative account of a non-defective process for learning epistemic norms.
- 43 It may be that, whenever we perceive something as *F*, we also have the background belief that what looks that way is *F*. It may even be that, whenever our perception epistemically supports our belief that something is *F*, this background belief has epistemic merit for us. My claim is that we need not include the background belief in our basis of believing that something is *F* in order for our perception to support our belief that something is *F*.
- 44 BonJour (2004, p. 365) objects that the support relation explicated by the practical approach is not epistemic, as it does not require that we be justified in believing that the experience is a reliable indicator of the belief's truth. Taken generally, this condition on the epistemic character of support relations contradicts his own account of how experiences epistemically support beliefs about our mental states. Markie (2004) and (2006) reply to additional objections.

Critics may charge that, since all our knowledge-how is actually knowledge-that (Stanley and Williamson, 2001), the background information argument is unanswered. I argue that knowledge-how is not a form of knowledge-that in "Knowing How Is Not Knowing That" (forthcoming).

- 45 Jon Kvanvig, Matthew McGrath, Matthias Steup and Paul Weirich have provided helpful comments on many of the ideas presented here.

## BIBLIOGRAPHY

- Alston, W.: *The Reliability of Sense Perception* (Ithaca, NY: Cornell University Press, 1993).
- Alston, W.: "Perceptual Knowledge," in *The Blackwell Guide to Epistemology*, ed. J. Greco and E. Sosa (Oxford: Blackwell, 1999).
- Bender, J.: *The Current State of Coherence Theory* (Boston, MA: Kluwer, 1989).
- Bergmann, M.: "Epistemic Circularity: Malignant and Benign," *Philosophy and Phenomenological Research* 3 (2004), 709–27.
- Bergmann, M.: "A Dilemma for Internalism," in *Knowledge and Reality: Essays in Honor of Alvin Plantinga*, ed. T. M. Crisp, Matthew Davidson and David Vander Laan (Secaucus, NJ: Springer, 2006).
- Bergmann, M.: *Justification without Awareness* (Oxford: Oxford University Press, 2006).
- BonJour, L.: *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1985).
- BonJour, L.: "Epistemological Problems of Perception," in *The Stanford Encyclopedia of Philosophy*, ed. E. Zalta, 2001 edn: <http://plato.stanford.edu>.
- BonJour, L.: "In Search of Direct Realism," *Philosophy and Phenomenological Research* 2 (2004), 349–67.
- BonJour, L. and Sosa, E.: *Epistemic Justification: Internalism vs. Externalism; Foundations vs. Virtues* (Oxford: Blackwell, 2003).
- Brewer, B.: *Perception and Reason* (Oxford: Oxford University Press, 1999).
- Brewer, B.: "Replies," *Philosophy and Phenomenological Research* 2 (2001), 449–64.
- Chisholm, R.: "The Problem of the Speckled Hen," *Mind* 51 (1942).
- Cohen, S.: "Basic Knowledge and the Problem of Easy Knowledge," *Philosophy and Phenomenological Research* 2 (2002), 309–29.
- Cohen, S.: "Why Basic Knowledge Is Easy Knowledge," *Philosophy and Phenomenological Research* 2 (2005), 417–30.
- Crane, T.: "The Problem of Perception," *Stanford Encyclopedia of Philosophy* (2005): <http://plato.stanford.edu>.



- Davidson, D.: "A Coherence Theory of Truth and Knowledge," in *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, ed. E. Lepore (Oxford: Blackwell, 1986).
- Dretske, F.: *Seeing and Knowing* (London: Routledge and Kegan Paul, 1969).
- Dretske, F.: "Perceptual Knowledge," in *A Companion to Epistemology*, ed. J. Dancy and E. Sosa (Oxford: Blackwell, 1993).
- Fumerton, R.: *Metaepistemology and Skepticism* (Lanham, MD: Rowman and Littlefield, 1995).
- Fumerton, R.: "Brewer, Direct Realism, and Acquaintance with Acquaintance," *Philosophy and Phenomenological Research* 2 (2001), 417–22.
- Goldman, A.: "Discrimination and Perceptual Knowledge," *Journal of Philosophy* (1976), 771–91.
- Goldman, A.: "What Is Justified Belief?," in *Justification and Knowledge*, ed. G. Pappas (Dordrecht: Reidel, 1979); also in *The Theory of Knowledge*, ed. L. Pojman (Belmont, CA: Thompson, 2003).
- Goldman, A.: *Epistemology and Cognition* (Cambridge, MA: Harvard University Press, 1986).
- Huemer, M.: *Skepticism and the Veil of Perception* (Lanham, MD: Rowman and Littlefield, 2001).
- Markie, P.: "Nondoxastic Perceptual Evidence," *Philosophy and Phenomenological Research* 3 (2004), 530–53.
- Markie, P.: "Easy Knowledge," *Philosophy and Phenomenological Research* (2005a), 406–17.
- Markie, P.: "The Mystery of Perceptual Justification," *Philosophical Studies* 3 (2005b), 347–73.
- Markie, P.: "Epistemically Appropriate Perceptual Belief," *Nous* 40 (2006), 118–42.
- Markie, P.: "Knowing How Isn't Knowing That," *Southwest Philosophical Review* (forthcoming).
- Peacocke, C.: *Thoughts* (Oxford: Oxford University Press, 1986).
- Plantinga, A.: *Warrant: The Current Debate* (Oxford: Oxford University Press, 1993a).
- Plantinga, A.: *Warrant and Proper Function*. (Oxford: Oxford University Press, 1993b).
- Pollock, J.: *Contemporary Theories of Knowledge* (Totowa, NJ: Rowman and Littlefield, 1986).
- Reid, T.: *An Inquiry into the Human Mind*, ed. T. Duggan (Chicago, IL: University of Chicago Press, 1970).
- Reynolds, S.: "Knowing How to Believe with Justification," *Philosophical Studies* 64 (1991), 273–92.
- Sosa, E. and Bonjour, L.: *Epistemic Justification: Internalism vs. Externalism; Foundations vs. Virtues* (Oxford: Blackwell, 2003).
- Stanley, J. and Williamson, T.: "Knowing How," *The Journal of Philosophy* 98 (2001).

- Steup, M.: *An Introduction to Contemporary Epistemology* (Upper Saddle River, NJ: Prentice Hall, 1996).
- Vogel, J.: "Reliabilism Leveled," *Journal of Philosophy* 97 (2000), 602–23.

PETER MARKIE

## Virtue Epistemology

The term "virtue epistemology" was first introduced into the literature by Ernest Sosa, who argued that epistemology might benefit by adopting an approach analogous to virtue theory in ethics.<sup>1</sup> By adopting a virtue-theoretic approach to justification and knowledge, Sosa argued, we can make progress on a variety of epistemological problems, including skepticism, the analysis of knowledge, the debate between foundationalism and coherentism, and the debate between internalism and externalism. Since Sosa's seminal work, epistemologists have defended a variety of positions and approaches that might properly be classified as "virtue epistemology". One important division among these concerns how we are to understand the notion of an intellectual virtue. Some epistemologists, including Sosa, understand the intellectual virtues as cognitive powers or abilities, such as accurate perception, reliable memory and sound reasoning (Sosa, 1991 and 2003a; Goldman, 1992; Kvanvig, 1992; Plantinga, 1993; Greco, 1999a and 2000a). Others understand the intellectual virtues as character traits more analogous to the moral virtues, such as intellectual courage, intellectual honesty, and fair-mindedness (Code, 1984 and 1987; Montmarquet, 1987 and 1993; Zagzebski, 1996). Another important division concerns the projects in which virtue epistemologists are engaged. Whereas some have argued that a virtue-theoretic approach helps to resolve traditional problems of epistemology, others have argued that the approach is best-suited to take epistemology in new and more interesting directions (Code, 1987; Kvanvig, 1992; Hookway, 1994 and 2003; Montmarquet, 1993; Zagzebski, 1996; Axtell, 1998; Dalmiya, 2001; McKinnon, 2003; Riggs, 2003).

Despite varied and deep diversity among virtue epistemologies, it is possible to identify some important common themes. This essay will focus on four of these. The final section of the essay reviews several objections that have been raised against virtue epistemology. Considering these will help to clarify both the nature and resources of a virtue-theoretic approach.

### 1. EPISTEMOLOGY AS A NORMATIVE DISCIPLINE

One theme that is common to any virtue epistemology is that epistemology is a normative discipline. When we say, for example, that someone knows something, or that some set of evidence is good, or that some belief is justified, we are making value judgments. We are, in some sense or other, making evaluative or normative judgments about persons, their evidence, their beliefs, etc. One task of epistemology is to investigate the sort of normativity that is involved in epistemic evaluations. By focusing inquiry here, some virtue epistemologists have argued, it is possible to make progress on perennial problems in epistemology, such as skepticism, the structure of justification, and the analysis of knowledge. Others have argued that a focus on epistemic normativity opens up new and interesting lines of investigation, such as the relations among epistemic and moral obligations, the moral and social dimensions of knowledge, and the ethics of inquiry.

### 2. VIRTUE-THEORETIC APPROACHES TO EPISTEMIC EVALUATION

One major task of epistemology is to investigate the sort of normativity involved in epistemic evaluations. In this regard, virtue epistemology looks to virtue theory in ethics for important insights and useful resources. The most common analogy that has been drawn concerns the direction of analysis among normative (or evaluative) concepts (Sosa, 1980 and 1991; Kvanvig, 1992; Zagzebski, 1996; Greco, 1999b). A defining character-

istic of virtue theories in ethics is that they take the normative properties of persons to be fundamental. A virtue theory then tries to define or explain other normative properties in terms of these. For example, it is common for virtue theories in ethics to define right action in terms of what a virtuous person would do, and to define good things in terms of what a virtuous person would desire. The analogous approach in epistemology takes the epistemically normative properties of persons as fundamental, and tries to define or explain such properties as justified belief and good evidence in terms of these. For example, a number of virtue epistemologists have defended the position that knowledge is true belief grounded in intellectual virtue (Sosa, 1991; Zagzebski, 1996; Lehrer, 2000; Riggs, 2002; Greco, 2004). Likewise, it has been suggested that good evidence and justified belief can be understood in terms of what an intellectually virtuous person would believe (Sosa, 1991; Montmarquet, 1993; Zagzebski, 1996; Greco, 2000a).

### 3. KNOWLEDGE AND CREDIT FOR TRUE BELIEF

A number of virtue epistemologists have endorsed the claim that knowledge is true belief produced by intellectual virtue. More specifically: In cases of knowledge, S has a true belief *because* S's belief is produced by intellectual virtue (Sosa, 1991; Zagzebski, 1996; Lehrer, 2000; Riggs, 2002; Greco, 2003b). This central idea has several attractive features.

First, the idea nicely organizes our pre-theoretical intuitions regarding which cases count as knowledge and which do not. For example, we think that true beliefs produced by good eyesight and sound reasoning, both of which are intellectual excellences, often qualify as knowledge, whereas true beliefs produced by lucky guesses and wishful thinking do not.

Second, the idea fits well with the widespread assumption that knowledge is non-accidentally true belief. On the present proposal, we have a straightforward way of



understanding the sense in which this is true. Specifically, on the current proposal knowledge is a kind of success through ability, and in general success through ability is no accident. Consider: If the archer is skilled, it is no accident that her arrow hits her target. If the painter is talented, it is no accident that his portrait resembles his subject. The current idea is that believing the truth is a kind of intellectual success. In cases of knowledge, this success is no accident in the sense that it is grounded in the knower's intellectual abilities.

Third, the current proposal nicely explains why knowledge is an achievement, and why knowledge attributions can function as credit attributions. Again, knowledge is a kind of success through ability, and in general success through ability is a kind of achievement for which the agent is properly credited. In cases of knowledge, it is to S's credit that S gets things right, in so far as S gets things right as the result of S's own abilities, as opposed to dumb luck or blind chance.

Finally, a virtue-theoretic account of knowledge and knowledge attributions nicely explains the value of knowledge. We turn to that issue next.

#### 4. THE VALUE OF KNOWLEDGE

In Plato's *Meno*, Socrates raises a question about the value of knowledge. Why is knowledge valuable? Or perhaps better: What explains the value of knowledge? In the course of his discussion, Socrates points out that knowledge has more than mere practical value. After all, a man who believes that *p* and who acts on that true information, is as well served in his purposes as the man who knows that *p* and who acts on that knowledge. The problem, then, is this: We think that knowledge has value over and above its practical value as true information. How do we explain that extra value?

Return now to the idea that knowledge is a kind of success through virtue. Put another way, knowledge is a kind of success through virtuous agency. This sort of success, we have seen, can be juxtaposed to mere lucky success: In the case of knowledge, S gets

things right *because* she is intellectually able and because she has exercised her abilities, and therefore her success is non-accidental in a familiar and relevant sense.

But now an answer to the value problem falls out of this account straightforwardly. In the *Nicomachean Ethics*, Aristotle makes a distinction between (a) achieving some end by luck or accident, and (b) achieving the end through the exercise of one's abilities (or virtues). It is only the latter kind of action, Aristotle argues, that is both intrinsically valuable and constitutive of human flourishing. "Human good", he writes, "turns out to be activity of soul exhibiting excellence" (*Nicomachean Ethics*, I.7). In this discussion Aristotle is clearly concerned with intellectual virtue as well as with moral virtue: his position is that the successful exercise of one's intellectual virtues is both intrinsically good and constitutive of human flourishing. If this is correct, then there is a clear difference in value between knowledge and mere true belief. In cases of knowledge, we achieve the truth through the exercise of our own intellectual abilities, which are a kind of intellectual virtue (Zagzebski, 1996, 1999, 2003 and 2004; Riggs, 2002; Greco, 2003b; Sosa, 2003b and 2005; see also Kvanvig, 2004).

#### 5. OBJECTIONS

A number of objections have been raised against a virtue-theoretic approach in epistemology. In this final section of the essay we shall consider some of the most important and persistent of these.

##### *a. Belief is not under voluntary control*

Some philosophers have objected to a virtue-theoretic approach in epistemology because, according to them, belief is not under voluntary control. Their line of reasoning is as follows.

1. Unlike actions, beliefs are not typically under one's voluntary control.
2. Virtue-theoretic evaluations apply only where there is voluntary control.

Therefore

3. Virtue-theoretic evaluations do not typically apply to beliefs.

Some virtue epistemologists have responded by challenging the first premise of the objection, arguing that our beliefs and cognitive dispositions are sufficiently under our control to warrant virtue-theoretic evaluations. A common version of this response is to distinguish between direct and indirect control, and to argue that the relevant evaluations require only the latter, which our beliefs enjoy. That is, even if I cannot directly control my believing that *p* now, I can control my efforts to think more carefully on the issue, gather more evidence, etc., thereby exercising a sort of indirect control over my belief. More generally, the strategy here is to argue that virtue-theoretic evaluations require only a weak kind of control, and that cognition is often under that weaker kind of control. A second part of this strategy is to emphasize the analogy between belief and action in this respect. That is, we think that virtue-theoretic evaluations often apply to actions that are not directly or otherwise strongly under our control. Such evaluations are deemed appropriate so long as the actions in question are under some weaker kind of control, often shared by beliefs as well (Montmarquet, 1987 and 1993; Zagzebski, 1996).

Other virtue epistemologists have challenged premise 2 of the present objection; i.e. that virtue-theoretic evaluations require voluntary control of any sort. This strategy follows Aristotle in making a distinction between moral virtues and intellectual virtues. Whereas moral virtues and the actions that flow from them essentially involve choice, the intellectual virtues and their acts do not. Likewise, moral evaluations are of a special sort, essentially implying some relation of control between the moral agent and the object of moral evaluation. Cognitive evaluation implies no such thing, although control of various sorts may in fact be present. In this respect, intellectual excellences are more like athletic excellences than moral excellences. Thus we praise an

athlete's skill, and give her credit for her achievements, without regard to whether her virtues and accomplishments involve choice. For example, a soccer goalie is credited with a spectacular save even if it results more from instinct and reflex than from deliberation and choice. The point is not that deliberation and choice cannot be present in athletic or cognitive achievements, but that they need not be for the relevant sorts of evaluation to be appropriate (Greco, 2000b and 2002).

#### *b. Natural faculties are not acquired habits*

Some philosophers have argued that the language of virtue is inappropriate in many paradigm cases of knowledge, such as perceptual knowledge and memory knowledge, since these arise from our natural faculties rather than as acquired habits, as virtues are thought to be. Once again, virtue epistemologists have replied by challenging both assumptions behind the objection. First, some virtue epistemologists point out that natural faculties such as perception and memory are developed and shaped over time, so that it is false to say that these faculties are merely natural and not acquired. Here again the analogy to action is emphasized: in paradigmatic cases of moral evaluation, our actions result from faculties and dispositions that are both naturally endowed and developed over time (Sosa, 1991 and 2004).

Other virtue epistemologists have challenged the assumption that intellectual faculties must be acquired to count as virtues. Again, the point is not that such faculties are never developed or acquired over time, but that they need not be to count as virtues and to be involved in the relevant sorts of evaluations. This second strategy for replying to the objection continues to insist on a distinction between moral virtues and intellectual virtues. Moral virtues, it is argued, must be acquired because they must result from voluntary actions to count as moral virtues in the first place. Intellectual virtues carry no such constitutive requirement, and therefore natural faculties can count as intellectual virtues. For example, mathematical acumen counts as an intellectual excellence whether

it is earned by hard study or is a God-given ability (Greco, 2000b and 2002).

*c. Knowledge does not require stability*

Above we saw that virtue epistemologists disagree about the nature of the intellectual virtues. Some understand these as character traits analogous to the moral virtues, while others understand the intellectual virtues as abilities or powers, more akin to athletic excellences than to excellences of moral character. On either understanding, however, the intellectual virtues are characterized as stable dispositions of the cognitive agent rather than as fleeting properties. This feature of the view is important for addressing certain counter-examples to reliabilism involving strange and fleeting processes. For example, a virtue-theoretic approach can agree that a true belief is not knowledge even if caused by a highly reliable but fleeting glitch in cognitive processing. The processes that give rise to knowledge must be grounded in the intellectual virtues of the knower, and these are by nature stable rather than fleeting.

Some critics have objected that this makes a virtue-theoretic approach too strong, however. For not all cases of knowledge, the objection goes, are grounded in stable dispositions of the knower (Cohen, 2003; Kvanvig, 2003; Lahroodi and Schmitt, 2003). For example, suppose that someone has excellent vision but also suffers from a degenerative ocular disease. Such a person's visual abilities are not stable, the objection goes, but this does not prevent him from having visual knowledge now. Or consider the condition of some Alzheimer's patients, who at times are lost in confusion but who at other times are perfectly lucid and cogent. The present objection, then, is that such victims lack stable dispositions or abilities. Nevertheless, it is false that such victims lack knowledge during times when they are lucid. For example, they have memory knowledge during times when their memory is functioning properly.

In order to answer this objection, it is important to be clear about what it means for a disposition or ability to be stable. The first thing to note is that dispositions in general and

abilities in particular cannot be understood merely in terms of what happens in the actual world. Rather, they must be understood in terms of what happens across relevantly close possible worlds. Let us focus on abilities. A person does not have an ability to hit baseballs just because he hits the ball in the actual world. After all, the person's success might be just a matter of good luck. The person does have the ability, however, if he is successful over a range of relevantly close possible worlds. When the pitch comes in a little lower, for example, a good hitter adjusts his swing and still hits the ball. Likewise, it does not count against a person's having an ability if the person is unsuccessful in worlds that are not relevantly close. Thus, it does not count against a baseball player's abilities that he would not hit the ball if it were thrown at two hundred miles per hour, or if he were drunk, or if it was dark. In general,

S has an ability to achieve result R in conditions C = across the range of (relevantly) close possible worlds where S is in C, S achieves R in C with a high rate of success.

The second thing to note is that the phrases "stable disposition" and "stable ability" are redundant. That is, dispositions in general and abilities in particular are by nature stable – they signify stable rather than fleeting properties of the agent who has them. Nevertheless, stability comes in degrees. On the present characterization, the stability of a disposition is a function of what the person does, or how the person acts, across the space of relevantly close possible worlds. The idea is that the further out into logical space the relevant behavior persists, the more stable the disposition. Thus, if the world must be almost exactly as it is for the person to act in accordance with the disposition, then the disposition in question is not very stable. If the person acts in accordance with the disposition even if the world is very different, then the disposition is very stable. It is important to remember, however, that dispositions are a function of what happens in worlds that are *relevantly close*. Again, whether a baseball player now has an ability to hit baseballs has

little to do with what he would do if he were blind or drunk. Neither would we fail to credit a healthy and sober player with a skillful hit, even if he were often drunk or soon to be blind.

We are now in a position to reply to the present line of objection. In sum, the requirement that knowledge be grounded in stable dispositions of the knower does not imply a requirement that the knower have the same dispositions through time. For example, attributing perceptual knowledge to S does not imply that S's perceptual abilities will not degenerate in the near future. Likewise, attributing memory knowledge to S does not imply that S's ability to remember will not come and go. That is because dispositions in general, and cognitive abilities in particular, must be understood in terms of what a person does in relevantly close possible worlds. But worlds where S's abilities are no longer operative, or where they are absent altogether, are not relevantly close.

To make the point a different way, there are two senses in which a disposition or ability might be stable: there is (a) stability through the logical space of close possible worlds, and (b) stability through time. According to a virtue-theoretic account, knowledge requires stability in the first sense rather than the second. These two kinds of stability often coincide, since the next moment in time is often close in logical space as well. The exception is when relevant changes occur over time. In that case, nearness in time does not imply nearness in logical space (Greco, 2003a).

*d. Not all knowledge is an achievement: the case of perceptual knowledge*

Recall the account of knowledge reviewed above: that knowledge is true belief resulting from intellectual virtue, and therefore a kind of intellectual achievement creditable to the knower. One objection to this position is that not all knowledge counts as an intellectual achievement. For example, my perceptual knowledge that there is a yellow school bus parked outside is no great achievement on my part – all I have to do is look out of my window, and I can't help but to see it there. It is

false, therefore, that all knowledge is an achievement for which the knower deserves credit. In the typical case, for example, perceptual knowledge is not.

In reply to this objection, it is important to distinguish between two kinds of credit for a success. In the first sense, to credit someone for a success is to praise them for it. In the second sense, to credit someone for a success is to attribute it to them – it is to say that it is their success rather than someone else's, or not a success at all. The first sort of credit implies praiseworthiness, whereas the second sort of credit implies ownership. Notice that the latter sort of credit does not imply the former. Specifically, S's success might be no big deal, and therefore not worthy of praise in any substantial sense.

To illustrate, consider two ways in which a person can travel across a room: one can walk across the room under one's own power, or one can stumble across it by accident. Getting across the room might constitute no great success, and yet there is a difference between doing so under one's own power and simply stumbling or falling across the room. In the case of walking across the room we have a success that is attributable to the agent. It is in that sense that the success can be credited to her.

And now the answer to the present objection is straightforward: on a virtue-theoretic approach, knowledge requires credit in the sense of ownership rather than credit in the sense of praiseworthiness. In cases of knowledge, S's believing the truth is creditable to her rather than to dumb luck, or blind chance, or something else. Of course, successfully believing the truth might be no big deal, no great achievement. Nevertheless, it is creditable to the agent, in the relevant sense, so long as it is a success grounded in the agent's own abilities.

*e. Not all knowledge is an achievement of the knower: the case of testimonial knowledge*

The last objection to virtue epistemology we shall consider charges that knowledge is not always an achievement that is creditable to the knower. That is, even if we understand credit

in the ownership sense, it is false that in all cases of knowledge the knower can be credited with the success. In particular, it is often the case that testimonial knowledge is not to the knower's credit (Lackey, 2004 and 2006). For example, consider the following case from Jennifer Lackey.

Having just arrived at the train station in Chicago, Morris wishes to obtain directions to the Sears Tower. He looks around, randomly approaches the first passerby that he sees, and asks how to get to his desired destination. The passerby, who happens to be a Chicago resident who knows the city extraordinarily well, provides Morris with impeccable directions to the Sears Tower.

Lackey writes:

What explains why Morris got things right has nearly nothing of epistemic interest to do with him and nearly everything of epistemic interest to do with the passerby. In particular, it is the passerby's experience with and knowledge of the city of Chicago that explains why Morris ended up with a true belief rather than a false belief. . . . Thus, though it is plausible to say that Morris acquired knowledge from the passerby, there seems to be no substantive sense in which Morris deserves credit for holding the true belief that he does. (Lackey, 2004)

This objection raises important issues about the epistemology of testimony in general. Often, theories of testimonial knowledge are divided into two camps, according to how they understand the requirement that the source of testimony be reliable. On the first kind of theory, what is important for testimonial knowledge is that the source of testimony is in fact reliable. On the second kind of theory, it is also important that the believer knows, or at least justifiably believes, that the source is reliable. From a virtue-theoretic perspective, however, a third kind of theory is plausible. Namely, testimonial knowledge requires that the *believer* is a reliable *receiver* of testimony. That is, what is important is not so much that the testifier is reliable, or that the believer know that he is, but that the believer herself is reliable in the way that she receives and evaluates testimony. This will plausibly

involve reliable capacities for discriminating reliable sources of testimony from unreliable sources.

Suppose that this third approach is correct. Then we have to divide Lackey's example into two cases: one where Morris is a reliable receiver of testimony and one where he is not. From the perspective of a virtue theory, it is only in the first sort of case that Morris knows the location of his destination. But in that sort of case it is also to Morris's credit that he forms a true belief. That is, his success is grounded in his ability to discriminate good from bad testimony and is therefore attributable to him.

#### NOTE

- 1 The earliest work in which Sosa defends this claim is (Sosa, 1980). The term "virtue epistemology" first appears in the introduction of (Sosa, 1991).

#### BIBLIOGRAPHY

- Axtell, G.: "The Role of the Intellectual Virtues in the Reunification of Epistemology," *The Monist* 81 (1998), 488–508.
- Code, L.: "Toward a 'Responsibilist' Epistemology," *Philosophy and Phenomenological Research* 45 (1984), 29–50.
- Code, L.: *Epistemic Responsibility* (Hanover, NH: University Press of New England and Brown University Press, 1987).
- Cohen, S.: "Greco's Agent Reliabilism," *Philosophy and Phenomenological Research* 66 (2003), 437–43.
- Dalmiya, V.: "Knowing People," in *Knowledge, Truth and Duty: Essays on Epistemic Justification, Responsibility and Virtue*, ed. M. Steup (Oxford: Oxford University Press, 2001).
- DePaul, M. and Zagzebski, L., eds: *Intellectual Virtue: Perspectives from Ethics and Epistemology* (Oxford: Oxford University Press, 2003).
- Goldman, A.: "Epistemic Folkways and Scientific Epistemology," in *Liaisons: Philosophy Meets the Cognitive and Social Sciences*, ed. A. Goldman (Cambridge, MA: MIT Press, 1992).
- Greco, J.: "Agent Reliabilism," *Philosophical Perspectives* (1999a), 273–96.
- Greco, J.: "Virtue Epistemology," in *The Stanford Encyclopedia of Philosophy*, ed. E. Zalta (1999b).



- Greco, J.: *Putting Skeptics in Their Place*. (Cambridge, MA: Cambridge University Press, 2000a).
- Greco, J.: "Two Kinds of Intellectual Virtue," *Philosophy and Phenomenological Research* 60 (2000b), 179–84.
- Greco, J.: "Virtues in Epistemology," in *The Oxford Handbook of Epistemology*, ed. P. Moser (Oxford: Oxford University Press, 2002).
- Greco, J.: "Further Thoughts on Agent Reliabilism," *Philosophy and Phenomenological Research* 66 (2003a), 466–80.
- Greco, J.: "Knowledge as Credit for True Belief" (2003b), in DePaul and Zagzebski, 2003.
- Greco, J. (ed.): *Sosa and His Critics*. (Oxford: Blackwell, 2004).
- Hookway, C.: "Cognitive Virtues and Epistemic Evaluations," *International Journal of Philosophical Studies* 2 (1994), 211–27.
- Hookway, C.: "How to Be a Virtue Epistemologist," in DePaul and Zagzebski, 2003.
- Kvanvig, J.: *The Intellectual Virtues and the Life of the Mind* (Savage, MD: Rowman and Littlefield, 1992).
- Kvanvig, J.: "Simple Reliabilism and Agent Reliabilism," *Philosophy and Phenomenological Research* 66 (2003), 451–6.
- Kvanvig, J.: *The Value of Knowledge and the Pursuit of Understanding* (Oxford: Oxford University Press, 2005).
- Lackey, J.: Review of DePaul, M. and Zagzebski, L. (eds) *Intellectual Virtue: Perspectives from Ethics and Epistemology* (Oxford: Oxford University Press, 2003) in *Notre Dame Philosophical Reviews* 08.01 (2004).
- Lackey, J.: "Why We Don't Deserve Credit for Everything We Know," *Synthese* 2006.
- Lahroodi, R. and Schmitt, F.: "Comment on John Greco's Putting Skeptics in Their Place," *Philosophy and Phenomenological Research* 66 (2003), 457–65.
- Lehrer, K.: *Theory of Knowledge*, 2nd edn (Boulder, Colo.: Westview Press, 2000).
- McKinnon, C.: "Knowing Cognitive Selves," in DePaul and Zagzebski, 2003.
- Montmarquet, J.: "Epistemic Virtue," *Mind* 96 (1987), 482–97.
- Montmarquet, J.: *Epistemic Virtue and Doxastic Responsibility* (Lanham, MD: Rowman and Littlefield, 1993).
- Plantinga, A.: *Warrant and Proper Function* (Oxford: Oxford University Press, 1993).
- Riggs, W.: "Reliability and the Value of Knowledge," *Philosophy and Phenomenological Research* 64 (2002), 79–96.
- Riggs, W.: "Understanding Virtue and the Virtue of Understanding," in DePaul and Zagzebski, 2003.
- Sosa, E.: "The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge," *Midwest Studies in Philosophy* (1980), 3–25. Reprinted in Sosa, 1991.
- Sosa, E.: *Knowledge in Perspective* (Cambridge: Cambridge University Press, 1991).
- Sosa, E.: "Beyond Internal Foundations to External Virtues" (2003a), in *Epistemic Justification: Internalism vs. Externalism; Foundations vs. Virtues*, ed. L. BonJour and E. Sosa (Oxford: Blackwell, 2003).
- Sosa, E.: "The Place of Truth in Epistemology" (2003b), in DePaul and Zagzebski, 2003.
- Sosa, E.: "Replies," in Greco, 2004.
- Zagzebski, L.: *Virtues of the Mind*. (Cambridge: Cambridge University Press, 1996).
- Zagzebski, L.: "From Reliabilism to Virtue Epistemology," *Proceedings of the Twentieth World Congress of Philosophy*, vol. 5, *Epistemology* (1999), 173–9.
- Zagzebski, L.: "Intellectual Motivation and the Good of Truth," in DePaul and Zagzebski, 2003.
- Zagzebski, L.: "Epistemic Value Monism," in Greco, 2004.

JOHN GRECO

## Social Epistemology

Social epistemology is an approach to epistemology, or branch of epistemology, that explores the social aspects of the knowledge enterprise. Traditional epistemology has a highly individualistic flavor. In the Cartesian mold, it considers what an isolated subject can know or be justified in believing entirely on her own, where her resources, at least at the outset, are confined to the contents of her mind. There is no prospect of help from anybody else, because the very existence of anybody else is problematic. Social epistemology, by contrast, examines the social and interpersonal dimensions of knowledge. Most of our intellectual enterprises are interwoven with the intellectual enterprises of others, and this should be reflected in at least one part of epistemology.

## THREE CONCEPTIONS OF SOCIAL EPISTEMOLOGY

Three different conceptions of social epistemology can be distinguished, each with its



own idea of how the project should be undertaken (Goldman, forthcoming a). The first conception, *revisionism*, holds that the traditional conception of epistemology is fundamentally misguided and needs to be replaced by a thoroughly socialized successor. Proponents of this conception include both social scientists and philosophers. They argue that empirical studies of, or philosophical reflections on, human intellectual enterprises show that the traditional paradigm of epistemology cannot be sustained. A new and highly socialized discipline is required.

A second conception of social epistemology finds nothing fundamentally wrong with the core of traditional epistemology; it is simply incomplete. Traditionally epistemology occupied itself with those parts or aspects of the quest for knowledge and justification that can be undertaken by an unaided epistemic agent. However, there is also ample space within traditional epistemology for an epistemic agent to obtain evidence from other agents: evidence of what they say and what they think. These kinds of social evidence can certainly be relevant to what an agent is justified in believing, what it is rational for her to believe, and what she knows. Two prime examples of such topics are testimony and peer disagreement (see below). Once these topics are added to traditional epistemology's agenda, all of the appropriate social elements will be on its plate. This approach to social epistemology might be called *preservationism*. It rejects the necessity of any major remodeling of epistemology's home office, and simply proposes to open a few new branches, a task essentially accomplished in the last few decades.

A third conception of social epistemology is less radical than revisionism but less conservative than preservationism. Like preservationism, it accepts the principal presuppositions of traditional epistemology but proposes to expand the social frontier beyond the topics of testimony and peer disagreement. A broader agenda for social epistemology would include these two topics but go beyond them. This approach to social epistemology might be called *expansionism* (Goldman, forthcoming a; forthcoming b).

#### REVISIONIST SOCIAL EPISTEMOLOGY

A thumbnail portrait of traditional epistemology (TE) might list the following tenets. First, the epistemic agents studied by TE – at least, the *primary* epistemic agents – are individual human beings. Second, TE seeks to analyze various types of epistemic desiderata, e.g. justification (warrant), rationality, and knowledge, and to determine suitable methods by which such desiderata can be realized or attained. Third, the normative standards of justification and rationality are not merely conventional or relativistic in a pejorative sense, but have universal, objective validity. Fourth, the central types of epistemic desiderata either entail or are linked to truth, and truth is an objective affair. Let us call these assumptions the *fundamental* assumptions of TE.

Revisionist brands of social epistemology reject many if not all of these assumptions. One brand of revisionist social epistemology is social constructivism, a radical version of which holds that truths or facts are not in or of the world. Facts are not “out there” to be discovered and revealed by investigators, but are the creations or fabrications of investigators (Latour and Woolgar, 1986). Social constructivists often claim that this is shown by empirical studies in the sociology of science – for example, studies of what transpires in a neuroendocrinology laboratory. “[W]e do not conceive of scientists . . . as pulling back the curtain on pre-given, but hitherto concealed truths. Rather, objects (in this case, substances) are constituted through the artful creativity of scientists” (Latour and Woolgar, 1986: 128–9). Similarly, as Steven Shapin puts it, “truth is a social institution” (1994: 6). In the lexicon of these writers, knowledge does not entail truth. Knowledge is simply whatever is believed, or perhaps “institutionalized” belief.

Another tenet of TE, the objectivity of rationality, has been challenged by members of the “strong programme” in the sociology of science. Barry Barnes and David Bloor declare that “there are no context-free or super-cultural norms of rationality” (1982: 27). In other words, there are no absolute facts about what makes a belief rational, or

what evidence justifies what hypotheses. Thomas Kuhn (1962) was an influential figure in the history of science who debunked the possibility of settling hardcore scientific disputes by purely objective considerations. Kuhn favored “social factors” as the non-rational causes or explanations of the retention or change in scientific paradigms. All these writers may be considered (revisionary) social epistemologists, although they themselves did not employ this phrase.

Mainstream epistemologists tend to pooh-pooh the views of the foregoing writers. Let us look more closely, therefore, at an argument for epistemic relativism that is more congenial to proponents of TE. This argument is formulated by Paul Boghossian (2006), based on discussions of Richard Rorty (1979) and Wittgenstein (1969). The first premise of the argument says that if there were absolute facts about justification it ought to be possible to arrive at justified beliefs about them. The second premise says that it isn't possible to arrive at justified beliefs about such justificational facts. What supports the second premise? Rorty contends that any justification about facts of justification would have to be (epistemically) circular. He cites the historical dispute between Galileo and his Vatican antagonist Cardinal Bellarmine, who refused to look through Galileo's telescope to get evidence about the structure of the heavens. Bellarmine contended that he already had better evidence about this matter – namely, scriptural evidence – than the telescopic evidence Galileo offered. Galileo and Bellarmine could each try to justify their beliefs by reference to their preferred principles of evidence. But how could they justify these preferred principles? Wouldn't each have to invoke the self-same principles, and wouldn't this be objectionably circular?

One possible response to this argument for relativism is to deny that justification of (putative) justificational facts has to be circular. Couldn't Cardinal Bellarmine (try to) justify a scriptural principle of evidence by appeal to past observations of miracles? The strength of such a justification is debatable, but it isn't obviously circular. A second response is that epistemic circularity is unavoidable

but doesn't necessarily undercut objectivist, non-relativist justification (Bergmann, 2004; Van Cleve, 2003). A third and different response to the argument is to concede the truth of relativism but deny that it leads to any distinctive, revisionist project for social epistemology. Many relativists are social scientists whose chief project is to study scientifically either science itself or other communities of inquiry. If epistemological relativism is correct, doesn't this undermine the relativists' project? Why study these matters *scientifically* if science has no claim to epistemic superiority over other methodologies? Rorty is more consistent in this respect, taking his relativism to support the “end of epistemology”.

A rather different basis for relativism is the dialectical approach to justification, advocated by writers like Rorty, David Annis (1978), Michael Williams (1999), and Martin Kusch (2002). On this approach, what it is for a person's belief or statement to be justified is a matter of its reception, or standing, within society, or among one's peers. Only if it a person's belief or statement is capable of surviving (actual or potential) objections from the person's peers is it justified. This avowedly social conception of justification is at odds with most parts of TE, although it borders on forms of contextualism that have recently been in fashion in mainstream epistemology. However, it faces two serious problems. First, what counts as a person's peers or society? Indefinitely many circles of peers or societies might, in principle, react to a person's statement. It might be acceptable in some circles but not in others. Second, what if the subject has private knowledge he cannot successfully communicate to his peers, or cannot persuade them of because of an undeservedly bad reputation among them (as a testifier). Even if his public reputation is irreparably undercut, can't he justifiably believe something about his own current thoughts or his memories of past actions, despite the inability to persuade his peers?

There are other kinds of revisionist proposals for social epistemology we haven't yet mentioned. For example, whereas individuals are the standard epistemic agents within TE, some social epistemologists propose that

knowers consist only of groups or communities, not individuals. Lynn Hankinson Nelson writes: “it is communities that construct and acquire knowledge . . . [T]he collaborators, the consensus achievers, and, in more general terms, the agents who generate knowledge are communities and subcommunities, not individuals” (1993: 124). A weaker form of communitarianism merely asserts that only within communities can epistemic norms exist. This communitarian thesis is endorsed by Martin Kusch (2002: 175), and is a variant of Wittgenstein’s (1968) argument against private languages. The idea is that socially isolated individuals are unable to generate normative phenomena. This is compatible, of course, with the thesis that individuals are the types of entities that conform or fail to conform to epistemic norms. This would be the main precept of traditional epistemology as far as epistemic agency is concerned.

#### PRESERVATIONIST SOCIAL EPISTEMOLOGY

We turn now to preservationist social epistemology and focus on its explorations of social dimensions of knowledge that stay within TE’s guidelines. Here we review the two topics previously mentioned – testimony and peer disagreement – that are widely accepted parts of the mainstream.

The problem of testimony is a problem about warrant or justification (see TESTIMONY in Part III). Presumably there are circumstances in which hearers are justified in believing propositions that speakers assert. What circumstances are these, exactly? This can be viewed as the core problem of social epistemology, “core” because the testimonial situation is a minimal case in which putative information is offered by one epistemic agent to another. It is minimal because only two people are involved and the communication consists of a single assertion.

Hume denied that testimony is a *basic* source of evidence or justification. Testimony-based justification must be derived from other sources of evidence – namely, perception,

memory, and inductive inference. Hume wrote: “The reason why we place any credit in witnesses and historians, is not derived from any *connexion*, which we perceive *a priori*, between testimony and reality, but because we are accustomed to find a conformity between them” (1957: 113). Hume’s view of testimonial justification is nowadays called *reductionism*, a view that can be spelled out in two ways. One interpretation is *global* reductionism. This says that a hearer is justified in accepting testimony from another if and only if she has non-testimonially based positive reasons for believing that testimony is *generally* reliable (truth-conducive), i.e. reliable across all speakers and all specimens of testimony. A second interpretation is *local* reductionism. It says that a hearer is justified in accepting testimony from a given speaker only if she has non-testimonially based positive reasons for believing that the speaker’s *present* testimony is reliable (E. Fricker, 1995).

A central worry about reductionism of either variety is that many epistemic agents – especially young children – are unable to satisfy its requirements. Young children presumably acquire lots of knowledge and justified belief from their mothers. Do these children satisfy the reductionist requirements for justified testimony-based belief? Does a young child (say, a 2-year-old) have positive reasons for believing in the *general* reliability of testimony? Most young children are exposed to a tiny sample of the world’s population, and one does not have the opportunity to check independently the reliability of many of their statements. Does the child have positive, non-testimonially based reasons for believing that even her mother is reliable most of the time? That, too, is questionable. These kinds of considerations led C. A. J. Coady (1992) to offer a general critique of reductionism about testimony. Coady specifically criticized Hume, challenging his idea that there is merely a contingent general correlation between truth and testimony, as opposed to a more intimate connection.

Recognizing the foregoing problems, many epistemologists opt for non-reductionism about testimony. Tyler Burge (1993), for example, holds that a person is *prima facie*

entitled to accept another's report (assertion) as long as he has no *defeating* evidence against it. In other words, a hearer doesn't need any independent positive evidence either for the general reliability of testimony or for the reliability of the particular speaker (on the present occasion). Unless the testimony is defeated by independent evidence, the hearer is justified in believing it.

As stated thus far, both views seem perfectly compatible with TE's tendency to view all epistemic activity as the making of doxastic "choices" by individual agents. Reductionism and non-reductionism tell different stories about what must be true of a hearer for her to be justified in forming a belief in someone else's report. Still, the only justification-maker (or un-maker) is the evidence possessed by the hearer, which encompasses evidence about other people and their utterances. The story conforms straightforwardly to the individualistic basis of TE, while admitting social evidence in the form of evidence *about* other people.

One approach to testimony, however, adds a further twist, taking the testimony scenario slightly beyond the strict orbit of TE. This approach invokes *transindividual* properties of testimony, in Frederick Schmitt's (2006) terminology. The basic idea is that both testimonial knowledge and testimonial justification are *social* in a fairly strong sense. Schmitt says that something in the vicinity of the following thesis holds for testimonial knowledge: "I know that *p* on *T*'s testimony that *p* only if *T* knows that *p*." This thesis makes testimonial knowledge social in the sense that my having testimonial knowledge entails that there is knowledge belonging to an individual other than myself. A similar thesis (with substantial qualifications) is said to hold for testimony-based justification. If the transindividual thesis is correct, we move more deeply into social territory, where one agent's justifiedness depends on the *actual* justifiedness of the testifying agent, not simply on the perceptions, memories, and prior beliefs of the hearer.

The problem of reasonable disagreement among peers can be introduced as follows. Suppose Marvin believes proposition *Q* while

Melvin rejects it. In short, they disagree. Further suppose that Marvin and Melvin each share with one other all the evidence in their possession relevant to *Q*. Finally, suppose that each regards the other as a "peer" in the sense of having both equal evidence and equal cognitive powers, such as intelligence, carefulness, fair-mindedness, and the like. None the less, after sharing their evidence and learning of the other's opposing opinion, both Marvin and Melvin "stick to their guns" on the question of *Q*. Each maintains his original attitude. Could they both be reasonable? Or do these facts imply that at least one of them has been unreasonable? If there is an objectively correct set of principles of justification – one of the fundamental assumptions of TE – how can a persisting disagreement between Marvin and Melvin be reasonable? Shouldn't they "split the difference" with one other and converge on the same attitude? This "conformist" position is defended by Richard Feldman (2007), David Christensen (2007), and Adam Elga (2007), among others.

Nonconformists, on the other hand, think that the foregoing analysis misses an important element. If there is a uniquely correct set of justificational rules, then prior to learning of the peer's different opinion one party to the disagreement may hold a justificational appropriate attitude whereas the other party does not. If so, why should the justified party be obliged to converge towards the unjustified party simply because she learns that the latter party disagrees? She should stick to her guns. This is roughly how "non-conformists" argue, including Peter van Inwagen (1996), Thomas Kelly (2005), Jennifer Lackey (forthcoming), and Ernest Sosa (forthcoming).

We don't have to settle this dispute for present purposes. Whether reasonability always requires peers to converge in doxastic attitudes, the situation with respect to social epistemology is the same. Having evidence about a peer's attitude towards *Q* can sometimes be *relevant* to the reasonability or justifiability of one's own attitude, even if it isn't decisive. Evidence about a peer's attitude – social evidence – is sometimes relevant. But this doesn't carry us beyond the

realm of TE, because the social evidence that helps determine reasonability is still possessed by each individual. This requires no fundamental change in TE's game.

#### EXPANSIONIST SOCIAL EPISTEMOLOGY: THE VERITISTIC VARIANT

In its original Cartesian mold, TE contemplated a single epistemic agent engaged in one activity of interest: *doxastic decision-making*. Doxastic decision-making is, roughly, the adoption (or retention) of doxastic attitudes towards propositions based on perceptions, memories, and prior beliefs. (The phrase "decision-making" is not intended to imply that the adoption of doxastic attitudes is a voluntary affair.) The attitudes can be classified in either a trichotomous fashion – belief, suspension, disbelief – or in a graded fashion – degrees of credence, or subjective probabilities. The preservationist approach to social epistemology is social in so far as it introduces into the scenario at least one additional human agent, someone who interacts with the doxastic decision-maker (DDM) by asserting something or revealing his opinions. Despite the introduction of this additional agent, the normative or evaluative enterprise of preservationist social epistemology still focuses on doxastic decision-making. What is evaluated is either the doxastic attitude adopted or the decision-making process that generates it. The attitude would typically be evaluated as justified or unjustified and the process might be evaluated as rational or less-than-rational. Preservationist social epistemology shows little interest in evaluating the speaker's activity.

Expansionist social epistemology, by contrast, is prepared to expand the list of activities or practices subject to epistemic evaluation. This is not the only proposed point of departure by expansionism, but it's a significant one. Even within current mainstream epistemology, a bit of attention is paid to norms of speech, especially the norm of assertion. It is generally agreed that people aren't allowed to assert just anything under any circumstances. It is inappropriate to

assert a proposition one believes to be false, or assert a proposition for which one lacks adequate evidence. Grice (1980) posits conversational maxims that prohibit assertions under such circumstances.

Epistemologists debate exactly which standard is the appropriate norm of assertion, but a popular view is that knowledge is the norm of assertion: one should assert that *p* only if one knows that *p* (Williamson, 2000; DeRose, 2002; Hawthorne, 2004; Stanley, 2005). The knowledge norm of assertion has not gone without its critics. Matthew Weiner (2005) endorses the truth norm of assertion, and Jennifer Lackey (2008) defends the reasonable belief norm of assertion. Each of these norms, however, is some sort of epistemic norm of speech. So even epistemologists who don't represent themselves as departing from TE (in the sense that interests us) endorse theses about epistemically evaluable activities that go beyond purely doxastic activity.

Going beyond assertion plain and simple, Goldman (1994) considers a more complex type of speech activity: argumentation. In argumentation a speaker typically endorses a certain conclusion on the basis of a set of asserted premises. Goldman posits a bevy of implicit rules that govern this kind of speech activity, and argues that they are based on expected epistemic consequences for the conversational participants. These rules are not doxastic decision-making rules, however, so they concern a different type of activity from that which TE highlights.

Still other types of speech-related activity may be subject to epistemically driven norms. Consider legal rules governing the admission or exclusion of evidence in court. Many of these rules are rationalized by theories about the impact that various types of evidence would have on jurors and consequently on the risk of producing inaccurate verdicts. The fear is that various kinds of testimony could "prejudice" or "mislead" the jury, thereby leading to erroneous verdicts. These kinds of rules are rationalized by epistemic considerations, viz. truth-conduciveness or error-avoidance. Since the activity being regulated (viz. the admission or exclusion of offering evidence) is not doxastic decision-making



activity, it goes beyond the ambit of TE and illustrates expansionism at work.

How might these rules or evaluations be systematized? Goldman (1999: 87–90) tries to systematize them by means of a *veritistic* approach. “Veritism” refers to a method of measuring a certain species of epistemic value associated with doxastic states and the truth-values of their propositional contents. Under this method, a person’s doxastic state vis-à-vis question  $Q$  – e.g. whether or not  $p$  is the case – has veritistic value 1.0 if the person categorically believes  $p$  (or has a degree-of-belief in  $p$  equal to 1.0) and it’s true. If the person has degree of belief 1.0 in  $p$  and it’s false – equivalently, has degree of belief 0 in not- $p$  and it’s true – the veritistic value of his doxastic state is 0. Veritistic values intermediate between 0 and 1 are associated with degrees of belief in a true answer to a question falling between full belief and full rejection. (Assignment of veritistic value also depends on the person’s having an interest in question  $Q$ .) Goldman posits a “weak” sense of knowledge in which it consists simply in true belief (defended more fully in Goldman and Olsson, forthcoming). So veritistically good states are also described as states of knowledge.

The foregoing presents the approach to “fundamental” veritistic value (V-value). Things other than doxastic states can be assigned “instrumental” V-value in virtue of their tendencies to produce new doxastic states, or changes in doxastic states, with fundamental V-value. In particular, various *practices*, or *activities*, are said to possess instrumental V-value. The practices in question can be either individual practices or social practices; the latter are the focus of *Knowledge in a Social World*. A social practice might have such impact on a given occasion by producing an increase (decrease) in a single person’s degree of belief towards (away from) the truth, or by producing an increase (decrease) of a group’s *mean* degree of belief towards (away from) the truth. There are two possible *modes* of veritistic evaluation under the proposed approach. One mode is the absolute mode. This assesses the instrumental V-value of a practice without reference to alternative practices. Does the target prac-

tice tend, on average, to produce positive or negative changes in fundamental V-value? The second mode of evaluation is comparative. Here the theorist asks which of two specified practices,  $\pi$  and  $\pi'$ , would have superior veritistic properties (applied over the same range of applications). For example, which set of legal rules for evidence admission and exclusion would yield a veritistically better set of verdicts? Needless to say, any attempt to determine the comparative veritistic properties of two practices will encounter many hurdles. None the less, this is often the kind of thing that policy-makers want to know, however difficult it may be to get an answer.

The central point for present purposes is that a wide assortment of social practices, not just doxastic decision-making practices, can be objects of veritistic evaluation. This is a principal point on which veritistic expansionism parts company with TE. Despite this difference, expansionism retains most of the fundamental assumptions of TE – for example, epistemic objectivism and continued (but not exclusive) emphasis on individual epistemic agents as opposed to collective agents. Acknowledgment of collective epistemic agents is a further departure of expansionism, however, addressed in the final section below.

#### EXPANSIONIST SOCIAL EPISTEMOLOGY: THE EVIDENTIAL VARIANT

The veritistic version of expansionism may not appeal to everyone. Qualms may center not on the new social practices brought under epistemology’s tent but on the veritistic style of evaluation. True belief may have some epistemic value, but doesn’t (“strong”) knowledge have greater epistemic value? So why should social epistemology focus narrowly on true belief? Why not focus on knowledge in the usual, “strong” sense, or on justified belief? To accommodate such preferences, let us explore a second way to motivate expansionism, one that centers on the creation and utilization of *social evidence*. The evidence focus will be neutral with respect to the standard of epistemic evaluation. It



would fit with either a knowledge-based standard or a justification-based standard (assuming that evidence is important under both standards).

Evidence-gathering is an epistemic activity that TE barely touches, despite its pervasiveness in everyday life and its salience in science. Evidence-gathering runs the gamut from ordinary acts of perceptual observation to the use of carefully crafted instruments of measurement and experimentation. In physical science, of course, the sought-after evidence is aimed at understanding physical objects and systems. So the evidence in question is usually physical evidence. In the social and policy sciences, however, the evidence of interest primarily concerns people, so social evidence predominates. This initial partition, however, underestimates the role of social evidence in all of science – and in scholarship and everyday life, for that matter. All science, even physical science, is a highly social activity in the sense that virtually every scientist relies on the scientific reports and opinions of others (Hardwig, 1985, 1991; Kitcher, 1993, ch. 8). Research begins with a literature review, which consists in gathering the reports of previous researchers. This is social evidence. To the extent that science is collaborative, most research projects feature reliance of each participant on the technical opinions of the other participants. These opinions are also specimens of social evidence (for the initial participant).

We should say more about how evidence in general and social evidence in particular is conceptualized here. There are many ways to conceptualize evidence (Kelly, 2006). Here I mean by evidence something in the world, outside the mind of the epistemic subject. Evidence is something the subject discovers in the world, and his mental representation of the evidence is what he uses in subsequent deliberation. Social evidence consists in actions or states of mind of other persons. For present purposes, social evidence consists of either doxastic states or (declarative) acts of communication: very roughly, what they believe and what they assert. This might be called *social-cognitive* evidence; but I won't burden readers with this cumbersome expression.

Like any other evidence, there are different things that social evidence might be evidence *for*. A default assumption, however, is that someone's asserting  $p$  is evidence that she *believes*  $p$ , and her believing  $p$  is evidence for the truth of  $p$ . The *quality* of such evidence, however, may vary considerably from case to case. Now, if a person is interested in the answer to a certain question but cannot personally obtain probative physical evidence, it is natural to seek social evidence that bears on the question. In matters extending beyond one's personal spatio-temporal environment, probative physical evidence is often difficult to obtain. So people naturally seek social evidence: to hear from travelers what foreign countries are like, to read the writings of ancient people to learn about the past. Acquiring such evidence, when it exists and is accessible, helps one get justified belief and even knowledge with respect to the questions of interest.

Edward Craig (1990) paints a similar portrait of our basic human situation by telling a "genealogical" story according to which the basic need for truth drives our practice of seeking out "good informants" to multiply our epistemic resources, i.e. people who will tell us the truth as to whether  $p$ . However, given the possibility that informants may be incompetent in their belief formation and/or deceptive in their speech, there must be some public means of distinguishing good informants. So societies often select "indicator properties" to help epistemic agents identify who is or isn't a competent believer and trustworthy speaker. This social selection process is plausibly considered a social-epistemic activity, and it belongs under the umbrella of activities to be studied by expansionist social epistemology. As Miranda Fricker (1998) points out, it is distinctly possible for societies to err in assigning indicator-properties. A society may deem a certain class of people to be good informants although they aren't, or fail to assign credibility to people who actually deserve it. Because of the relationships between credibility and power, this kind of epistemic practice, says Fricker, inflicts on some people a kind of injustice (Fricker, 2007). An example of this sort, explored by Steven

Shapin (1994), was the assignment of special credibility, in seventeenth-century England, to members of the “gentlemanly” class. Being a gentleman was considered an “indicator property” in that culture. Elizabeth Anderson (2006) provides an analogous example in stressing the importance to democracies of listening to diverse informants.

For many segments of expansionist social epistemology a good general heading might be “the public task of pooling our collective epistemic resources”. One such task, analyzed by Philip Kitcher (1990), is how science should divide the labor among multiple inquirers. Funding agencies must often decide where to place their bets, which specific lines of scientific research they should support. Kitcher shows that, even if an individual chemist’s chances of discovering the structure of an important molecule were greater using method A rather than method B, it doesn’t follow that the optimal distribution of community effort (and hence funding) is to have all chemists use method A. Instead, it is better to diversify efforts: let some chemists use A and others use B.

Other questions about pooling collective resources concern the optimal uses of digital technologies to create and disseminate social evidence. *Wikipedia*, the online encyclopedia, is an example of such a project. It makes an enormous quantity of social evidence readily available about a myriad of subjects, with the intent of improving the epistemic situations of its many readers. *Wikipedia* is the product of mass collaboration; it can be edited by anyone with Internet access. Don Fallis (2008) offers a balanced evaluation of the epistemic strengths and weaknesses of *Wikipedia*. What a given entry says concerning its subject presumably constitutes *some* (positive) evidence of its truth, but the quality of the evidence is debatable. One skeptic writes “it’s the blind leading the blind – infinite monkeys providing infinite information for infinite readers, perpetuating the cycle of misinformation and ignorance” (Keen, 2007: 4). The assessment Fallis offers of *Wikipedia*’s epistemic prowess is more positive (though guarded).

Another arena in which choices about social evidence abound is the law. Courts

sometimes choose to *compel* certain witnesses to testify even when they are reluctant to do so. A different choice made by courts, i.e. judges, is whether to *allow* a witness to testify who is offered as an expert by one of the parties. If the witness purports to be a scientific expert, the *Daubert* (1993) ruling declares that the presiding judge must decide whether he is qualified to testify. Because judges are usually lay people with respect to the sciences, the epistemic quality of this procedure raises controversial issues (Brewer, 2006).

In their role as admitters or excluders of certain testimony, judges serve as “gatekeepers” to speech within the courtroom. Gatekeeping is a role in the communication process that is fairly ubiquitous and an appropriate subject for epistemic analysis. Gatekeeping is managing a communication channel, such as a newspaper op-ed page, an academic journal, or a television news show. Gatekeepers decide which works to publish or which experts to voice their opinions; in other words, which social evidence to present to a large audience. Some theorists would say that there should not be gatekeepers at all, that society’s epistemic ends are best-served by a free market for speech, in which speech opportunities are allocated by the usual market mechanisms. This is an issue to which expansionist social epistemology can contribute (Goldman and Cox, 1996; Goldman, 1999, ch. 7).

#### CAN COLLECTIVE EPISTEMIC AGENTS BE RATIONAL?

A final theme for expansionist social epistemology concerns the existence and character of collective epistemic agents. Both TE and preservationism restrict themselves to individual agents. Revisionism (or some versions thereof) advocate communities as the only epistemic agents. Expansionism proposes to admit both individual and collective epistemic agents.

In recent years numerous philosophers have defended the metaphysical legitimacy of group entities with intentional attitudes (see especially Pettit, 2003). This accords with everyday language and formal discourse

about governments, legal tribunals, academic committees, and the like. If group entities have intentional attitudes, beliefs and judgments, another question is whether their patterns of judgment can meet standards of epistemic *rationality*. There are intriguing problems in this terrain that belong on the agenda of (expansionist) social epistemology.

This issue is the locus of a new sub-field called *judgment aggregation*. Philip Pettit (2003) identifies and generalizes a certain paradox, which he calls the “discursive dilemma”. It concerns the relationship between factual judgments of individual members of a group and judgments of the group *per se*. When a committee or tribunal must render a judgment on a set of logically related propositions, there are two ways to proceed. Consider an argument in which propositions A and B are premises and C is the conclusion. If the group’s judgment on the truth of these propositions is fixed by majority vote, one procedure would have all members vote on each premise, let the majority vote on each premise determine the group judgment on each premise, and let the group’s premise judgments determine the group’s conclusion judgment. Alternatively, each member could first fix his/her own conclusion judgment and let the individual conclusion judgments determine the group’s conclusion judgment. In some circumstances, however, the two methods will yield different results. Which is the preferable method for aggregating the group’s judgments?

Problems get more challenging in light of an “impossibility” theorem proved by Christian List and Pettit (2002; *see also* List and Pettit, forthcoming). This theorem establishes the impossibility of group rationality analogous to Kenneth Arrow’s (1963) impossibility theorem that launched social choice theory. Call the judgments about a set of propositions rational just in case the set is consistent and complete. Assume that each group member makes an individually rational set of judgments on these propositions. According to the theorem, however, there exists no possible aggregation procedure that would generate collective judgments from member judgments that satisfies the rationality constraint plus the following three conditions: (a) universal

domain, (b) anonymity, and (c) systematicity (*see* List, 2005, for an explanation and exploration of these conditions).

What, then, are the prospects for group rationality in light of these formal findings? The question is analogous to questions that interest epistemologists when confronted with other paradoxes, such as the lottery paradox or the paradox of the preface. Rationality is a central concept of epistemic evaluation, so it is highly appropriate for social epistemologists to evaluate the prospects for rational or almost-rational procedures for aggregating factual judgments. Both TE and preservationist social epistemology confine their attention to doxastic decision-making by individual agents. There is a clear continuity, however, between issues of rational decision-making by individuals and issues of rational decision-making by collectives. The latter is a topic that naturally falls within the purview of social epistemology under its expansionist variant.

#### BIBLIOGRAPHY

- Anderson, Elizabeth: “The Epistemology of Democracy,” *Episteme, a Journal of Social Epistemology* 3: 1–2 (2006), 8–22.
- Annis, David: “A Contextualist Theory of Epistemic Justification,” *American Philosophy Quarterly* 15 (1978), 213–19.
- Arrow, Kenneth: *Social Choice and Individual Values* (New York: Wiley, 1963).
- Barnes, Barry and Bloor, David: “Relativism, Rationalism, and the Sociology of Knowledge,” in *Rationality and Relativism*, ed. M. Hollis and S. Lukes (Cambridge, MA: MIT Press, 1982).
- Bergmann, Michael: “Epistemic Circularity: Malignant and Benign,” *Philosophy and Phenomenological Research* 69 (2004), 709–27.
- Boghossian, Paul: *Fear of Knowledge: Against Relativism and Constructivism* (Oxford: Oxford University Press, 2006).
- Brewer, Scott: “Scientific Expert Testimony and Intellectual Due Process,” in *The Philosophy of Expertise*, ed. E. Selinger and R. P. Crease (New York: Columbia University Press, 2006).
- Burge, Tyler: “Content Preservation,” *Philosophical Review* 102 (1993), 457–88.
- Christensen, David: “Epistemology of Disagreement: The Good News,” *Philosophical Review* 116 (2007), 187–217.

- Coady, C. A. J.: *Testimony: A Philosophical Study* (Oxford: Oxford University Press, 1992).
- Craig, Edward: *Knowledge and the State of Nature* (Oxford: Oxford University Press, 1990).
- DeRose, Keith: "Assertion, Knowledge, and Context," *The Philosophical Review* 111 (2002), 167–203.
- Elga, Adam: "Reflection and Disagreement," *Nous* 41 (2007), 478–502.
- Fallis, Don: "Toward an Epistemology of Wikipedia," *Journal of the American Society for Information Science and Technology* (2008).
- Feldman, Richard: "Reasonable Religious Disagreements," in *Philosophers without Gods: Meditations on Atheism and the Secular Life*, ed. L. Antony (Oxford: Oxford University Press, 2007).
- Fricker, Elizabeth: "Telling and Trusting: Reductionism and Anti-Reductionism in the Epistemology of Testimony," *Mind* 104 (1995), 393–411.
- Fricker, Miranda: "Rational Authority and Social Power: Towards a Truly Social Epistemology," *Proceedings of the Aristotelian Society*, NS 98 (1998), 159–77.
- Fricker, Miranda: *Epistemic Injustice: Power and the Ethics of Knowing* (Oxford: Oxford University Press, 2007).
- Goldman, Alvin: "Argumentation and Social Epistemology," *Journal of Philosophy* 91 (1994), 27–49.
- Goldman, Alvin: *Knowledge in a Social World* (Oxford: Oxford University Press, 1999).
- Goldman, Alvin: "Three Approaches to Social Epistemology" (forthcoming a), in *Readings in Social Epistemology*, ed. A. Goldman and D. Whitcomb (New York: Oxford University Press, forthcoming).
- Goldman, Alvin: "Social Evidence and Expansionist Social Epistemology" (forthcoming b), in *Social Epistemology*, ed. D. Pritchard (forthcoming).
- Goldman, Alvin and Cox, James: "Speech, Truth, and the Free Market for Ideas," *Legal Theory* 2 (1996), 1–32.
- Goldman, Alvin and Olsson, Erik: "Reliabilism and the Value of Knowledge," in *Epistemic Value*, ed. D. Pritchard, A. Millar and A. Haddock (Oxford: Oxford University Press, forthcoming).
- Grice, H. P.: *Studies in the Way of Words* (Cambridge, MA: Harvard University Press, 1980).
- Hardwig, John: "Epistemic Dependence," *Journal of Philosophy* 82 (1985), 335–49.
- Hardwig, John: "The Role of Trust in Knowledge," *Journal of Philosophy* 88 (1991), 693–704.
- Hawthorne, John: *Knowledge and Lotteries* (Oxford: Oxford University Press, 2004).
- Hume, David: *An Inquiry Concerning Human Understanding*, ed. L. A. Selby-Bigge, 2nd edn (Oxford: Oxford University Press, 1957).
- Keen, Andrew: *Cult of the Amateur* (New York: Doubleday, 2007).
- Kelly, Thomas: "The Epistemic Significance of Disagreement," in *Oxford Studies in Epistemology*, ed. J. Hawthorne and T. Gendler, vol. 1 (Oxford: Oxford University Press, 2005).
- Kelly, Thomas: "Evidence," *Stanford Encyclopedia of Philosophy*, 2006 edn.
- Kelly, Thomas: "Peer Disagreement and Higher-order Evidence," in *Disagreement*, ed. Richard Feldman and Ted Warfield (Oxford: Oxford University Press, forthcoming).
- Kitcher, Philip: "The Division of Cognitive Labor," *Journal of Philosophy* 87 (1990), 5–22.
- Kitcher, Philip: *The Advancement of Science* (Oxford: Oxford University Press, 1993).
- Kuhn, Thomas: *The Structure of Scientific Revolutions* (Chicago, IL: University of Chicago Press, 1962).
- Kusch, Martin: *Knowledge by Agreement* (Oxford: Oxford University Press, 2002).
- Lackey, Jennifer: *Learning from Words: Testimony as a Source of Knowledge* (Oxford: Oxford University Press, 2008).
- Lackey, Jennifer: "A Justificationist View of Disagreement's Epistemic Significance," in *Social Epistemology*, ed. D. Pritchard (forthcoming).
- Latour, Bruno and Woolgar, Steve: *Laboratory Life: The Construction of Scientific Facts*, 2nd edn (Princeton, NJ: Princeton University Press, 1986).
- List, Christian: "Group Knowledge and Group Rationality: A Judgment Aggregation Perspective," *Episteme, a Journal of Social Epistemology* 2: 1 (2005), 25–38.
- List, Christian and Pettit, Philip: "Aggregating Sets of Judgments: An Impossibility Result," *Economics and Philosophy* 18 (2002), 89–110.
- List, Christian and Pettit, Philip: *Group Agency: The Possibility, Design, and Status of Corporate Agents* (forthcoming).
- Nelson, Lynn Hankinson: "Epistemological Communities," in *Feminist Epistemologies*, ed. L. Alcoff and E. Potter (New York: Routledge, 1993).
- Pettit, Philip: "Groups with Minds of Their Own," in *Socializing Metaphysics: The Nature of Social Reality*, ed. F. Schmitt (Lanham, MD: Rowman and Littlefield, 2003).
- Rorty, Richard: *Philosophy and the Mirror of Nature* (Princeton, NJ: Princeton University Press, 1979).
- Schmitt, Frederick: "Testimonial Justification and Transindividual Reasons," in *The Epistemology of*



- Testimony*, ed. J. Lackey and E. Sosa (New York: Oxford University Press, 2006).
- Shapin, Steven: *A Social History of Truth* (Chicago, IL: University of Chicago Press, 1994).
- Sosa, Ernest: "... Disagreement . . .," in *Social Epistemology*, ed. D. Pritchard (forthcoming).
- Stanley, Jason: *Knowledge and Practical Interests* (Oxford: Oxford University Press, 2005).
- Van Cleve, James: "Is Knowledge Easy – or Impossible? Externalism as the Only Alternative to Skepticism," in *The Skeptics*, ed. S. Luper (Aldershot: Ashgate, 2003).
- Van Inwagen, Peter: "It Is Wrong, Everywhere, Always, and for Anyone, to Believe Anything on Insufficient Evidence," in *Faith, Freedom, and Rationality: Philosophy of Religion Today*, ed. J. Jordan and D. Howard-Snyder (Totowa, NJ: Rowman and Littlefield, 1996).
- Weiner, Matthew: "Must We Know What We Say?," *The Philosophical Review* 114 (2005), 227–51.
- Williams, Michael: *Groundless Belief: An Essay on the Possibility of Epistemology*, 2nd edn (Princeton: Princeton University Press, 1999).
- Williamson, Timothy: *Knowledge and Its Limits* (Oxford: Oxford University Press, 2000).
- Wittgenstein, Ludwig: *Philosophical Investigations*, 3rd edn (Oxford: Blackwell, 1968).
- Wittgenstein, Ludwig: *On Certainty*, trans. G. E. M. Anscombe and G. H. von Wright (New York: Harper, 1969).

ALVIN GOLDMAN

## Bayesian Epistemology

### 1. INTRODUCTION

Bayesianism is our leading theory of *uncertainty*. Epistemology is defined as the theory of *knowledge*. So "Bayesian Epistemology" may sound like an oxymoron. Bayesianism, after all, studies the properties and dynamics of degrees of belief, understood to be probabilities. Traditional epistemology, on the other hand, places the singularly non-probabilistic notion of knowledge at center stage; and, to the extent that it traffics in belief, that notion does not come in degrees. So how can there be a Bayesian epistemology?

According to one view, there cannot: Bayesianism fails to do justice to essential aspects of knowledge and belief, and as such

it cannot provide a genuine epistemology at all. According to another view, Bayesianism should supersede traditional epistemology: where the latter has been mired in endless debates over skepticism and Gettierology, Bayesianism offers the epistemologist a thriving research program. We shall advocate a more moderate view: Bayesianism can illuminate various long-standing problems of epistemology, while not addressing all of them; and, while Bayesianism opens up fascinating new areas of research, it by no means closes down the staple preoccupations of traditional epistemology.

The contrast between the two epistemologies can be traced back to the mid-seventeenth century. Descartes regarded belief as an all-or-nothing matter, and he sought justifications for his claims to knowledge in the face of powerful skeptical arguments. No more than four years after his death, Pascal and Fermat inaugurated the probabilistic revolution, writ large in the *Port-Royale Logic*, in which the many shades of uncertainty are represented with probabilities, and rational decision-making is a matter of maximizing expected utilities (as we now call them). Correspondingly, the Cartesian concern for knowledge fades into the background, and a more nuanced representation of epistemic states has the limelight. Theistic belief provides a vivid example of the contrasting orientations. Descartes sought certainty in the existence of God grounded in apodeictic demonstrations. Pascal, by contrast, explicitly shunned such alleged "proofs", arguing instead that our situation with respect to God is like a gamble, and that belief in God is the best bet – thus turning the question of theistic belief into a *decision* problem (which he, unlike Descartes, had the tools to solve). Bayesian epistemology owes its name to the Reverend Thomas Bayes, who, a century later, published an important theorem that underwrites certain calculations of conditional probability central to confirmation theory – more on this shortly. But really "Bayesian epistemology" is something of a misnomer; "Kolmogorovian epistemology" would be far more appropriate, as we shall see.

Caveats: When we speak of “traditional epistemology”, we lump together a plethora of positions as if they form a monolithic whole. Other articles in this volume distinguish carefully among various positions that our broad banner conflates. For our purposes, they start out regarding knowledge and belief as the central concepts of epistemology, and then to go on to study the properties, grounds, and limits of these binary notions. We also speak of “Bayesianism” as if it is a unified school of thought, when in fact there are numerous intra-mural disputes. I. J. Good (1971) calculates that there are (at least) 46,656 ways to be a Bayesian, while we shall mostly pretend that there is just one. By and large, the various distinctions among Bayesians will not matter for our purposes. As a good (indeed, a Good) Bayesian might say, our conclusions will be robust under various precisifications of the position. Many traditional problems can be framed, and progress can be made on them, using the tools of probability theory. But Bayesian epistemology does not merely re-create traditional epistemology; thanks to its considerable expressive power, it also opens up new lines of enquiry.

## 2. WHAT IS BAYESIAN EPISTEMOLOGY?

Bayesian epistemology is the application of Bayesian methods to epistemological problems. Bayesianism models degrees of belief as *probabilities* along the lines of Kolmogorov’s (1933) axiomatization. Let  $\Omega$  be a non-empty set. A *field (algebra)* on  $\Omega$  is a set  $\mathcal{F}$  of subsets of  $\Omega$  that has  $\Omega$  as a member, and that is closed under complementation (with respect to  $\Omega$ ) and union. Let  $P$  be a function from  $\mathcal{F}$  to the real numbers obeying:

- (1)  $P(a) \geq 0$  for all  $a \in \mathcal{F}$ . (Non-negativity)
- (2)  $P(\Omega) = 1$ . (Normalization)
- (3)  $P(a \cup b) = P(a) + P(b)$  for all  $a, b \in \mathcal{F}$  such that  $a \cap b = \emptyset$ . (Finite additivity)

Call  $P$  a *probability function*, and  $(\Omega, \mathcal{F}, P)$  a *probability space*.

One could instead attach probabilities to members of a collection of sentences of a formal language, closed under truth-functional combinations; this is more common in Bayesian confirmation theory. A lively area of current debate concerns just how finely grained such contents of probability attributions should be. For example, various problems of “self-location” suggest that probabilities should attach to “centered propositions”, e.g.  $\langle \text{possible world, individual, time} \rangle$  triples.

Kolmogorov extends his axiomatization to cover infinite probability spaces, requiring  $\mathcal{F}$  to be closed under *countable* union, and strengthening (3) to *countable* additivity. He defines the *conditional probability of a given b* by the ratio of unconditional probabilities:

$$P(a \mid b) = \frac{P(a \cap b)}{P(b)}, \text{ provided } P(b) > 0.$$

If  $P(a \mid b) = P(a)$ , then  $a$  and  $b$  are said to be *independent* (relative to  $P$ ).

Versions of *Bayes’ theorem* can now be proven:

$$\begin{aligned} P(a \mid b) &= \frac{P(b \mid a)P(a)}{P(b)} \\ &= \frac{P(b \mid a)P(a)}{P(b \mid a)P(a) + P(b \mid \neg a)P(\neg a)} \end{aligned}$$

More generally, suppose there is a partition of hypotheses  $\{h_1, h_2, \dots, h_n\}$ , and evidence  $e$ . Then for each  $j$

$$P(h_i \mid e) = \frac{P(e \mid h_i)P(h_i)}{\sum_{j=1}^n P(e \mid h_j)P(h_j)}$$

The  $P(e \mid h_i)$  terms are called *likelihoods*, and the  $P(h_i)$  terms are called *priors*. See Joyce (2008) for a more detailed discussion of varieties and uses of Bayes’s theorem.

Bayesianism offers a natural analysis of the relation of *confirmation* between a piece of evidence  $e$  and a hypothesis  $h$ :

$e$  confirms  $h$  (relative to  $P$ ) iff  $P(h \mid e) > P(h)$ .



We may also define various probabilistic notions of *comparative* confirmation, and various *measures of evidential support* (see Eells and Fitelson, 2007; and Fitelson, 1999).

Bayesianism can be understood as combining a *synchronic* thesis about the degrees of belief or *credences* of a rational agent at a given time, and a *diachronic* thesis about how they evolve in response to evidence. Synchronically, the agent's credences are probabilities. Diachronically, her credences update according to the rule of *conditionalization*. Suppose that initially her credences are given by probability function  $P_{initial}$ , and that she becomes certain of  $e$  (where  $e$  is the strongest such proposition). Then her new credence function  $P_{new}$  is related to  $P_{initial}$  as follows:

(Conditionalization)  $P_{new}(x) = P_{initial}(x | e)$   
(provided  $P_{initial}(e) > 0$ ).

Jeffrey conditionalization allows for less decisive learning experiences in which her probabilities across a partition  $\{e_1, e_2, \dots\}$  change to  $\{P_{new}(e_1), P_{new}(e_2), \dots\}$ , where none of these values need be 0 or 1:

$P_{new}(x) = \sum_i P_{initial}(x | e_i) P_{new}(e_i)$  (provided  $P_{initial}(e_i) > 0$ ).

(Jeffrey, 1983.  $P_{new}(x)$  is called the *posterior* probability function).

According to Bayesian orthodoxy, an agent begins with a "prior" probability function and repeatedly updates by (Jeffrey) conditionalization as evidence comes in. This combines a striking permissiveness about the starting point of an agent's epistemological odyssey with considerable rigidity about how the agent should respond to evidence. But Bayesianism is a theme that admits of many variations – see Good (1971) and Hájek and Hall (2002) for discussion of some of them. Here, let us consider several further constraints on "priors" that have been proposed.

A probability function is said to be *regular* if it assigns probability 1 only to tautologies, and probability 0 only to contradictions – to all other sentences it assigns intermediate values. It seems to be an epistemological

desideratum that a prior be regular, reflecting an open-mindedness appropriate to an agent who is a *tabula rasa*. A rationale is that to rule out (probabilistically speaking) a priori some genuine logical possibility would be to pretend that one's evidence was stronger than it really was. The *principle of indifference* also enjoins you to reflect the poverty of your evidence in your credences: you are required to give equal probability to all possibilities among which your evidence does not discriminate (and, in a state of total evidential innocence, that is all of them). A sophisticated version of the principle of indifference, favoured by so-called *objective* Bayesians, has been explored by Jaynes (2003): maximize the probability function's *entropy*, which for an assignment of positive probabilities  $p_1, \dots, p_n$  to  $n$  possibilities equals  $-\sum_i p_i \log(p_i)$ . See also Williamson (2005).

Then there are two principles that are meant to codify one's epistemic commitment to aligning one's credences to certain probabilistic hypotheses. In the first, Lewis's *Principal Principle* (here simplified), the hypotheses concern the *objective chance* of the relevant propositions (Lewis, 1980):

$C_0(a | ch_t(a) = x) = x$  (for all  $a$  and  $t$  for which this is defined).

Here  $C_0$  is some reasonable prior,  $a$  an arbitrary proposition, and  $ch_t(a) = x$  the claim that the chance at time  $t$  of  $a$  is  $x$ . The idea is that one should align one's credences with what one takes the corresponding objective chances to be, where the latter are genuine probabilities *in the world*. In the second, Van Fraassen's *Reflection Principle* (Van Fraassen, 1984), the hypotheses concern one's own future credences for the relevant propositions

$C_t(a | C_{t'}(a) = x) = x$  (for all  $t, t', a$  and  $x$  for which this is defined).

Here  $C_t$  is one's probability function at time  $t$ , and  $C_{t'}$  one's function at later time  $t'$ . The idea is that rationality requires a certain commitment to one's future opinions; when all is going well, one's future selves are better-informed versions of one's current self.

### 3. CONTRASTS BETWEEN TRADITIONAL EPISTEMOLOGY AND BAYESIAN EPISTEMOLOGY

We can now bring out several points of contrast between traditional and Bayesian epistemology. We have noted that “knowledge” and “belief” are binary notions, to be contrasted with the potentially infinitely many degrees of “credence” (corresponding to all the real numbers in the  $[0, 1]$  interval). “Knowledge” is not merely “justified true belief”, as Gettier has famously shown, but many epistemologists hope that some “fourth condition” will complete the analysis – some kind of condition that rules out cases in which one has a justified true belief by luck, or for some anomalous reason. Some epistemologists (e.g. Sosa, 1999) advocate versions of *safety* as a condition on knowledge – roughly, at the closest worlds in which a given agent believes  $p$ ,  $p$  is true. Others (e.g. Nozick, 1981) advocate *sensitivity* – roughly, at the closest worlds in which  $p$  is false, the agent does not believe  $p$ . (And some advocate both.) Note well: the *closest* worlds. Here we find another disjuncture between traditional and Bayesian epistemology: nothing in the standard Bayesian apparatus reflects the “similarity” of worlds that has taken center-stage in the analysis of knowledge.

Notice that truth, justification, and these anti-luck conditions may be characterized as at least partially *objective*, with “belief” providing the only purely *subjective* component. This is in sharp contrast to orthodox Bayesianism, which refines and analyzes this doxastic notion, but which has no clear analogue of the “objective” conditions. Most importantly, Bayesianism apparently has nothing that corresponds to the *factivity* of knowledge: that one can only know truths. And, even when our beliefs fall short of knowledge, still it is a desideratum that they be true; but the Bayesian seems to have no corresponding desideratum for intermediate credences, which are its stock-in-trade. When you assign, for example, probability 0.3 to it raining tomorrow, what sense can be made of this assignment being *true*? It is also dubious whether Bayesianism

can capture “justification” (cf. Shogenji, 2009, for an interesting proposal) or any “anti-luck” condition on knowledge – more on this shortly.

Relatedly, all Bayesian claims must be relativized to a probability function or, more precisely, to a probability space – an entire probability model. We saw this above in the definitions of “independence” and “confirmation” – they came with parenthetical references to  $P$ . Many authors suppress these references, encouraging one to forget their inherent subjectivity (and even we secreted them away in slightly disingenuous parentheses!). Traditional epistemologists, by contrast, conduct much of their discussions in terms free of any such relativization – they speak of one proposition being *evidence* for another, of a process of belief acquisition being *reliable*, and so on without any qualification. And, again, truth, justification, and “anti-luck” conditions are typically supposed to hold or not independently of whether some agent *thinks* that they do, or whether some *model* says they do.

The synchronic requirement that an agent’s credences obey the probability calculus may be regarded as generalizing the requirement familiar from traditional epistemology that one’s beliefs should be consistent. The diachronic requirement of conditionalization is reminiscent of the Quinean principle of “minimum mutilation” (change beliefs as little as the evidence allows) – a conservative recipe for belief revision. But nothing in traditional epistemology corresponds to Jeffrey conditionalization – it is *essentially* a probabilistic revision rule. The principle of indifference corresponds very roughly to a Cartesian admonition to suspend judgment when one’s evidence is lacking, but it is far more specific. And there are no traditional analogues of the various additional constraints on priors. Going in the other direction, Bayesianism is silent about some of the cornerstones and more recent concerns of traditional epistemology – we shall discuss this at greater length at the end.

Given the striking differences between traditional and Bayesian epistemology, are there reasons to prefer one to the other?

#### 4. THESIS: BAYESIAN EPISTEMOLOGY IS SUPERIOR TO TRADITIONAL EPISTEMOLOGY

Jeffrey, a famous Bayesian, suggests two main benefits accrued by the Bayesian framework in his (1992):

1. Subjective probabilities figure in *decision theory*, an account of how our opinions and our desires conspire to dictate what we should do. The desirability of each of our possible actions is measured by its *expected utility*, a probability-weighted sum of the utilities associated with that action. To complete Jeffrey's argument, we should add that traditional epistemology offers no decision theory (recall Descartes versus Pascal). The analysis of rational action surely needs to advert to more fine-grained mental states than binary belief and knowledge. (See Eriksson and Hájek, 2007, for more discussion of why the intermediate credences that are necessary for that analysis cannot be reduced to these binary notions.)
2. Observations rarely deliver certainties – rather, their effect is typically to raise our probabilities for certain propositions (and to drop our probabilities for others), without any reaching the extremes of 1 or 0. Traditional epistemology apparently has no way of accommodating such less-than-conclusive experiential inputs, whereas Jeffrey conditionalization is tailor-made to do so.

We may continue the list that Jeffrey has started of putative advantages of Bayesianism over traditional epistemology at some length.

3. Knowledge is unforgiving. Its standards are so high that they can rarely be met, at least in certain contexts. (This is related to the fact that knowledge does not come in degrees – near-knowledge is not knowledge at all.) This in turn plays into the hands of skeptics. But it is harder for skeptical arguments to get a toehold against the Bayesian. For example, the mere *possibility* of error regarding some proposition *X*

undermines a claim of knowledge regarding *X*, but it is innocuous from a probabilistic point of view: an agent can simply assign *X* some suitable probability less than 1. Indeed, even an assignment of probability 1 is consistent with the possibility of error – under plausible assumptions, it can be shown that a dart thrown at random at a representation of the  $[0, 1]$  interval has probability 1 of hitting an irrational number, even though it might fail to do so.

4. Moreover, it is a platitude that doxastic states come in degrees, and the categories of “belief” and “knowledge” are too coarse-grained to do justice to this fact. You believe, among other things, that  $2 + 2 = 4$ , that you have a hand, that London is in England, and (say) that Khartoum is in Sudan. But you do not have the same confidence in all these propositions, as we can easily reveal in your betting behavior and other decision-making that you might engage in. The impoverished nature of “belief” attributions is only exacerbated when we consider the wide range of propositions for which you have less confidence – that this coin will land heads, that it will rain tomorrow in Novosibirsk, and so on. We may conflate your attitudes to them all as “suspensions of belief” (as Descartes would), but that belies their underlying structure. Such attitudes are better-understood as subjective probabilities.
5. Relatedly, the conceptual apparatus of deductivism is impoverished, and comparatively little of our reasoning can be captured by it, either in science or in daily life (*pace* Popper and Hempel). After all, whether we like it or not, our epistemic practices constantly betray our commitment to relations of support that fall short of entailment (Chater and Oaksford, 2007). We think that it would be irrational to deny that the sun will rise tomorrow, to project “grue” rather than “green” in our inductions, and to commit the *gambler's fallacy*. Probability theory helps us to understand such relations.
6. Bayesianism has powerful mathematical underpinnings. It can help itself to a

century of work in probability theory and statistics. Traditional epistemology may appeal to the occasional system of epistemic or doxastic logic, but nothing comparable to the formidable formal machinery that we find in the Bayesian's toolkit.

7. Bayesian methods, in turn, have much wider application than any formal systematization of "knowledge" or "belief". Look at the sciences, social sciences, engineering, and artificial intelligence if you need any convincing of this.
8. There are many arguments for Bayesianism, which collectively provide a kind of triangulation to it. For example, "Dutch book arguments" provide an important defense of the thesis that rational credences are probabilities. An agent's credences are identified with her betting prices; it is then shown that she is susceptible to sure losses iff these prices do not conform to Kolmogorov's axioms. There are also arguments from various decision-theoretic representation theorems (Ramsey, 1931; Savage, 1954; Joyce, 1999), from calibration (Van Fraassen, 1984), from "gradational accuracy" or minimization of discrepancy from truth (Joyce, 1998), from qualitative constraints on reasonable opinion (see Earman, 1996, for a discussion of results of Cox and others), and so on. Moreover, there are various arguments in support of conditionalization and Jeffrey conditionalization – e.g., Dutch book arguments (Armendt, 1980; Lewis, 1999) and arguments from minimal revision of one's credences (Diaconis and Zabell, 1982). Again, there is nothing comparable in traditional epistemology.
9. Finally, a pragmatic argument for Bayesianism comes from an evaluation of its fruits. As we show at greater length in section 7, Bayesianism is highly explanatory with minimal resources – a simple, fecund theory if ever there was one. Traditional epistemology is hard-pressed to offer the same rewards. For example, we shall see how various important intuitions about confirmation can be

vindicated by a Bayesian analysis, and some erroneous intuitions can be corrected. It seems that no analysis couched purely in terms of "knowledge" and "belief" could pay such dividends.

So we see various advantages that Bayesianism apparently has over traditional epistemology. But this does not tell the whole story. For starters, the triumphs of Bayesian confirmation theory just touted are supposedly offset by the so-called *problem of old evidence* (Glymour, 1980). If  $P(e) = 1$ , then  $e$  apparently cannot confirm anything by Bayesian lights: in that case,  $P(h | e) = P(h \cap e)/P(e) = P(h)$ . Yet we often think that such "old evidence" can be confirmatory. Consider the evidence of the advance of the perihelion of Mercury, which was known to Einstein at the time that he formulated general relativity theory, and thus (we may assume) was assigned probability 1 by him. None the less, he rightly regarded this evidence as strongly confirmatory of general relativity theory. The challenge for Bayesians is to account for this (see Zynda, 1995, for discussion).

Bayesianism, then, is not without problems of its own. So let us revisit the contest between traditional epistemology and Bayesianism, this time looking at arguments that incline in favor of the former.

## 5. ANTITHESIS: BAYESIAN EPISTEMOLOGY IS NOT SUPERIOR TO TRADITIONAL EPISTEMOLOGY

1. Bayesians introduce a new technical term, "degree of belief", but they struggle to explicate it. To be sure, the literature is full of nods to betting interpretations, but these meet a fate similar to that of behaviorism – indeed, a particularly *localized* behaviorism that focuses solely on the rather peculiar kind of behavior that is mostly found at race-tracks and casinos. Other characterizations of "degree of belief" that fall out of decision-theoretic representation theorems are also problematic (see Eriksson and Hájek, 2007). "Belief", by contrast, is so familiar to the folk that it needs no explication.

2. Recall the absence of any notion of *truth* of an intermediate degree of belief. Yet truth is the very aim of belief. It is usually thought to consist in *correspondence* to the way things are. Moreover, we want our methods for acquiring beliefs to be *reliable*, in the sense of being *truth-conducive*. What is the analogous aim, notion of correspondence, and notion of reliability for the Bayesian? The terms of her epistemology seem to lack the success-grammar of these italicized words. For example, one can assign very high probability to the period at the end of this sentence being the creator of the universe without incurring any Bayesian sanction: one can do so while assigning correspondingly low probability to the period *not* being the creator, and while dutifully conditionalizing on all the evidence that comes in. Traditional epistemology is not so tolerant, and rightly not.

3. Relatedly, the Bayesian does not answer the skeptic, but merely ignores him. Bayesianism doesn't make skeptical positions go away; it merely makes them harder to state.

4. The Bayesian similarly lacks a notion of "justification" – or, to the extent that she has one, it is too permissive. At least on what we have called Bayesian orthodoxy, any prior is a suitable starting point for a Bayesian odyssey, yet mere conformity to the probability calculus is scant justification.

Now, the Bayesian will be quick to answer this and the previous objections in a single stroke. She will appeal to various *convergence theorems*. For example:

If observations are precise . . . then the form and properties of the prior distribution have negligible influence on the posterior distribution. From a practical point of view, then, the untrammelled subjectivity of opinion . . . ceases to apply as soon as much data becomes available. More generally, two people with widely divergent prior opinions but reasonably open minds will be forced into arbitrarily close agreement about future observations by a sufficient amount of data.

(Edwards et al., 1963: 201)

Call this *convergence to intersubjective agreement*; such agreement, moreover, is often

thought to be the mark of objectivity. The "forcing" here is a result of conditionalizing the people's priors on the data. Gaifman and Snir (1982) similarly show that for each suitably open-minded agent there is a data set sufficiently rich to force her arbitrarily close to assigning probability 1 to the true member of a partition of hypotheses. Call this *convergence to the truth*. The Bayesian might even try to parlay these theorems into providing surrogates for that "fourth condition" for knowledge, insisting that such convergences do not happen by luck, or for some anomalous reason, but are probabilistically guaranteed.

These are beautiful theorems, but one should not overstate their epistemological significance. They are "glass half-full" theorems, but a simple reversal of the quantifiers turns them into "glass half-empty" theorems. For each data set, there is a suitably open-minded agent whose prior is sufficiently perverse to thwart such convergence: after conditionalizing her prior on the data set, she is still nowhere near assigning probability 1 to the true hypothesis, and still nowhere near agreement with other people. And strong assumptions underlie the innocent-sounding phrases "suitably open-minded agent" and "sufficiently rich data set". No data set, however rich, will drive anywhere at all a dogmatic agent who concentrates all credence on a single world (maximally specific hypothesis). Worse, an agent with a wacky enough prior will be driven *away* from the truth. Consider someone who starts by giving low probability to being a brain in a vat, but whose prior regards all the evidence that she actually gets as confirming that she is. And we can always come up with rival hypotheses that no courses of evidence can discriminate between – think of the irresolvable conflict between an atheist and a creationist who sees God's handiwork in everything.

5. A proponent of Bayesianism may describe it as "fecund" (and we did); but an opponent may describe it as "empty". With so little constraint on priors, it is not surprising that Bayesianism accounts for so much. Indeed, it is irrational to deny that the sun will rise tomorrow, to project "grue" rather than "green" in our inductions, and to commit



the *gambler's fallacy*. The trouble is that Bayesianism condones all such inferential practices – for all are licensed by suitably perverse priors.

6. The traditional epistemologist may protest that Bayesians distance themselves from the world. Recall our discussion of the relativization of Bayesian claims to the subjective probability functions of agents. Rather than hooking up directly with the world, the terms of their epistemology are all *internal to probabilistic models* of the world. Moreover, the Bayesian apparently does not have much of a story about what makes a model good, or one model better than another. This is related to the concern that the Bayesian does not do justice to truth, justification, and the “anti-luck” conditions.

What are we to make of these conflicting considerations for and against Bayesian epistemology? At this stage of the dialectic, any good Hegelian will insist that it's time for a:

## 6. SYNTHESIS

Should we really prefer one approach to epistemology over the other? Should one of the two approaches be jettisoned? We shall argue that we should not regard them as in competition. In fact, the two approaches complement each other in both subject matter and method.

Traditional epistemologists sometimes stress that philosophy differs from science and insist that philosophy has its own distinct method of enquiry, namely conceptual analysis. Bayesians, on the other hand, typically consider their work more in line with scientific theorizing. This is reflected in the many connections Bayesian epistemology has with Bayesian statistics, decision theory, and the literature on causal discovery in artificial intelligence. It is also reflected in the importance Bayesians give to solving real problems. As we shall see in the next section, Bayesianism is tremendously successful in this respect. Given these successes, Bayesians should hardly be expected to give up their framework just because it is not a complete panacea for all epistemological ills. By way

of analogy, scientists rightly held on to Newtonian mechanics even in the face of some theoretical problems (for example, its commitment to action at a distance). And, despite its being superseded by Relativity Theory, we understand precisely how Newtonian mechanics is still approximately true. We believe that the successor of Bayesianism will stand in a similar relation to Bayesianism as Relativity Theory stands to Newtonian mechanics.

So let us revisit the charges leveled at Bayesianism in the previous section, keeping in mind this view of it as a work-in-progress that none the less is clearly earning its theoretical keep – like any good scientific theory. We may happily take its fundamental concept, “degree of belief” as a primitive in the absence of a successful analysis of it. It has earned *its* theoretical keep by its contribution to a virtuous total theory – like any primitive scientific concept. And it can live peacefully alongside traditional epistemology's primitive concept of *belief*, without any expectation of reduction of one to the other. Indeed, the prospects for such reduction strike us as unpromising (see Eriksson and Hájek, 2007).

Against the charge of Bayesianism being empty, it can plead the good company of deductive logic. To be sure, crazy sets of belief can be consistent, and inferences from absurd premises can be valid – the slogan “garbage in, garbage out” is as true in epistemology as it is in computer science. Bayesianism, like logic, can nevertheless play a salutary role in keeping our degrees of belief, like our beliefs, in harmony, and in policing our elicited inferences. After all, deductive logic is never regarded as a *complete* set of constraints on belief; similarly, the Bayesian constraints on degrees of belief should not be regarded as complete. Some additional constraints may well find their inspiration in traditional epistemology.

Against the charge of Bayesianism's verdicts being model-relative, it can plead the good company of science. After all, our best methods of enquiry in the physical and social sciences work like this. Arguably we should not expect epistemology to be different.

Nor need answering skeptical challenges be part of Bayesianism's job description, just



as it is not part of traditional epistemology's job description to underwrite rational decision-making, confirmation, and the use of probabilistic and statistical methods in the sciences. There is no harm in their labor being divided. They are two different ways to approach epistemology and they often answer different questions.

And where their questions are shared their approaches can be complementary rather than mutually exclusive. We see that Bayesian epistemology helps to address some questions that we find in traditional epistemological debates. The Bayesian treatment of issues such as testimony and the coherence theory of justification (see section 7) are cases in point. It turns out that the formal machinery of Bayesianism is well suited to make certain questions more precise and to provide answers when our intuitions don't give clear verdicts.

Now let us see what one can do with the machinery of Bayesianism. It should be assessed by the problems it solves and how much it unifies, for example, the methodology of science.

## 7. ACHIEVEMENTS OF BAYESIAN EPISTEMOLOGY

Many of the Bayesian success stories are from confirmation theory. But Bayesianism has much more to offer as its domain of applicability also includes other parts of epistemology and philosophy of science. Here are five highlights.

1. *Confirmation Theory.* As we saw in section 2, Bayesians begin with the idea that confirmation is a matter of *probability-raising*. They then show how important intuitions about confirmation can be vindicated. Suppose that  $h$  entails  $e$ , so that  $P(e | h) = 1$ . Then the posterior probability of  $h$  is  $P(h | e) = P(h)/P(e)$ . Hence, for a fixed prior probability of  $h$ , the posterior probability of  $h$  increases if  $P(e)$  decreases. From this we can immediately account for the methodological insight that *more surprising evidence confirms better*. Similarly, Bayesians have provided a rationale for the *variety-of-evidence thesis* – the

more varied the evidence is, the better – and have provided illuminating discussions of the Duhem–Quine thesis (Earman, 1996, ch. 3).

To address these issues, several model assumptions have to be made. In the case of the variety-of-evidence thesis, for example, “more varied” has to be explicated in probabilistic terms. This can be done in different ways. Many Bayesian discussions of the variety-of-evidence thesis assume that the evidence is certain. But we saw already that this is not always the case. It speaks in favor of the Bayesian framework that it provides the tools to model more complicated testing scenarios. See Bovens and Hartmann (2003) for more realistic Bayesian models of the variety-of-evidence thesis and the Duhem–Quine thesis.

Bayesian confirmation theory connects naturally with empirical psychology as a wealth of work in the psychology of reasoning under uncertainty demonstrates. See, for example, Chater and Oaksford (2007, 2008) for sophisticated Bayesian models that account for empirical findings. Crupi et al. (2008) show how the presence of the conjunction fallacy (i.e. that experimental subjects assign a higher probability to a conjunction than to one of the conjuncts), as famously demonstrated in psychological experiments by Tversky and Kahneman (1983), can be explained in confirmation-theoretical terms. Crupi et al. (2007) argue on normative and experimental grounds for a specific measure of evidential support (the so-called Z-measure). And Bayesian confirmation theory provides a flexible framework to reconstruct specific episodes rationally from the history of science. While traditional epistemology does not have the resources to study such episodes, the Bayesian framework is ideally suited for these purposes (Franklin, 1990). It is also better-suited than some system of epistemic or doxastic logic – imagine trying to illuminate some scientific episode solely with “K” and “B” operators!

2. *Dynamics of Belief.* Traditional epistemology, with its focus on the analysis of knowledge, is relatively silent about questions of belief dynamics. If there is talk about belief change, it is generally assumed that it takes place on the basis of learned evidence that

is certain. Traditional epistemology shares this assumption with logical theories of belief revision such as the AGM theory (Gärdenfors and Rott, 1995). However, Jeffrey taught us that learning often does not come in the form of certainties. To address these cases of learning and belief change, philosophers as well as researchers in artificial intelligence have formulated new updating rules (such as Jeffrey conditionalization) and developed powerful tools such as the theory of Bayesian networks (Neapolitan, 2003).

3. *Applications.* Bayesianism has a symbiotic relationship with causation, and powerful algorithms have been developed to learn causal relations from probabilistic data (Korb and Nicholson, 2004; Pearl, 2000; Spirtes et al., 2001). These algorithms use the theory of Bayesian networks. A Bayesian network organizes a set of variables into a *Directed Acyclic Graph* (DAG). A DAG is a set of nodes and a set of arrows between some of the nodes. The only constraint is that there are no closed paths formed by following the arrows. A *root node* is a node with outgoing arrows only, and a *parent* of a given node is a node from which an arrow points into the given node. Each node represents a propositional variable, which can take any number of mutually exclusive and exhaustive values. To make a DAG into a Bayesian network, one more step is required: we need to specify the prior probabilities for the variables in the root nodes and the conditional probabilities for the variables in all other nodes, given any combination of values of the variables in their respective parent nodes. The arrows in a Bayesian network carry information about the independence relations between the variables in the network. This information is expressed by the Parental Markov Condition: *A variable represented by a node in the Bayesian network is independent of all variables represented by its non-descendent nodes, conditional on all variables represented by its parent nodes.* In the causal modeling literature, this condition is called the Causal Markov Condition.

4. *The Coherence Theory of Justification.* Confronted with the Cartesian skeptic, coherentists point out that when our belief systems hang together well, with their different parts

supporting each other, this is an indication of the truth of the systems (BonJour, 1985). However, the corresponding theory – the coherence theory of justification – suffers from several problems. Here are two. First, the theory is vague, as it is difficult to make precise what coherence is. Second, coherence is not necessarily truth-conducive. For example, fairy tales are made up, although the stories they tell may be highly coherent. Hence, the coherence of a set of propositions is at best truth-conducive *ceteris paribus*. But what goes in the *ceteris paribus* clause? This question is hard to address if we only have the toolbox of traditional epistemology. Bayesians can be of real help here. They have proposed and analyzed various measures of coherence and analyzed in detail under which conditions, if at all, coherence is truth-conducive (Bovens and Hartmann, 2003; Douven and Meijs, 2007; Olsson, 2009).

5. *Sources of Knowledge/Belief.* Traditional epistemology examines sources of knowledge and belief such as our senses, memory and testimony. All three have inspired Bayesian model-building. First, the uncertainty of the evidence from our senses has prompted the development of a more realistic updating rule than strict conditionalization – Jeffrey conditionalization. Second, conditionalization represents an idealized version of the epistemological role of memory – one who updates only by conditionalization never forgets – while Bayesian models of bounded rationality allow for memory loss (Mehta et al., 2004). Third, Bayesians have the resources to model the effect of combining the testimony of several witnesses (Bovens and Hartmann, 2003). There is also a growing literature on self-knowledge and self-location, as exemplified by the Sleeping Beauty problem (Elga, 2000), and whether “centered” information can rationally induce changes in opinions about “uncentered” propositions concerning how the world is.

## 8. AVENUES FOR FUTURE RESEARCH

In this closing section, we briefly point to some topics that we would like to see

addressed in future research. While some of them concern the relation of Bayesian epistemology to traditional epistemology and philosophy of science, others are internal to the Bayesian program.

1. *More bridges between Bayesian epistemology and traditional epistemology.* There is no harm in labor being divided between the two kinds of epistemology, as we have argued – but it would be all the better if they could become more cooperative enterprises. Think of some of the time-honored debates in traditional epistemology: skepticism, the analysis of knowledge, reliabilism, internalism vs. externalism. Think of some of the currently hot topics: contextualism, subject-sensitive invariantism, contrastivism, relativism, luminosity, “knowledge how” (as opposed to “knowledge that”), knowledge “wh-” (who, where, when, which). . . . Where are the counterpart debates in Bayesian epistemology? Going in the other direction, think of some of the time-honored debates in Bayesian epistemology: constraints on priors, updating rules, the extension of subjective probabilities to infinite spaces. And think of some currently hot topics: credences about chances (as in the Principal Principle), credences about one’s future credences (as in the Reflection Principle), updating credences on “centered” or “indexical” propositions. . . . Where are the counterpart debates in traditional epistemology? Each of these topics suggests a bridge waiting to be built.

To some extent, such progress awaits a better understanding of the relationship between traditional claims about belief/knowledge and Bayesian claims about degrees of belief, which is still controversial. As we have said, we are not sanguine about the prospects of a reduction in either direction, although reduction is surely not the only way to offer illumination. And, even if the two epistemologies continue on separate tracks, still developments in one can provide inspiration or heuristic guidance for the other.

2. *More bridges between Bayesian epistemology and philosophy of science.* Bayesianism started as a confirmation theory. And, indeed, the formal machinery to address confirma-

tion-theoretical questions is highly developed. Our ambitious goal, however, should be to develop a full-fledged Bayesian philosophy of science. Here is an incomplete list of questions that should be addressed to this end:

- (i) Which stance does Bayesianism take in the realism debate? Is it neutral to the debate, or does it favor a version of scientific realism or antirealism (Douven, 2005; Earman, 1996)?
- (ii) Can scientific theory change be understood in Bayesian terms (Earman, 1996)?
- (iii) Can Bayesianism help to characterize the overall structure of science? Is it epistemically advantageous to aim for unified theories (Myrvold, 2003)?
- (iv) Can a Bayesian reading of Inference to the Best Explanation be given? Lipton (2004) argues that explanatory considerations are encoded in the likelihoods, and not in the priors.
- (v) How can scientific idealizations be understood in Bayesian terms? Addressing this question is important as idealizations are ubiquitous in science. The trouble begins when we attach a prior probability of zero to an idealized (hence false) statement. The posterior is then arguably also zero, which arguably renders Bayesianism useless.

3. *Bayesian Social Epistemology.* Bayesian epistemology, as we have presented it so far, shares one important feature with traditional epistemology: it is individualistic, i.e. it is concerned with one agent, who has beliefs and who updates her beliefs in the light of new evidence. However, the doxastic unit could well be a community comprising several individuals, or more. Kuhn (1962) argued that it is the entire scientific community that accepts or rejects a paradigm. Or think of a jury that has to come up with a consensual verdict in a murder case. In light of examples such as these, a new field – social epistemology – has been established (Goldman, 1999; Kitcher, 1993). While much work in this field is informal, formal tools have recently been developed that address issues in social epistemology.

Especially noteworthy is the work on judgment aggregation (List and Puppe, 2009) that comprises investigations inspired by the discursive dilemma. Bovens and Rabinowicz (2006) and Hartmann and Sprenger (2009) have given epistemic analyses of various aggregation rules studied in this context. It is hoped that this work will eventually develop into a full Bayesian account of group judgment and group decision-making. Other topics of current interest include the debate about rational disagreement (Feldman and Warfield, 2009) and, related to this, theories about consensus and compromise formation.

To sum up: Bayesian epistemology is an exciting and thriving research program. There's plenty more work for Bayesians to do.<sup>1</sup>

# NOTE

- 1 We thank John Cusbert, Kevin Korb, Ralph Miles, Jonah Schupbach, Declan Smithies, and Jan Sprenger for very helpful comments.

# BIBLIOGRAPHY

- Armendt, B.: "Is There a Dutch Book Argument for Probability Kinematics?," *Philosophy of Science* (1980), 583–8.
- BonJour, L.: *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1985).
- Bovens, L. and Hartmann, S.: *Bayesian Epistemology* (Oxford: Oxford University Press, 2003).
- Bovens, L. and Rabinowicz, W.: "Answers to Complex Questions: An Epistemic Perspective," *Synthese* 150 (2006), 131–53.
- Chater, N. and Oaksford, M.: *Bayesian Rationality: The Probabilistic Approach to Human Reasoning* (Oxford: Oxford University Press, 2007).
- Chater, N. and Oaksford, M. (eds): *The Probabilistic Mind: Prospects for Bayesian Cognitive Science* (Oxford: Oxford University Press, 2008).
- Crupi, V., Tentori, K. and Gonzalez, M.: "Bayesian Measures of Evidential Support: Theoretical and Empirical Issues," *Philosophy of Science* 74: 2 (2007), 229–52.
- Crupi, V., Fitelson, B. and Tentori, K.: "Probability, Confirmation and the Conjunction Fallacy," *Thinking and Reasoning* 14 (HYPERLINK

- "<http://www.informaworld.com/smpp/title~content=g792699768~db=all>" 2) (2008), 182–99.
- Diaconis, P. and Zabell, S. L.: "Updating Subjective Probability," *Journal of the American Statistical Association* 77: 380 (1982), 822–30.
- Douven, I.: "Wouldn't It Be Lovely: Explanation and Scientific Realism," *Metascience* 14 (2005), 338–43.
- Douven, I. and Meijs, W.: "Measuring Coherence," *Synthese* 156 (2007), 405–25.
- Earman, J.: *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory* (Cambridge, MA: MIT Press, 1996).
- Edwards, W., Lindman, H. and Savage, J.: "Bayesian Statistical Inference for Psychological Research," *Psychological Review* 70 (1963), 193–242.
- Eells, E. and Fitelson, B.: "Measuring Confirmation and Evidence," *Journal of Philosophy* 97 (2007), 663–72.
- Elga, A.: "Self-Locating Belief and the Sleeping Beauty Problem," *Analysis* 60: 2 (2000), 143–7.
- Eriksson, L. and Hájek, A.: "What Are Degrees of Belief?," *Studia Logica* 86, Special Issue Formal Epistemology, ed. I. B. Fitelson (2007), 183–213.
- Feldman, R. and Warfield, T. (eds): *Disagreement* (Oxford: Oxford University Press, 2009).
- Fitelson, B.: "The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity," *Philosophy of Science* 66, Proceedings (1999), S362–78.
- Franklin, A.: *Experiment, Right or Wrong* (Cambridge: Cambridge University Press, 1990).
- Gaifman, H. and Snir, M.: "Probabilities over Rich Languages," *Journal of Symbolic Logic* 47 (1982), 495–548.
- Gärdenfors, P. and Rott, H.: "Belief Revision," in *Handbook of Logic in Artificial Intelligence and Logic Programming, Vol. IV, Epistemic and Temporal Reasoning*, ed. D. M. Gabbay, C. J. Hogger and J. A. Robinson (Oxford: Oxford University Press, 1995), 35–132.
- Glymour, C.: *Theory and Evidence* (Princeton, NJ: Princeton University Press, 1980).
- Goldman, A.: *Knowledge in a Social World* (Oxford: Clarendon Press, 1999).
- Good, I. J.: Letter to *The American Statistician* 25 (1971), 62–3.
- Hájek, A. and Hall, N.: "Induction and Probability," in *The Blackwell Guide to the Philosophy of Science*, ed. P. Machamer and M. Silberstein (Oxford: Blackwell, 2002), 149–72.
- Hartmann, S. and Sprenger, J.: "Judgment Aggregation and the Problem of Tracking the Truth," *Synthese* (2009).

- Howson, C. and Urbach, P.: *Scientific Reasoning: The Bayesian Approach* (Chicago, Ill.: Open Court, 2006).
- Jaynes, E.: *Probability Theory: The Logic of Science* (Cambridge: Cambridge University Press, 2003).
- Jeffrey, R.: *The Logic of Decision*, 2nd edn (Chicago, Ill.: University of Chicago Press, 1983).
- Jeffrey, R.: "Probable Knowledge," in *Probability and the Art of Judgment* (Cambridge: Cambridge University Press, 1992).
- Joyce, J.: "A Nonpragmatic Vindication of Probabilism," *Philosophy of Science* 65 (1998), 575–603.
- Joyce, J.: *The Foundations of Causal Decision Theory* (Cambridge: Cambridge University Press, 1999).
- Joyce, J.: "Bayes' Theorem," in *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), ed. E. Zalta.
- Kitcher, P.: *The Advancement of Science: Science without Legend, Objectivity without Illusions* (New York: Oxford University Press, 1993).
- Kolmogorov, A. N.: *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Berlin: Springer, 1933). English translation: *Foundations of the Theory of Probability* (New York: Chelsea, 1950).
- Korb, K. and Nicholson, A.: *Bayesian Artificial Intelligence* (Boca Raton, FL: CRC Press, 2004).
- Kuhn, T.: *The Structure of Scientific Revolutions* (Chicago, IL: University of Chicago Press, 1962).
- Lewis, D.: "A Subjectivist's Guide to Objective Chance," in R. Jeffrey (ed.), *Studies in Inductive Logic and Probability*, Vol II (Berkeley, CA: University of California Press, 1980), 263–93.
- Lewis, D.: "Why Conditionalize?," in *Papers in Metaphysics and Epistemology* (Cambridge: Cambridge University Press, 1999).
- Lipton, P.: *Inference to the Best Explanation* (London: Routledge, 2004).
- List, C. and Puppe, C.: "Judgment Aggregation: A Survey," in P. Anand, C. Puppe and P. Pattanaik (eds), *Oxford Handbook of Rational and Social Choice* (Oxford: Oxford University Press, 2009).
- Mehta, N., Rajiv, S. and Srinivasan, K.: "Role of Forgetting in Memory-Based Choice Decisions: A Structural Model," *Quantitative Marketing and Economics* 2: 2 (2004), 107–40.
- Myrvold, W.: "A Bayesian Account of the Virtue of Unification," *Philosophy of Science* 70 (2003), 399–423.
- Neapolitan, R.: *Learning Bayesian Networks* (Upper Saddle River, NJ: Prentice Hall, 2003).
- Nozick, R.: *Philosophical Explanations* (Cambridge, MA: Harvard University Press, 1981).
- Olsson, E.: *Against Coherence: Truth, Probability, and Justification* (Oxford: Oxford University Press, 2009).
- Pearl, J.: *Causality* (New York: Cambridge University Press, 2000).
- Ramsey, F. P.: "Truth and Probability," in R. B. Braithwaite (ed.), *Foundations of Mathematics and Other Essays* (London: Routledge, 1931), 156–98; reprinted in D. H. Mellor (ed.), *F. P. Ramsey: Philosophical Papers* (Cambridge: Cambridge University Press, 1990).
- Savage, L. J.: *The Foundations of Statistics* (Chichester: John Wiley, 1954).
- Shogenji, T.: "The Degree of Epistemic Justification and the Conjunction Fallacy," *Synthese* (2009).
- Sosa, E.: "How to Defeat Opposition to Moore," *Philosophical Perspectives* 13 (1999), 141–53.
- Spirtes, P., Glymour, C. and Scheines, R.: *Causation, Prediction and Search* (Cambridge, MA: MIT Press, 2001).
- Tversky, A. and Kahneman, D.: "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment," *Psychological Review* 90 (1983), 293–315.
- Van Fraassen, B.: "Belief and the Will," *Journal of Philosophy* 81 (1984), 235–56.
- Williamson, J.: *Bayesian Nets and Causality: Philosophical and Computational Foundations* (Oxford: Oxford University Press, 2005).
- Zynda, L.: "Old Evidence and New Theories," *Philosophical Studies* 77 (1995), 67–95.

ALAN HÁJEK AND STEPHAN HARTMANN





## Part II

# Twenty Epistemological Self-profiles



## Robert Audi

Epistemology as I conceive it is not only the theory of knowledge but also the theory of justification and indeed of rational belief. I take justification to entail rationality where both notions are applicable, but I deny the converse. In this sketch, however, I cannot address rationality specifically. But I take it to have the same kinds of grounds as justification, and much of what I say about justification will hold for rationality.

### THE GROUNDS OF JUSTIFICATION AND KNOWLEDGE

A full-scale epistemology should provide (among other things) an account of knowledge and justification, of how they are grounded, of their sources, and of their interconnections. I have supplied many central elements of such an account in *Epistemology* (2003), others in *The Architecture of Reason* (2001), and still others in *The Structure of Justification* (1993) and elsewhere. Let me begin with grounds.

The basic grounds of justification are *broadly* experiential. My sensory impressions of shape and sound, for instance, justify my believing that there is print before me and traffic noise outside. This belief could be justified, yet false. Since, by contrast, I cannot know the proposition in question without its being true, what grounds knowledge differs from what grounds justification in a way that reflects this contrast. In my view, knowledge that *p* (propositional knowledge) must be grounded in some appropriate connection to the fact that *p* (or whatever it is in virtue of which *p* is true). This conception, however, does not entail that knowledge must have any experiential ground, and I have argued that in special cases it need not (2003, ch. 8; and, for memory knowledge, 1995). If, as this view implies, knowledge is possible without justification, we face the interesting question of how the two are related and why paradigms of knowledge are also cases of justified belief. "Justification, Truth, and Reliability"

(1988) addresses this question, and I continue to reflect on it.

### MODERATE RATIONALISM VERSUS EMPIRICISM

If the basic grounds of justification are experiential, one might think that empiricism, at least about justified belief, is implied. If reflection and intuitive apprehension are properly considered experiential, it is not. What justifies me in believing that (e.g.) nothing is round and square is a certain kind of understanding of that proposition itself. I consider this proposition self-evident (though not analytic, for reasons given in 2003, ch. 4), and I regard the self-evident as the base of the *a priori*. A truth is self-evident, I hold, iff it is such that any adequate understanding of it meets two conditions: (a) in virtue of having that understanding, one is justified in believing the proposition (i.e., has justification for believing it, whether one in fact believes it or not); and (b) if one believes the proposition on the *basis* of that understanding of it, then one knows it. Regarding what constitutes the appropriate understanding, there is much to be said. I make a number of basic points in (e.g.) "Self-Evidence" (1999).

Given that a true belief justified in the *a priori* way just outlined is *based on* understanding, it represents *a priori* knowledge. (I have explicated this basis relation in "Belief, Reason, and Inference," 1986). Suppose, however, that one takes the relation of the fact that *p* to knowledge that *p* to be causal. Can the kind of understanding that grounds *a priori* justification for believing *p* causally connected with the fact that *p* in an appropriate way? The question is especially urgent if *a priori* facts obtain in virtue of properties or relations of abstract entities, which are commonly and not implausibly considered to be outside the causal order. I present a partial solution to this problem in "Skepticism about the *A Priori*" (2007). Here I can say only that I do give causal relations an important place in grounding *a priori* justification and knowledge, and that I do not consider

accounting for that place inconsistent with a moderately rationalist view of a priori justification or knowledge.

#### THE STRUCTURE OF KNOWLEDGE AND JUSTIFICATION

If we take the common-sense view that we each possess much knowledge and justified belief, it is important to have an account of the structure of a body of justification beliefs or knowledge. For convenience I will speak only of knowledge here. The leading positions on this structural question have been foundationalism and coherentism. I have defended a moderate foundationalism. In outline, this is the view that the structure of a body of knowledge is foundational in the sense that any indirect (hence non-foundational) knowledge it contains depends on direct (thus in a sense foundational) knowledge, such as perceptual or a priori knowledge. A similar thesis holds for justified belief. Both theses are explicated in (1993b) and (2003, ch. 7). The position is moderate rather than Cartesian because (among other things) it does not require (a) indefeasibility in foundational elements, (b) deductivism for transmission of knowledge or justification from foundations to superstructure, or (c) denying an epistemologically important role for coherence – particularly in the theory of concepts and their acquisition.

This is a good place to emphasize that throughout my epistemological theorizing I seek to be psychologically realistic – for instance, to avoid representing knowers in ways that imply their having an infinite set of beliefs or making far more inferences than they do. Given that a body of knowledge is also a body of belief, I have formulated psychological counterparts of epistemological foundationalism and coherentism and compared these – for instance, in “Psychological Foundationalism” (1978) and (2003, ch. 7). In arguing for foundationalism I have formulated the venerable epistemic regress argument in a way that allows coherentism a good hearing but also facilitates comparison of foundationalism and coherentism in relation to their commitments regarding

our psychology. Here, quite apart from its purely epistemological advantages, foundationalism seems superior. There are apparently empirically basic sources of belief; and there is much plausibility in taking a kind of *psychological foundationalism* to hold regarding the structure of our belief system. Epistemological and psychological foundationalism cohere well with each other; psychological coherentism (formulated in my 1978) is not plausible, and epistemological coherentism seems at best to have psychologically implausible implications. This does not imply that coherentism is epistemologically unimportant; but, far from grounding justification, coherence is apparently consequential on the same elements as justification (1993c).

Grounds of knowledge and justification are provided by their basic sources. A *basic* source of knowledge or justification is roughly one that grounds it without epistemic (as opposed to conceptual) dependence on any other source. I will speak here only of justification, though the same sources play essential roles in grounding knowledge. In many writings I argue that sensory experience (the phenomenal element in perception) is a basic ground of justification, where this implies that it grounds non-inferential justification for belief. I take memorial and introspective experience to be similarly basic for justification. These three represent the traditional experiential basic sources. Reflection and intuitive apprehension have traditionally been described as constituting a source in “reason”. I have used “reason” for a similar purpose, but have noted (e.g. in 2001, ch. 1) that we can conceive ratiocinative experience, with (occurrent) intuition as a momentary case, as a broadly experiential source. I take these four sources to have an a priori status in relation to justification and knowledge. For instance, that my clear and steadfast impression of a chair gives me *prima facie* justification for believing there is one here is not empirical. I have not posited other basic sources, but also do not represent the concepts of justification or knowledge as such that there *cannot* be any other basic sources.

## CONCEPTIONS OF JUSTIFICATION AND KNOWLEDGE

Given the framework now briefly sketched, it is natural to hold what I call a well-groundedness conception of justification (and indeed of rationality, though I have argued, e.g. in 2001, that rationality, which I conceive in contrast to *irrationality*, has weaker grounding requirements than justification, which I conceive in contrast to what is *unjustified*). Well-groundedness may be inferential or non-inferential (and I provide for coherence – and especially incoherence – to play a significant role in understanding well-groundedness). This conception is developed in (2001), (2003), and many of my papers.

Knowledge I view as, roughly, true belief based in the right way on the right kind of ground. This is not meant as an analysis but as a conception that can be developed in detail depending on one's theories of appropriate grounds and of what it is for a belief to be based on them in the right way. My account of justification is not only experientialist but also internalist, in a sense I characterize in (2002) and elsewhere. In part because I do not take knowledge to entail justification, my account of knowledge is externalist (2003, ch. 8).

## EPISTEMOLOGY AND THE PHILOSOPHY OF MIND

The psychological realism that my epistemology preserves goes with an enduring concern with the philosophy-of-mind side of epistemology. Given this concern – spurred in part by my work in the theory of action and practical reasoning (of which 1993a and 2006 are representative) – I have been particularly concerned with the notion of belief. Although there is a concept of knowledge on which the existence of knowledge that *p* does not entail belief that *p* (a point explained in 2003, ch. 9), “standard” cases of knowledge are instances of belief. My epistemological writings do not contain an overall account of belief (something I provided in 1972), but they do say much about what constitutes

belief and about the distinction – still too often overlooked – between *dispositional beliefs* *dispositions to believe* (1982, 1994). We do not believe every proposition we are *disposed to believe*, in a sense implying that we would immediately assent to it upon considering it – say, that there are fewer than 1,000,003 words on this page. If so, we do not know it, either (prior to forming the belief). Similarly, though I was justified *in believing* this proposition before it occurred to me, I did not *justifiedly believe* it until it did – and I then believed it on an appropriate ground.

If belief – and indeed knowledge – is ascribed too liberally, so is inference. Do you instantly infer that this page is not maroon when you consider the question – say, from seeing that it is white and believing it cannot be both colors? And how can we tell? It turns out to be very difficult to say what an inference is (and the same holds for reasoning). I addressed this in (1985a), and distinguished there and elsewhere between a belief's being simply *based on* another, in the sense important for foundationalism, and its arising from one or more others by *an inference*, conceived roughly as a kind of conscious tokening of an argument (2003, ch. 6; 2006, chs 4 and 5). Once again, I seek to represent justification and knowledge in a psychologically realistic way, positing no more inferences than needed to account for them.

In several places (e.g., 1983a, 1985a, and 2001) I defend a causal-basis requirement for knowledge and indeed for justified belief. This is important not only for the foundationalism-coherentism controversy and for understanding justification and knowledge, but also for psychological realism. Our beliefs are normally based on experience, or on other beliefs, in a partly causal sense that implies an explanatory relation and is crucial for securing an appropriate connection between knowledge and the facts in virtue of which what is known is true. The connection must not be “wayward”, and (1985a) provides a partial theory of the right kind of causal connections. I have also distinguished (e.g., in 1985b, 2001a, and 2006) between, on the one hand, the justifiedness and rationality of beliefs – which are subject to certain causal



requirements connecting beliefs to their experiential grounds – and, on the other, the justifiability and rationalizability of beliefs, which need not meet these requirements.

#### SPECIAL TOPICS IN EPISTEMOLOGY

I have painted in broad strokes here and must continue to. But it is appropriate to conclude with a sketch of some of the special topics in epistemology that have engaged me and remain among my research interests. I shall not discuss further the special topics already mentioned in sufficient detail: self-evidence and the *a priori*; internalism and externalism; inference and the basis relation; the concept of believing; and consciousness as a basis of justification and knowledge.

*Perception* is epistemologically fundamental. I offer a theory of it in (2003, ch. 1) and treat it in many places, including later chapters in (2003) that compare perception, as a source of belief, with introspection (fruitfully conceived, up to a point, as inner perception), with memory, and with inference. I distinguish three cases: simple perception – say, seeing *a tree*; objectual perception – for instance, seeing a tree *to be blue*; and propositional perception – say, seeing *that a tree is blighted*. Each case is connected differently with belief and knowledge. My overall account of perception gives due weight to its phenomenal element, but remains a direct realism. The account is, however, open-ended as to what *kinds* of phenomenal elements can represent the propositions perceptually known. This point, together with others in my overall position, is important for religious epistemology. I have accordingly contended (e.g., in 2003, ch. 10) that a kind of theistic perception is possible. Still, if knowledge is possible without justification, then that possibility does not by itself entail the possibility of *justified* theistic beliefs. Even if none could be perceptually justified, however, given my view that rationality is a weaker normative status than justification, it would not follow that none could be *rational*. If, as I also hold, faith that *p* does not entail belief that *p* (though it is incompatible with believing

*p* false), then religious epistemology must countenance a wider range of cognitive attitudes than has been traditionally supposed.

*Memory* has also been a major topic in my epistemology. It is a basic source of justification, but not of knowledge (1995; 2003, ch. 2). Here a major problem is to account for the grounding of memorial knowledge where the knower simply has the knowledge with no phenomenal basis in a memory impression. Given how psychologically subtle a memory impression can be, I doubt that there are as many such cases as some apparently think. In any event, my double-barreled view is fruitful here: if knowledge does not entail justification, the way is open to argue that one can remember that *p* (a way of knowing it) without having justification for *p* (1995).

*Social epistemology* is also a major concern in my work, especially in my papers on testimony (e.g., 1997, 2004a). I consider testimony an *essential*, not basic, of justification and knowledge. It is not a *basic source* because it has an operational dependence on perception. It is none the less a source of *basic beliefs* (1997). There may even be a sense in which it is an *a priori* source of justification, though here I distinguish between its grounding of justified *belief* and its grounding of justified *acceptance* (2004b).

*Parallels between epistemology and the theory of practical reason.* Many of my epistemological works are written with the general theory of rationality in mind, hence with a sense of parallels between theoretical and practical reason (developed most extensively in 2001) and of how epistemology bears on ethics. Moral epistemology is a major topic in my work, especially in the past decade. I argue that certain moral principles are plausibly considered self-evident, particularly if we distinguish hard (strongly axiomatic) from soft self-evidence (1999; 2004b, chs 1 and 2). I also argue for a distinct source of justification that is naturally called moral experience (1998). This is important both for the justification of singular moral judgments and for the kind of understanding required for knowledge or justified belief of basic moral principles.

*Skepticism.* I have spoken as if I presupposed that there is a great deal of knowledge and justification belief. I do, but I also take skepticism seriously. In responding to it I have stressed a distinction between *defeasibility*, which I grant for all or virtually all of our justification, and *epistemic dependence*, which I deny holds for all our justified beliefs (1983b). The former is a negative dependence, a vulnerability to defeat by possible counter-evidence; the latter is an actual positive dependence on some ground(s). I have shown how skepticism can gain unearned plausibility from missing this distinction and, related to this, how skepticism raises a dialectical regress problem that bedevils certain kinds of defenses of common sense. Another distinction crucial in dealing with skepticism is between *rebutting* it, in the sense of showing that the case for it is inconclusive, and *refuting* it, which requires showing that skeptical theses are false and hence that there is knowledge or justified belief of certain kinds. I have argued that the prospects for rebuttal are good and that a plausible case can be made – at least on broadly rationalist, foundationalist assumptions – for refutation of some skeptical positions (2003, ch. 10).

A number of the positions sketched here are still being developed in my ongoing work, and in this endeavor, as in the past, I take this occasion to acknowledge with deep appreciation both the excellent example of my teachers and the astute responses of many colleagues and students in the field. The epistemology I have constructed is experientialist, rationalist, realist, foundationalist, common-sensist, and consonant with a plausible philosophy of mind. I shall continue to develop it both as a wide-ranging epistemology and as a basis for work in normative ethics and the theory of practical reason.

#### WRITINGS BY ROBERT AUDI

- "Skepticism about the A Priori," in *Handbook of Skepticism*, ed. J. Greco (Oxford: Oxford University Press, 2007).
- Practical Reasoning and Ethical Decision* (London: Routledge, 2006).
- The Good in the Right: A Theory of Intuition and Intrinsic Value* (Princeton, NJ: Princeton University Press, 2004a).
- "The A Priori Authority of Testimony," *Philosophical Issues* (Nous supplementary volume) 14 (2004b), 18–34.
- Epistemology: A Contemporary Introduction to the Theory of Knowledge*, 2nd edn (London: Routledge, 2003).
- "An Internalist Theory of Normative Grounds," *Philosophical Topics* 29 (2002), 19–46.
- The Architecture of Reason: The Structure and Substance of Rationality* (Oxford: Oxford University Press, 2001).
- "Self-Evidence," *Philosophical Perspectives* 13 (1999), 205–28.
- "The Axiology of Moral Experience," *Journal of Ethics* 2 (1998), 355–75.
- "The Place of Testimony in the Fabric of Knowledge and Justification," *American Philosophical Quarterly* 34 (1997), 404–22.
- "Memorial Justification," *Philosophical Topics* 23 (1995), 31–45.
- "Dispositional Beliefs and Dispositions to Believe," *Nous* 28 (1994), 419–34.
- Action, Intention, and Reason* (Ithaca, NY: Cornell University Press, 1993a).
- "The Foundationalism–coherentism Controversy: Hardened Stereotypes and Overlapping Theories" (1993b), in *The Structure of Justification* (Cambridge: Cambridge University Press, 1993).
- The Structure of Justification* (Cambridge: Cambridge University Press, 1993c).
- "Justification, Truth, and Reliability," *Philosophy and Phenomenological Research* 49 (1988), 1–29; reprinted in *The Structure of Justification* (Cambridge: Cambridge University Press, 1993).
- "Belief, Reason, and Inference," *Philosophical Topics* 14 (1985a), 27–65; reprinted in *The Structure of Justification* (Cambridge: Cambridge University Press, 1993).
- "Rationalization and Rationality," *Synthese* 65: 2 (1985b), 159–84; reprinted in *The Structure of Justification* (Cambridge: Cambridge University Press, 1993).
- "The Causal Structure of Indirect Justification," *Journal of Philosophy* (1983a), 398–415; reprinted in *The Structure of Justification* (Cambridge: Cambridge University Press, 1993c).
- "Foundationalism, Epistemic Dependence, and Defeasibility," *Synthese* 55 (1983b), 119–39; reprinted in *The Structure of Justification* (Cambridge: Cambridge University Press, 1993c).
- "Believing and Affirming," *Mind* 91 (1982), 115–20.

"Psychological Foundationalism," *The Monist* 62 (1978), 592–610; reprinted in *The Structure of Justification* (Cambridge: Cambridge University Press, 1993c).

"The Concept of Believing," *The Personalist* 53 (1972), 43–62.

## Laurence Bonjour

### 1. INTRODUCTION

My epistemological thinking has focused primarily on a set of related issues that I take to be the central issues of epistemology, both historically and substantively. Do we have good reasons for thinking that our various beliefs about the world (primarily about the common-sense world of material objects, including its history and scientific nature) are true? If we have such reasons, what is their detailed nature and structure, and how ultimately cogent are they?

In recent epistemology, issues in this vicinity have been standardly formulated in terms of the concepts of *knowledge* and *epistemic justification*; and my own discussion has often been couched in these terms. I have lately come to think, however, for reasons that are briefly suggested in the final section of this self-profile, that such formulations are inessential and, to a significant degree, misleading. What the great historical epistemologists (here I have especially Descartes and Locke in mind) were asking more than anything else was, I believe, just the questions I have mentioned, even though they often couched them in terms of knowledge (though rarely, if ever, in terms of the somewhat technical notion of justification).

Much recent epistemological discussion has been devoted to the issue between internalist and externalist theories of justification and knowledge. Here I shall only say that, as I understand the issues listed above, externalist views are simply irrelevant to them: externalism may offer conceptions of knowledge or of justification or perhaps even (in what I can understand only as a stipulated sense) of a reason for a belief, but *having* a reason is an essentially internalist notion.

### 2. THE REGRESS PROBLEM, COHERENTISM, AND FOUNDATIONALISM

One main problem that arises in the attempt to understand how we could have good reasons for our beliefs is, of course, the epistemic regress problem. It is easy to understand how one propositional belief (or a set of such beliefs) might provide reasons for thinking that a further propositional belief is true. But this notoriously just raises the question of what reason there is for the prior belief(s), and so on, leading to a regress that threatens to be vicious.

Since neither beginning with beliefs for which there is no reason, nor an actually infinite regress of reasons, nor a circle of reasons appears acceptable, this problem was usually taken – correctly, as I now think – to be the basis of an argument for a foundationalist account of reasons: one according to which, as I would now put it, there are "basic" beliefs for which there are, in some way, good reasons that do not take the form of further propositional beliefs, thus ending the regress.

But the problem is to understand how this is possible – something about which many versions of foundationalism have little to say. My early thinking was shaped by the conviction that it is *not* possible, a conviction that was largely based on a dilemma derived from Sellars: If one says that the reason for a basic belief is sensory experience, the question becomes how the character of that experience is apprehended by the believer, which it seemingly must be if it is to yield a reason for him. If it is apprehended in propositional terms, then this apprehension seems to constitute just a further propositional belief, itself as much in need of a reason, and the regress continues; whereas if the apprehension of experience is non-propositional and so presumably non-conceptual in character, then it is difficult (and might, in something of a leap, be claimed to be impossible) to understand how the apprehended content can provide a reason for a propositional belief (*SEK*, pt I<sup>1</sup>).

Having convinced myself that foundationalism is impossible, at least for reasons of an empirical sort, and being unwilling either

to abandon the project of finding reasons for our beliefs or to adopt various ways (such as contextualism or externalism) of sweeping such issues under the rug, I found myself, as philosophers do with amazing frequency, backed into a corner and forced to take seriously a view that initially seems quite implausible: a coherence theory of the reasons for empirical beliefs.

I have no space here to recount the details of my attempts to make sense of coherentism. It would involve the ideas of non-linear justification, explanatory coherence, a would-be coherentist account of sensory observation (beliefs caused by the external world via the senses but justified by internal coherence), the attempt at a “metajustification” of the appeal to coherence, and of course the infamous “doxastic presumption” (SEK, pt II). There are, as it now seems, many, many reasons why the whole project was doomed to failure, but the main one is the problem of making sense of the believer’s own access to his system of beliefs and its coherence: a problem that can, I believe, only be dealt with in an essentially foundationalist way, thereby eliminating any hope that coherentism can provide a genuine alternative to foundationalism (EJ, ch. 3).

At this point, I was genuinely stymied: I still saw no way to make foundationalism work but also became ever less hopeful that some way could be found to salvage coherentism. Fortunately, simultaneous work on the a priori (see the next section) came indirectly to my rescue. In thinking about the kind of access to the content of a thought that is needed to provide a basis for an a priori insight, I came to see that making sense of the very idea of a conscious mental state of any sort requires that there be a kind of awareness of its content that is *constitutive* of the state as a conscious state (“built in”, as I like to say), rather than apperceptive or second-order in character. And it is this sort of awareness of content that, I claim, provides a basis for a solution to the Sellarsian dilemma.

The basic idea is that this intrinsic awareness of content, since it is what makes that state the very conscious state that it is, is not an awareness for which a reason can

meaningfully be demanded. And yet, being an awareness that is present in consciousness, it can itself provide the conscious subject with a reason for a belief, at the very least for a belief that merely describes the experiential content in conceptual terms (but see further below). The belief for which a reason of this sort exists will thereby be foundational in the sense specified above: there will be a reason for thinking that it is true, but that reason will not depend on a further propositional belief that would in turn require a further reason (EJ, ch. 4). (This means, of course, that it is the “built-in” awareness of content that is really the ultimate foundation.)

I remain convinced that an account along these lines is the only viable form of foundationalism for empirical beliefs, and that the sort of experience that it must appeal to, at least primarily, is immediate, non-conceptual sensory experience. But this still leaves two difficult problems outstanding. The first and more obvious one is the venerable problem of how from information about the content of immediate sensory experience it is possible to infer to claims about the external material world. This would be easier if, as is often assumed, a characterization of the content of such experience in physical object terms (something like “I am having an experience as of a large brown dog”) could be unproblematically taken for granted. But if, as I believe, any such characterization, however natural it may be, in effect simply begs the most interesting question, then things become much more difficult. For what is then seemingly required is an inference to the material world from the more narrowly phenomenal features of experience – in the case of vision, from something like patterns of color or shape. Showing that such an inference is cogent is a task for which I cannot claim so far to have provided more than the merest sketch, with the basic idea being that material object claims offer the best *explanation* of the features of such experience (mainly fairly large-scale patterns of experience over time) (EJ, ch. 5).

The other, somewhat less obvious but equally serious problem is whether a foundationalist view along these lines requires beliefs

explicitly about the phenomenal features of sensory experience as an intermediate step between the “built-in” awareness of the content of sensory experience and eventual beliefs couched in material object terms. It is (as suggested above) beliefs of this sort for which the “built-in” awareness of content most straightforwardly provides reasons, but the obvious problem is that we do not ordinarily, if ever, have such explicitly phenomenal beliefs, at least not enough of them of a sufficiently detailed sort. I have lately come to think, however, that it may be plausible to hold that, if the truth of a material object belief indeed provides the best explanation for an extended pattern of non-conceptual experience, that experience may serve as a reason for that belief directly, without the need for a belief couched in phenomenal terms. The crucial point is that I am aware of the specific character of such experience merely by having it, so that, if that specific, non-conceptual character is indeed best-explained by supposing that a certain sort of physical object is present, this can be recognized and serve as a reason for a corresponding physical object belief with no essential need for a conceptual description of the experience itself.<sup>2</sup>

### 3. A RATIONALIST CONCEPTION OF THE A PRIORI

In contrast to the radical changes in my views concerning empirical reasons, my position on a priori reasons has remained essentially unchanged. In opposition to the radical empiricism that denies the very existence of a priori reasons and the moderate empiricism that insists that they are confined to claims that are analytic, I have defended the traditional rationalist view that a priori insight yields genuinely cogent reasons for accepting non-analytic claims about the world.

My main argument for such a view is extremely simple, but also, I believe, quite compelling. It begins with two premises that only a very extreme skeptic can deny: first, that experience or observation provides, in some way, direct reasons for accepting certain

empirical claims; and, second, that the class of broadly empirical claims for which we have good reasons is much larger than that for which there are reasons of this directly experiential sort. (The former class would include at least claims about unobserved situations in the past and present, claims about the future, claims about unobservable entities of various sorts, and claims about laws of nature.) The argument is then that we can have a good reason for some claim in this former class only if we have a logically prior good reason for a conditional proposition having some claim (or conjunction of claims) supported by directly experiential reasons as the antecedent and the claim in question as the consequent. And the reason for this conditional proposition can only be a priori, since it is obviously not a matter of direct experience (*IDPR*, ch. 1).

This argument provides an immediate reason for rejecting radical empiricism: it shows that on a radical empiricist view we can have good reasons for at most directly observational claims, a deeply skeptical result (*IDPR*, ch. 3). The application to moderate empiricism is almost as immediate: if extremely implausible reductionist theses are set aside, the needed conditionals linking directly experiential claims with those that are not directly experiential are clearly not analytic in any epistemologically useful sense of that term.

Though these arguments are dialectically compelling, they would have little persuasive power were it not for the existence of many plausible examples of claims for which there are a priori reasons – claims that cannot, I have argued, be accounted for in the moderate empiricist way. Indeed, a further argument against moderate empiricism is that the moderate empiricist view can seem even initially plausible in relation to such examples only by shifting tacitly and illegitimately among different and non-equivalent senses of the key term “analytic” (*IDPR*, ch. 2).

Here I can sketch only briefly my rationalist account of a priori reasons. The concept of an a priori reason has in my view two sides, one negative and one positive, but both equally essential. Negatively, an a priori



reason is one whose rational force does not derive from *experience*, either directly (as in sense perception) or indirectly (as by inference of any sort whose premises derive their acceptability from experience). Here the paradigm cases of experience are the various sorts of sense experience, including such things as kinesthetic experience. But, in opposition to a number of recent discussions, I would argue that *introspective* awareness of one's mental states should also count as a variety of experience, simply because such awareness is far closer in its essential character to sensory experience than it is to the kind of insight we have into, for example, mathematical claims.

Turning to the positive aspect of the concept of an a priori reason, the traditional view, which I believe to be essentially correct, is that in the most basic cases such reasons result from direct or immediate insight into the truth, indeed the necessary truth, of the relevant claim. A derivative class of a priori reasons results from similar insights into the derivability of a claim from one or more premises for which such a priori reasons exist or from a chain of such derivations. And a *partially a priori* reason may result from an a priori insight into the derivability of a claim from others established on broadly empirical grounds.

Here it is important to be clear that insights of this sort are not supposed to be merely brute convictions of truth, on a par with hunches that may be psychologically compelling. On the contrary, a priori insights purport to reveal not just *that* the claim in question must be true but also, at some level, *why* this is so. They are thus putative insights into the essential nature of things or situations of the relevant kind, into the way that reality in the respect in question *must* be. But, contrary to the most standard historical views, the idea of an a priori reason does not imply either: (i) that experience could not *also* count for or against the claim in question; or (ii) that an a priori reason could not be overridden by experience; or still less (iii) that an a priori reason renders the claim certain or infallible.

One other essential point about the nature of a priori insights should be briefly mentioned. For a variety of reasons, but most fundamentally because of the role that such insights are supposed to play in deductive inference, it is often and quite possibly always a mistake to construe them as *propositional* in form. The problem here, pointed out long ago by Lewis Carroll, is that at least in the most fundamental sorts of cases (think here of *modus ponens*), the application of a propositional insight concerning the cogency of such an inference would require either a further inference of the very sort in question or one equally fundamental, thereby leading to a vicious regress. Instead, I suggest, the relevant logical insight must be construed as non-propositional in character, as a direct grasping of the way in which the conclusion is related to the premises and validly flows from them.

#### 4. JUSTIFICATION AND KNOWLEDGE

I close with the discussion promised earlier of the concepts of justification and knowledge. About justification, there is relatively little to say. *Epistemic justification* is a technical philosophical concept that has no very explicit standing in ordinary thought. It can be understood or developed in a variety of ways, and there is in my judgment simply no point in arguing about which of these understandings is correct (as internalists and externalists have so long been doing).

Despite its undeniable presence in common-sense thought, on the other hand, the concept of *knowledge* seems to me much more problematic. What I am increasingly skeptical about is a conception of knowledge that I take to be essentially an invention of philosophers, and one whose very coherence is highly questionable. According to this concept, knowledge requires at least belief, truth, and some sort of condition of "justification" or "warrant". Whatever this last condition amounts to, it is supposed to be relevant to the truth of the belief in question, but to fall short of guaranteeing truth. And the problem

is that no very clear specification of this lesser level of “justification” is ever offered. Nor does there seem to be any that would be consistent with the intuitive role of knowledge, most importantly the seeming fact that one can deductively infer further conclusions from one’s knowledge and thereby come to know them as well.

My conviction, which I have no space either to elaborate or to defend here, is that all of this is largely a mistake, that there simply is no well-defined concept having these features – and that it is the myth of such a concept that has given rise to many of the characteristic problems and issues that have been the concern of recent epistemologists: the Gettier problem, doubts about closure, the lottery paradox, contextualist views, etc. My suggestion is that none of these problems is genuine. In my view, the only well-defined concept of knowledge is belief for which there is a conclusive reason, one that does guarantee truth – and that this just is the common-sense concept. Of course, it is obvious on good philosophical grounds that we have little knowledge, thus understood. The ordinary person is, I think, unlikely to find this very alarming, but philosophers, determined to make the epistemic claims of common sense literally true, have invented a concept that is really a monstrosity – devouring much time and effort and philosophical cleverness and giving almost nothing back in return (*EJ*, ch. 1).

## NOTES

- 1 While abundant references seem inappropriate here, a few may prove useful, especially where there are points that cannot be developed very fully. Most of these will be inserted in the text, using the following abbreviations for my books: “*SEK*” for *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1985); “*IDPR*” for *In Defense of Pure Reason* (Cambridge: Cambridge University Press, 1998); and “*EJ*” for *Epistemic Justification* (jointly with Ernest Sosa; Oxford: Blackwell, 2004).
- 2 See my paper “Are Perceptual Beliefs Properly Foundational?”, forthcoming in a *Festschrift* for Robert Audi.

## Stewart Cohen

## THE BASIC VIEW

For the past fifteen years, my main focus in epistemology has been developing and defending Contextualism – a view about the nature of knowledge ascriptions. In this essay, I shall outline the main components of my view along with what I take to be the principal problem the view faces.

The best way to understand Contextualism is to see what thesis it denies. Consider two knowledge ascriptions. Smith says, “Mary knows her car is parked in Lot 2”. Jones says about the same Mary at the very same time, “Mary does not know her car is parked in Lot 2”. A view we can call “Invariantism” says that these two knowledge ascriptions necessarily conflict. If what Smith says is true, then what Jones says is false. And if what Jones says is true, then what Smith says is false.

This is precisely what Contextualism denies. According to Contextualism, it is possible that both what Mary says is true and what Smith says is true. This is possible because the content of “know” and its cognates, and so the content of ascriptions of the form “S knows P”, can vary across contexts of ascription. So if the content of “Mary knows her car is parked in Lot 2” is different at Smith’s context than it is at Jones’s context, then it is possible that both what Smith says and what Jones says is true. That is to say, it is possible that their knowledge ascriptions do not conflict.

So, according to Contextualism, the truth-value of a knowledge ascription is sensitive to the context of ascription. This follows from the claim that the content of “know” is sensitive to the context of ascription. This raises the question: What precisely about the content of “know” can vary across contexts?

On my view, in different contexts, the content of “know” can involve different standards for justification. Consider an ascription of the form “S knows P”. How strong must S’s justification for P be in order for that ascription to be true? While Invariantist views might differ on how they answer this question, all Invariantists hold that, whatever the level required, it is the same for all

contexts of ascription. But, on my view, the level of justification required for the truth of an ascription of the form “S knows P” can vary with the context of ascription. Certain contexts will require more justification than others. Thus, the reason that Smith’s and Jones’s ascriptions can be compatible is that the standard of Jones’s context can require Mary to have a stronger level of justification than the standard of Smith’s context. So it may be that Mary’s level of justification is sufficient for Smith’s claim, “Mary knows that her car is parked in Lot 2”, to be true. But that same level of justification may not be sufficient for the truth of that same ascription at Jones’s context. This would allow that Jones can be speaking truly when saying “Mary does not know her car is parked in Lot 2”.

#### MOTIVATION FOR THE VIEW

There are two kinds of motivations for Contextualism. One has to do with how to treat certain kinds of puzzling cases. The other concerns how to respond to certain skeptical paradoxes.

##### *Puzzling Cases*

Consider an example of the sort of puzzling case I have in mind:<sup>1</sup>

Mary and John are at the LA airport contemplating taking a certain flight to New York. They want to know whether the flight has a layover in Chicago. They overhear someone ask a passenger Smith if he knows whether the flight stops in Chicago. Smith looks at the flight itinerary he got from the travel agent and responds, “Yes, I know – it does stop in Chicago.” It turns out that Mary and John have a very important business contact they have to make at the Chicago airport. Mary says, “How reliable is that itinerary? It could contain a misprint. They could have changed the schedule at the last minute. Smith does not know the plane stops in Chicago.” John agrees with Mary, and they decide to check with the airline agent.

What should we say about this case? Smith claims to know the plane stops in Chicago. Mary and John claim that Smith does not

know the plane stops in Chicago. Mary and John seem to be using a stricter standard for how strong one’s justification must be in order to know than Smith is using. According to Invariantism, only one of the standards can be the correct standard and so only one of them can be speaking the truth. But it is difficult to defend either the view that Smith speaks the truth or the view that Mary speaks the truth.

Consider the view that Smith speaks the truth when he says, “I know the plane stops in Chicago.” So Smith’s weaker standard is the correct standard. If that view is correct, then it is puzzling exactly how John and Mary should think of their situation. Surely, given the importance of their meeting, they are rational in not relying on the flight itinerary to determine whether the plane stops in Chicago. But it is hard to square this fact with the claim that Smith knows the plane stops in Chicago. For then John and Mary could truly say, “Smith knows the plane stops in Chicago. But that’s not good enough for us; we need to check further.” To my ear, it is very hard to make sense of such a claim. Moreover, if what is contained in the flight itinerary is good enough for Smith to know the plane stops in Chicago, then it is good enough for John and Mary as well. But then it seems that John and Mary could truly say, “We know the plane stops in Chicago. But that’s not good enough for us; we need to check further.” Again, it seems very hard to make sense of such a claim.<sup>2</sup>

These considerations suggest that perhaps Smith speaks falsely when he claims to know the plane stops in Chicago. So perhaps the correct standard is John and Mary’s stricter standard. This would mean that Mary speaks the truth when she says, “Smith does not know the plane stops in Chicago.” But if we view matters this way, then we are forced to deny that we know much of what we ordinarily claim to know. In everyday life, we readily allow that we can come to know things on the basis of written information contained in things like flight itineraries. Moreover we could describe a case where even John and Mary’s stronger standard seems too weak. Suppose John has ingested a slow-acting

poison, and the only antidote is in Chicago. If that were the case, then John and Mary might not be willing to ascribe knowledge that the plane stops even on the basis of the testimony of the airline agent. They might insist on checking with the pilot.

So there are difficulties with either way of viewing the apparent conflict between Smith and Mary. Whether we say Smith speaks truly or whether we say Mary speaks truly, untoward consequences seem to ensue. It is at this point that the Contextualist suggests that perhaps we are wrong to assume that there is a fixed correct answer. Perhaps each standard is correct relative to the context at which it is employed. In Mary's context of ascription, a stronger standard is in effect than in Smith's context of ascription. This would allow that Smith speaks truly when he says, "I know the plane stops in Chicago," and Mary also speaks truly when she says, "Smith does not know the plane stops in Chicago."

### *Skeptical Paradoxes*

A second motivation for Contextualism is that it can provide a resolution of certain skeptical paradoxes. We can begin to construct such a paradox by considering what seems to be a mundane case of knowledge. Suppose that one hour ago, I parked my car in Lot 2 and I am now in my office. In such a situation I can surely know that my car is parked in Lot 2. But now consider the proposition that my car has not been stolen. Do I know that to be true? While I have very good inductive grounds for supposing that my car has not been stolen, those grounds do not seem sufficient for me to know that it has not been stolen.

If this is correct, then I know my car is parked in Lot 2, but I do not know it has not been stolen. The problem of course is that it seems compelling that I know my car is parked in Lot 2 only if I know it has not been stolen. This is a consequence of the deductive closure principle for knowledge.

DC: If S knows P, and S knows P entails Q, then S knows (or at least is in a position to know) Q.

This leaves us with a paradox – an inconsistent set of propositions, each of which seems true:

- (1) I know my car is parked in Lot 2.
- (2) I do not know my car has not been stolen.
- (3) I know my car is parked in Lot 2 only if I know my car has not been stolen.

How should we respond to this paradox? A skeptic will argue from (2) and (3) to the denial of (1). According to the skeptic, we are just wrong to think that I can know my car is parked in Lot 2. Of course there is nothing special about the particular propositions that constitute this paradox. For virtually any proposition like (1) that we intuitively find to be true, there is a corresponding proposition like (2) that we also find intuitively to be true. A few more examples:

I know I shall be in Chicago tomorrow, but I do not know that I shall not die of a heart attack before I leave.

I know that I see a zebra, but I do not know that I do not see a cleverly disguised mule.

I know I'll never get rich, but I don't know I'll lose the lottery.

So this kind of paradox in fact threatens much of what we ordinarily claim to know.<sup>3</sup>

How can we respond to this paradox in a way that avoids this kind of skepticism? We might try to argue from (1) and (3) to the denial of (2). But this view seems hard to sustain. Again, while I do have inductive evidence that my car has not been stolen, it seems wrong to go so far as to say that I know my car has not been stolen.

The other option is to try to deny (3) (and so deny DC). The problem with this way of proceeding is that it seems incoherent to suppose that I could know that my car is parked in Lot 2 and yet fail to know that it has not been stolen. DC strikes us as axiomatic.

So none of these options for denying one member of the inconsistent triad seems palatable. Of course, this just underscores that we are confronting a paradox. A successful resolution of the paradox must do more than argue that one of the propositions is false. This would leave us with the question of why

it seems intuitively to be true. A successful resolution of the paradox must not only deny one proposition from the inconsistent triad; it must also explain why we mistakenly thought the proposition was true.

This is just what Contextualism tries to do. According to Contextualism, our acceptance of (1) is explained by the fact that at everyday contexts the standards for justification are low enough so that (1) can be true at those contexts. At that same context (2) is false. Why, then, does (2) strike us as true? Because, when we consider (2), we shift to a context with stricter standards. And at that stricter context (2) is true. Moreover, at that stricter context (1) is false. So at no context are both (1) and (2) true. This means that (3) is true at every context, which reflects the apparent axiomatic status of DC. Contextualism can thus explain why each member of the inconsistent triad strikes us as true, while still holding that at every context at least one member of the triad is false.

Contextualists' use the same strategy for skeptical paradoxes involving global skeptical hypotheses:

- (4) I know I have a hand.
- (5) I do not know I am not a brain-in-a-vat.
- (6) If I know I have a hand, then I do not know I am not a brain-in-a-vat.

According to Contextualism, at everyday contexts (4) is true and (5) is false. At stricter contexts, (4) is false and (5) is true. And (6) is true at every context.

#### THE MECHANISM OF CONTEXT-SHIFTING

In my discussion of both the airport case and the skeptical paradoxes, I argued that otherwise puzzling phenomena can be explained by supposing that different contexts of ascriptions have different standards for knowledge. At ordinary, everyday contexts, the standards are relatively low, but at other contexts the standards are higher. What explains why the standards are higher at some contexts?

On my view, it is the salience of error possibilities that drives the standards up. Where

S believes P on the basis of e, an error possibility is a possibility consistent with e, but inconsistent with P. Let us return to the airport case. When Smith looks at his flight itinerary and says, "I know the plane stops in Chicago," he is not thinking about the possibility that the itinerary contains a mistake, or has changed, etc. At his context, no error possibility is salient. Thus, the standards are relatively low at his context of ascription. This explains why what Smith says is true.

Now consider John and Mary. They have a very important meeting in Chicago. This leads them to think about error possibilities, e.g., the possibility that the itinerary has changed. Because these kinds of error possibilities are salient, the standards for knowledge rise. Thus, when Mary says, "Smith does not know the plane stops in Chicago," what she says is true.

The salience of error possibilities also explains the contextual shifts that give rise to the skeptical paradoxes. Consider again our example. When we consider whether I know my car is parked in Lot 2, we are not initially thinking about any error possibilities. Thus, we are at an ordinary low standards context at which (1) is true. Given closure, at this same context, (2) is false. But when we actually consider (2) an error possibility becomes salient, viz., the possibility that my car has been stolen. This shifts us into a higher standards context at which (1) is false and (2) is true. So when we go from considering (1) to considering (2) an error possibility becomes salient which shifts us to a higher standards context. Once we are at the higher standards context, (1) will seem false. But the situation is not stable. At times, common sense reasserts itself and we shift back to a context where (1) is true and (2) is false. Positing these kinds of contextual shifts can explain why we often vacillate between skepticism and common sense.

#### SEMANTIC BLINDNESS

A central feature of the Contextualism I defend is that we mistakenly assume certain knowledge ascriptions conflict which in fact do not. So we mistakenly assume that "I know my



car is parked in Lot 2" uttered at a low standards context conflicts with "I do not know my car is parked in Lot 2" uttered at a high standards context. And in the airport case we mistakenly judge that what Mary says when she utters, "Smith does not know the plane stops in Chicago," conflicts with what Smith says when he utters, "I know the plane stops in Chicago." Thus, I am committed to the view that, although ascriptions of knowledge are context-sensitive, competent speakers can be unaware of this, and so can be misled by it. Although their knowledge ascriptions track the shifts in context, they are unaware that these shifts are occurring. John Hawthorne has referred to this alleged phenomenon as "semantic blindness".<sup>4</sup>

Some philosophers find this semantic blindness thesis to be implausible.<sup>5</sup> How could competent speakers be unaware of these contextual shifts? And perhaps it does seem strange to suppose that competent speakers could, in this way, be blind to the workings of their own language.

What can the Contextualist say in response? If there exists an uncontroversial case of this kind of semantic blindness in connection with another term in the language, then it will not be so implausible for the contextualist to argue that this kind of semantic blindness exists in the case of "know".

Consider, then, the term "flat". We readily allow that many of the surfaces that we encounter in our daily lives are flat. But, as Peter Unger has noted, by calling attention to microscopic irregularities that exist in all physical surfaces, many competent speakers can be led to judge that "No surface is flat".<sup>6</sup> Should we conclude that no surface is flat? That seems like a radical conclusion. On the other hand, it seems unduly harsh to judge that these speakers who claim that nothing is flat are just wrong. Exactly what kind of mistake are they making? These kinds of reflections will convince most that ascriptions of flatness are context sensitive. The standards for how flat a surface must be in order to be flat simpliciter can vary across contexts. In everyday contexts, the standards are such that our utterances of the form "X is flat" can be true. But when we are in the kind

of strict context induced by calling attention to microscopic irregularities our utterance "Nothing is flat" can be true as well.

But notice that when we are in the grips of flatness skepticism, i.e., when we are in a context with strict standards, we think that our previous utterances of the form "X is flat" were false. That is to say, we mistakenly judge that our utterance of "Nothing is flat" conflicts with our previous utterances of the form "X is flat". And this is precisely the same sort of mistake that the contextualist attributes to us in the case of "know". So it appears that competent speakers can be blind to the context-sensitivity of a term in their language. This lends credibility to the Contextualist's claim that competent speakers are semantically blind with respect to "know".

That being said, I must concede that there is an important disanalogy between "know" and "flat". Competent speakers, upon reflection, will readily accept that ascriptions of flatness are context sensitive. But the same does not seem to be true for ascriptions of knowledge. So the claim that ascriptions of knowledge are context sensitive requires that we posit a higher degree of semantic blindness than does the claim that ascriptions of flatness are context sensitive. And this would appear to be a theoretical cost for the Contextualist. For one wonders whether it is plausible to attribute such a high degree of semantic blindness to competent speakers.

Keith DeRose has argued that the semantic blindness hypothesis has no theoretical cost to the Contextualist.<sup>7</sup> He reports that when he presents cases like the airport case to students a significant number of them will respond by saying that the claims of Smith and Mary do not conflict. They will claim that both are true. But this is impossible if "know" has invariant truth conditions. Thus, the Invariantist will have to posit semantic blindness on the part of these students who mistakenly think that the claims do not conflict. And if the Invariantist is stuck with positing semantic blindness, then the Contextualist's posit of semantic blindness cannot be seen as a theoretical disadvantage.

## NOTES

- 1 This example is from Cohen, 1999.
- 2 Hawthorne (in 2004) develops an Invariantist view that he argues can account for the airport case. For my objection to Hawthorne's view, Hawthorne's reply and my reply back, see Cohen, 2005.
- 3 For an illuminating discussion of this kind of paradox, see Hawthorne, 2004.
- 4 *ibid.*
- 5 Schiffer, 1996.
- 6 Unger, 1975.
- 7 DeRose, forthcoming.

## BIBLIOGRAPHY

- Cohen, Stewart: "Knowledge, Speaker, and Subject," *Philosophical Quarterly* 55 (2005), 199–212.
- Cohen, Stewart: "Contextualism, Skepticism, and the Structure of Reasons," *Philosophical Perspectives* 13, *Epistemology*, ed. J. Tomberlin (1999), 57–89.
- DeRose, Keith: "'Bamboozled by Our Own Words': Semantic Blindness and Some Objections to Contextualism," *Philosophy and Phenomenological Research* (forthcoming).
- Hawthorne, John: *Knowledge and Lotteries* (Oxford: Clarendon Press, 2004).
- Schiffer, Stephen: "Contextualist Solutions to Skepticism," *Proceedings of the Aristotelian Society* 96 (1996), 317–33.
- Unger, Peter: *Ignorance: A Case for Skepticism* (Oxford: Clarendon Press, 1975).

## Earl Conee and Richard Feldman

Our work in epistemology, both jointly and individually, has focused on developing and defending evidentialism and on criticizing rival theories. In this essay we shall highlight the central theses of the version of evidentialism that we support, identify what we regard as the main problems the theory faces, and sketch our responses to some of those problems. Also, we shall present a few critical comments on rival theories and highlight the connections evidentialism has to some prominent epistemological issues.<sup>1</sup>

## I. EVIDENTIALISM

The basic thrust of our evidentialism is a platitude – the justification needed for knowledge has to do with evidence. More specifically and more contentiously, we hold that any of a person's doxastic attitudes is epistemically justified if and only if it fits the evidence that the person has.<sup>2</sup> When one's evidence is counterbalanced, suspension of judgment is the justified attitude. When the evidence better supports a proposition than its denial, belief is the justified attitude; when it is the denial that is better supported, disbelief is justified.<sup>3</sup>

More detailed versions of this evidentialist thesis can be formulated. They differ about exactly what counts as evidence, what it is for a person to have something as evidence, and what propositions or attitudes a particular body of evidence supports.

Our evidentialist thesis about knowledge is that justification, understood in evidentialist terms, is necessary for knowledge. Versions of evidentialism can differ about what degree of support is required for knowledge and what other conditions are required for knowledge.

More than minimal evidential support for a proposition is required to satisfy the justification condition for knowledge of the proposition. Our view is that the degree of support required to satisfy the justification condition is best-captured by *the criminal standard*: a belief is well-enough justified for knowledge provided that the believer has strong evidence that supports it beyond all reasonable doubt. One has evidence that supports a proposition beyond all reasonable doubt just in case one has strong evidence in support of the proposition and no undefeated reason to doubt the proposition.

We hold that a person's evidence is restricted to the person's mental states. Evidence in this sense amounts to truth-oriented reasons broadly conceived. The evidence one has consists in the reasons of this sort that one has to go on in forming beliefs. These reasons include current experiential states, such as those involved in perception, introspection,

and occurrent memory. In addition, the experiential states involved in rational thought can serve as evidence. The word “reasons” is sometimes taken to apply only to believed propositions, not to the sorts of states just mentioned. But, in our view, beliefs are only some of one’s evidence and they need evidential support to serve as evidence. Stored memories might also serve as evidence. It is difficult to specify just which memories do count as evidence a person has at a particular time. Evidentialist theories can differ about such details while remaining in the spirit of the view.

Any adequate version of evidentialism will identify two distinct epistemic evaluations. On the one hand, there is the doxastic attitude that is supported by the evidence one has. This is the *justified* attitude. But one might not adopt the attitude that is justified. When one does adopt the justified attitude, it might be adopted for the wrong reasons. The right reasons consist in justifying evidence. In our terminology, when one believes a proposition, but not on the basis of proper reasons, one has a belief that is not *well founded*. We hold that one’s having a well-founded belief in a proposition is required for one to know the proposition.<sup>4</sup>

We take evidentialism to be a version of epistemological internalism. Internalism, as we understand it, is the view that justification strongly supervenes on internal, or mental, facts. Consequently, if any two possible individuals are mentally alike, then they are justificationally alike.<sup>5</sup> Although we endorse internalism, our primary allegiance is to evidentialism. Some philosophers have argued that internalism requires more than the strong supervenience thesis.<sup>6</sup> They hold that internalism implies that one is justified in believing a proposition on the basis of a body of evidence only if one knows, or justifiably believes, that the evidence supports that proposition. For reasons we shall briefly describe below, we reject this sort of meta-level requirement. (For a defense, see Fumerton (1995).) However, we see no basis for resolving disputes about the precise extension of the technical philosophical term “internalism”, nor any reason to debate what internalism “really” is.

As we just noted, we deny that a person’s evidence supports a proposition only if the person justifiably believes the meta-level proposition that the evidence does support the proposition. Simple examples illustrate why. Suppose that someone familiar with trees sees a typical sugar maple in good light and, on the basis of this visual perception, believes that there is a tree there. In the typical case, this will be a well-founded belief. The person need not also believe anything to the effect that having a tree-ish visual experience justifies, or makes probable, that there is a tree present. The person need not have any meta-level beliefs like this at all. Perhaps, however, in the typical case the person does have a justification for these meta-level beliefs.

It is possible to develop a view similar to our evidentialism without endorsing the strong supervenience thesis. One such possibility relies on a weak supervenience thesis, holding only that two individuals in the same possible world who have the same evidence must be justificationally alike. In effect, such versions of evidentialism make the evidential support relation contingent. We think that this is a liability. Our main reason for preferring the strong supervenience thesis is that it accommodates a strong and significant intuition. Victims of deception intuitively can have justified beliefs no matter what proportion of their beliefs are false. In a typical skeptical scenario, the victims get the same evidence ordinary people actually have. By the strong supervenience thesis, the same beliefs are justified in spite of the deception making them mostly false. Given that ordinary people have numerous justified beliefs, it follows, using the strong supervenience thesis, that the victims’ beliefs are also justified.

There has been considerable discussion of *the truth connection* in recent epistemology.

As Stewart Cohen (1984) argues, there is some especially close and important connection between epistemic justification and truth. It is the kind of justification that is intuitively “all about truth” rather than practical matters. Evidentialism gives a good account of such a connection: the epistemic justification for believing a proposition consists in evidence

for its truth. Notably, this same connection exists when the justified belief is false. The view we endorse allows that most of one's justified beliefs are false. Hilary Kornblith (2003, pp. 591–2) argues that this is a defect because “all connection between justification and truth has been severed”. In our view, this feature of evidentialism is a virtue since, as the skeptical scenarios illustrate, it is possible for most justified beliefs to be false. Yet justification requires evidence of truth, and this is enough of a truth connection.

Our strong supervenience thesis implies that an epistemically justified doxastic response to a body of evidence is necessarily a justified response to that evidence. Externalists are bound to deny this. It is contingent that the external conditions accompanying the evidence make the doxastic response that fits the evidence one that externalists count as justified. A simple example will help to bring out the plausibility of the evidentialist view we prefer and to illustrate the truth lurking in the thought that the justified response to a body of evidence is a contingent matter.

Assume that you know that the *New York Times* is a reliable source of information about a particular topic. You see a report on that topic in the *Times* and you believe what it says. Our evidentialist theory allows this to be a justified response. It is, of course, a contingent fact that you have justification that the *Times* is a reliable source. There are other possible situations in which it would not be appropriate to believe what one reads there. The strong supervenience thesis permits this possibility. Your background evidence about the reliability of the *Times* makes the crucial difference. If the “source” of your justification is construed narrowly, to refer to just the report in the *Times*, then the fittingness of the belief to its source is a contingent matter. This can make it appear that the support you have for the belief justifies it only contingently. But, intuitively, that narrow basis is not justifying at all. Your basis for justified belief includes all the background factors that contribute to your evidence on the matter. Given this, it is plausible that the fittingness of your belief to its whole basis is not a contingent matter.<sup>7</sup>

Thus, the core of evidentialism is the idea that the sort of justification needed for knowledge of a proposition is provided by internally possessed evidence that bears a necessary epistemic support relation to that proposition.

## II. SOME QUESTIONS

A fully worked out version of evidentialism must answer some admittedly difficult questions. Three such questions arise from the brief account of the theory just given. The first question concerns the nature of the support that perceptual states supply to external world beliefs. For instance, one's experience of feeling warm often plays some central role in justifying one's belief that one is in a warm place. Does this experience justify in virtue of its having propositional content, or by some pure phenomenal character that it has? Whichever answer one gives, further difficult questions arise. If the experience has propositional content, exactly what is that content? How does the content relate epistemically to the phenomenal character of the state, on the one hand, and to the justified belief on the other? If instead the experience of feeling warm helps to justify the belief that one is in a warm place in virtue of its qualitative character, how do internal phenomenal qualities manage to provide reasons for a belief concerning external temperature?

These questions do not arise only for evidentialist theories. Any acceptable theory of justification must acknowledge that perceptual experiences can make a justifying contribution, and it must do this without imposing an unreasonable meta-level requirement. We think that asking how perceptual experience provides evidence places the focus on the most fundamental epistemic question about perception.

A second challenge for evidentialism concerns the evidential status of stored beliefs and other memories. It seems clear that deeply buried memories that could only be retrieved through extensive psychoanalysis do not bear on the epistemic status of one's current beliefs. However, some stored information does seem relevant. Thus, someone who has

come to have a method for justified identifications of oak trees by the shape of their leaves can have a justified belief that there is an oak tree present when seeing an appropriate shape. It is not necessary that the person consciously entertain a proposition to the effect that the shape is indicative of oaks, nor must the person think of any reasons supporting this proposition. However, the justification of the oak tree beliefs does depend on using the remembered method, rather than, say, guessing on the spur of the moment. The difficult issue is to specify in a general and informative way which memories affect justification and which do not.

A related issue has been raised by Goldman (1999). He points out that we often know propositions on the basis of memory without remembering the original source for that information. For example, a person might remember that George Washington was the first president of the United States without remembering where that fact was learned. It is possible that there is nothing in the person's current states that reveals a relevant difference between this memory and another one – say, that George Washington once visited the person's home town, which is the only persisting element of a story the person made up as a child. There is a clear epistemological difference between the two cases: even if the latter belief is true, it is not a case of knowledge. Goldman contends that internalists have a hard time explaining the difference between the two memories.

Evidentialism has the resources to make the needed epistemic distinctions in this area. It is an assumption of the example just described that there is nothing in the evidence available to the believer to differentiate in a relevant way the two beliefs about George Washington. Given this, we think that the two beliefs are equally well justified. The person would be unreasonable in giving up one while retaining the other without having a basis for doing so. However, since the memory that has as its source a made-up story happens to be true, the example is a Gettier case. That is, it is a justified true belief that is only coincidentally true, and thus is not knowledge.

Again, these problems about memory are not restricted to evidentialist theories. Any acceptable theory of justification must account for the distinction between those memories that do affect justification and those that do not.

A third difficult question concerns the basing relation. Suppose one's overall evidence supports a proposition, and one believes that proposition. What, exactly, is required for one to believe that proposition *on the basis of* proper evidence? Is believing on the basis of evidence some purely causal relation between the evidence and the belief? Is it some disposition to use the evidence in defense of the belief? The challenge is to explain the basing relation in a way that captures the distinction between those cases in which someone believes the right things *for the right reasons* and those cases in which the person believes the right things but not for a right reason (where *right* is epistemically understood).

In this case, too, the challenge exists regardless of one's approach to the nature of justification. No adequate theory can fail to accommodate this difference.

### III. EVIDENTIALISM AND SKEPTICISM

One significant virtue of evidentialism is that it simultaneously helps to explain the appeal of skepticism and provides the tools for a satisfying response to it. One way in which evidentialism explains the appeal of skepticism depends upon the strong supervenience thesis. People with our actual evidence might have lacked knowledge of the external world because they were victims of some sweeping deception. By the strong supervenience thesis, we are no better justified in our external world beliefs than these possible victims are. It might be plausibly (but mistakenly) thought that this shows that our actual evidence is not strong enough for us to have knowledge of the external world. The possibility of error allowed by our actual evidence establishes that there is a kind of weakness, or inconclusiveness, in our evidence. The evidence does not entail the truth of the belief. One can easily see why those worried by skepticism might worry that this weakness shows that we



lack the justification needed for knowledge. They are concerned that having inconclusive information about the truth of a proposition is incompatible with knowing the proposition. It is not just obvious that this is a mistake. This makes sense of the skeptical worries.

An evidentialist response to these skeptical worries depends upon the claim that the inconclusiveness in our evidence is compatible with its being sufficient to justify our beliefs. This response raises at least two related questions: how well justified must our beliefs be in order to qualify as knowledge (when the other conditions for knowledge are met); and how well justified are ordinary external world beliefs? Our answer to the first question, as noted earlier, is that the relevant standard is similar to the standard for conviction of the accused in criminal trials in the US. One must have evidence that places the proposition beyond reasonable doubt. Evidence allowing the possibility of error can place a proposition beyond reasonable doubt. Our account of this, in a few words, is that, when a proposition is strongly supported by one's evidence, reasonable doubt requires having evidence that a contrary possibility does in fact obtain. Concerning ordinary strongly supported external world beliefs, people often have no evidence for the truth of any contrary possibility. Many typical perceptual and memorial beliefs are thus beyond reasonable doubt, and can be instances of knowledge on this account.

Notably, many beliefs in scientific theories do not meet the criminal standard. But here the evidentialist view seems to get things exactly right. Scientific theories are often overthrown. Even for well-confirmed theories, there is often recalcitrant experimental evidence supporting their falsity. These facts make many theoretical beliefs subject to reasonable doubt. Thus, even when they are well justified, they fall short of knowledge.

It is sometimes seen as a virtue of rival theories that they provide a good basis for rejecting skepticism. This alleged virtue seems to amount to the idea that it would be good if we did have knowledge, and so a theory about knowledge that unquestionably implies that we do have knowledge (when coupled with

some safe assumptions about our place in the world) is therefore a good theory. But surely this is a mistake. Obviously unsatisfactory theories, such as the theory that knowledge is mere true belief, also have this implication. It is true that a good theory of knowledge must provide for the knowledge we actually have. But an optimal theory will also reveal knowledge to be the sort of thing whose existence is at least sensibly thought to be called into question by the possibility (or actuality) of errors. Evidentialism does this. In contrast, externalist theories and the versions of evidentialism endorsing a weak supervenience thesis do not. These theories make the justifying relation contingent, leaving it difficult to see why merely possible skeptical scenarios even raise a question about actual justification and knowledge.

#### IV. EVIDENTIALISM AND OTHER THEORIES AND ISSUES

##### *IVA. Reliabilism and other theories*

Among the prominent rival theories of justification are reliabilism, proper functionalism, virtue theory, and responsibilism.<sup>8</sup> In a nutshell, these theories go as follows. Reliabilism holds that a belief is justified provided it results from a belief-forming process that reliably leads to true beliefs. Proper functionalism holds that a belief is justified provided it results from the proper function of the believer's cognitive system. Virtue theory holds that a belief is justified when it results from the exercise of cognitive virtue. And responsibilism implies that a belief is justified just in case it is formed in an epistemically responsible way. (Able defenders of these theories have offered many subtle refinements which we must ignore in this brief discussion.)

Our fundamental view about proper functionalism, virtue theory, and responsibilism is that either they are roundabout statements of evidentialism or else they are subject to decisive counter-examples. Consider proper functionalism. Either a properly functioning system is one that responds properly to evidence or it is not. If the latter is the case, then



there are counter-examples to the theory. Proper function that delivers beliefs with weak supporting evidence leaves the beliefs intuitively unjustified; malfunction that delivers beliefs with good evidence yields beliefs that are intuitively justified. If proper function is responding properly to evidence, then the theory is equivalent to evidentialism. Evidentialism provides the more basic explanation of justification. What renders a belief justified is not, in the final analysis, the fact that it results from the proper function of the cognitive system but rather that it is based on justifying evidence. Essentially the same point applies to virtue theory and responsibilism.

Our principal doubt about reliabilism is of a different nature. We believe that the central problem with reliabilism is the generality problem. Reliability is a property of types of belief-forming processes, not of concrete sequences of events. However, every such actual sequence is an instance of many types. To have a theory that can even be assessed, there must be some specification of those types whose reliability determines the epistemic status of beliefs. There are many ways to specify those types but, in our view, none that yields a plausible account of epistemic justification.

#### *IVB. Contextualism*

Contextualism in epistemology is roughly the thesis that “knows” gives the sentences in which it occurs truth conditions that vary with some differences in the context of use. Epistemic contextualists typically contend that skeptical arguments are sound because the uses of “knows” in those arguments occur in contexts governed by highly demanding standards, while ordinary knowledge claims are often true, too, because the standards in ordinary contexts are sufficiently lax. We have written separately in opposition to contextualism (Conee 2005; Feldman 2001). We have argued that the view is inadequately defended as a semantic thesis and that it does not have any real advantage as a ground for objecting to skeptical reasoning. This opposition to contextualism is independent of our basic evidentialism. The relevant basic

evidentialist thesis is that knowledge requires justifying evidence. This thesis is compatible with there being a contextual variation in the strength of justification required for true “knowledge” attributions.<sup>9</sup>

Our more specific evidentialist proposal about justification for knowledge, the criminal standard, also allows contextual variation. The proposal is that knowledge-level justification for a proposition requires strong evidential support for the proposition and the absence of an undefeated reason to doubt it. The strength of evidence needed to count as “strong evidential support” could vary with context.<sup>10</sup> We see no sufficient reason to believe that there is any such contextual variation, but evidentialists can consistently be contextualists.

#### *IVC. Epistemic Deontology*

Some philosophers (e.g., Plantinga (1993), Goldman (1999)) have linked internalism (and evidentialism) closely with deontological conceptions of epistemic justification. These conceptions characterize justification in terms of obligations, duties, requirements, permissions, or praise and blame. We do not endorse any such analysis. To the extent that we understand such terms, we think that the deontological theories yield incorrect results. A proposition can be justified by one’s evidence without one thereby having any obligation to believe the proposition and without one being subject to blame for a failure to believe. The proposition may be so inconsequential that one has no obligation to turn one’s attention to it or the pain that would result from belief so great that the failure to believe is blameless. Furthermore, it is possible for one to be unable for psychological reasons to believe what is supported by one’s evidence. If this incapacity removes any duty to believe a supported proposition, then what’s justified differs from what’s dutiful. We grant that there may be uses of phrases such as “ought to believe” and “has a duty to believe” that are tacitly epistemically modified. These uses may be equivalent to “is justified in believing”. Perhaps that is enough to render true an anemic version of deontology.

## V. CONCLUSION

Our development of evidentialism to date leaves the theory pretty abstract. Important questions remain about the nature of evidence and about why a body of evidence justifies what it does. Still, we contend that our view correctly answers some fundamental epistemic questions and it correctly frames some fundamental epistemic issues.

## NOTES

- 1 Many of the ideas presented here are developed in our papers collected in Conee and Feldman, 2004.
- 2 We take the doxastic attitudes subject to epistemic evaluation to be belief, disbelief, and suspension of judgment. However, evidentialism can easily be adapted to views that recognize degrees of belief. That theory holds that the justified degree of belief is the one that matches the strength of evidential support.
- 3 Some versions of evidentialism hold that merely having evidence that better supports a proposition than its denial is not sufficient to make belief the justified attitude, and that something more decisive is needed to justify belief. Our view is that any modest overall support is enough to make belief the justified attitude.
- 4 In addition, knowledge requires truth, belief, and some condition adequate to deal with examples of the sort Gettier made famous. We shall not discuss solutions to the Gettier problem here.
- 5 Semantic externalism complicates matters. If semantic externalism is correct, then individuals who are internally alike can believe different propositions. Our view is that if this is correct, then epistemological internalism should be modified slightly. One possible modification begins by introducing the idea of *counterpart propositions*. If two possible individuals are internally alike, but external factors map their shared internal state on to different propositions, then those propositions are counterpart propositions for those individuals. Where the internal states do not yield different propositions, that proposition is a counterpart of itself for the two individuals. Epistemological internalism then holds that if two individuals are internally alike, then for any proposition and attitude that is justified for one of them, the same attitude towards its counterpart proposition is justified for the other.

We note that semantic externalism does not clearly require modifying the thesis that individuals who are *mentally* alike are justificationally alike.

- 6 See Feldman (2004) for discussion of what internalism is.
- 7 We hold that background evidence about the *Times* is required in a case such as this one. We do not hold that a belief based on a report in the *Times* is justified only if the believer consciously forms a belief about the reliability of the *Times*.
- 8 Sometimes these theories are offered as accounts of something other than epistemic justification.
- 9 Fantl and McGrath (2002) have objected to evidentialism on the grounds that it is inconsistent with a certain kind of contextualism about the justification needed for knowledge. We deny that there is any inconsistency. We note that the kind of contextualism that Fantl and McGrath consider is one in which pragmatic factors concerning the importance of the putative item of knowledge affect the truth conditions of the attribution.
- 10 There is no room for contextual variation in the rest of what is required to meet the criminal standard: having no undefeated reason to doubt. Having the slightest undefeated evidence against a proposition is enough to fail to meet this condition.

## BIBLIOGRAPHY

- Cohen, S.: "Justification and Truth," *Philosophical Studies* 46 (1984), 279–95.
- Conee, E.: "Contextualism Contested," in *Contemporary Debates in Epistemology*, ed. M. Steup and E. Sosa (Oxford: Blackwell, 2005), 47–56, 62–6.
- Conee, E. and Feldman, R.: *Evidentialism: Essays in Epistemology* (Oxford: Oxford University Press, 2004).
- Fantl, J. and McGrath, M.: "Evidence, Pragmatics and Justification," *The Philosophical Review* 111 (2002), 67–94.
- Feldman, R.: "Skeptical Problems, Contextualist Solutions," *Philosophical Studies* 103 (2001), 61–85.
- Feldman, R.: "In Search of Internalism and Externalism," in *The Externalist Challenge*, ed. R. Schantz (Berlin: De Gruyter, 2004), 143–56.
- Fumerton, R.: *Metaepistemology and Skepticism* (Lanham, MD: Rowman and Littlefield, 1995).

- Goldman, A.: "Internalism Exposed," *The Journal of Philosophy* 96 (1999), 271–93.
- Kornblith, H.: "Roderick Chisholm and the Shaping of American Epistemology," *Metaphilosophy* 34 (2003), 582–602.
- Plantinga, A.: *Warrant: The Current Debate* (Oxford: Oxford University Press, 1993).

## Fred Dretske

My views about knowledge have been shaped by my work on visual perception. Very early (*circa* 1966) I became convinced that if I knew anything about the world – that my wife is on the sofa, for instance – then one of the ways of knowing this is by seeing her there. If that doesn't qualify as a way of finding out where my wife is, nothing does. If that doesn't count as a way of knowing she is on the sofa, skepticism (about the external world) is true.

In *Seeing and Knowing* (1969) I tried to describe the conditions that convert seeing my wife on the sofa (I called this non-epistemic perception since one can see one's wife on the sofa without knowing she is there) into knowledge that she on the sofa. In developing this theory of perceptual knowledge I became convinced of two things. I am still convinced of them.

The first thing I learned is that perceptual knowledge, if we have it, derives from the circumstances in which one comes to believe, not one's justification for the belief. To put it in currently fashionable jargon, I learned:

- (I) *Externalism*: If skepticism is false, externalism is true.

If, as a result of seeing my wife on the sofa, I come to know she is there, this knowledge cannot depend on my justification for thinking she is there – at least, not if justification is understood (as most internalists understand it) as evidence I have, things I know to be true, that support – to a degree sufficient for knowing – that my wife is on the sofa. Seeing her there does not give me that kind of justification. It does not give me evidence for thinking the person on the sofa is not an imposter, tricked up to look like my wife, and

put there to deceive me. Or, if I actually am justified in thinking the person is not an imposter (why, after all, would anyone take the trouble to fool me in this way? it would be most unusual), my justification is certainly not strong enough for me to *know* that no such deception is taking place. In most circumstances, I have no evidence that bears on the question of whether an unusual kind of deception of this sort is taking place. I simply assume it isn't. I simply assume that things are (as they usually are) completely normal. The general sort of background regularities I depend on in order to justify this assumption do not justify my belief that nothing strange is occurring on this particular occasion and it is this belief, a belief about what is happening now, that would have to be justified for me to know that nothing funny was going on this time.

Bertrand Russell's skeptical hypothesis about the past – how do we know the entire world, complete with fossils, memory traces, history books, etc., was not created minutes ago? – illustrates this general point about justification and knowledge. It displays a fact, the fact that a certain possibility did *not* materialize, that knowledge of the past depends on that our evidence for what happened in the past is powerless to justify. If my knowledge that I had granola for breakfast this morning requires me to have evidence, a justification, that the world was not created minutes ago (complete with all the misleading evidence that it was not), then I am incapable of knowing, incapable (therefore) of remembering, what I had for breakfast. I don't have a justification – certainly not one good enough for knowledge – that Russell's conjecture is false. The hypothesis, after all, was deliberately constructed so as to deny me evidence for thinking it false. So, if I remember (hence, know) that I had granola for breakfast this morning, this knowledge must not require a justification (good enough for knowledge) for the existence of a morning – much less a morning in which I had granola.

This is externalism. In the case of Russell's hypothesis about the past, externalism in epistemology is the view that one doesn't

have to be justified in thinking that the world was not created in the way Russell imagines in order to remember – hence, know – facts about the past. It is the fact that Russell's conjecture is false, not one's justification for thinking it false, that allows one to remember (hence, know) what one had for breakfast. And it is the fact that the circumstances in which I see my wife on the sofa are completely normal (non-deceptive), not the fact that I am justified in believing this, that enables me to see who is on the sofa.

The second thing I learned is something about the external circumstances that make such knowledge possible. These circumstances must, in some yet-to-be-clarified sense of possibility, eliminate the possibility of error. If the evidence on which one's belief that P is based is evidence that one might have even when P is false, then such evidence isn't good enough to know that P is true. One can't know P is true and, at the same time, concede that one's evidence leaves open the possibility that P is false. That sounds inconsistent to me. If this is indeed so, then knowledge that is based on evidence (I allow for the possibility that some knowledge isn't) requires what I (in Dretske, 1971) called *conclusive reasons* to believe.

- (II) *Conclusive Reasons*: If S knows that P on the basis of evidence E, S would not have E (E would not have existed) unless P were true.

(II) gives expression to the external condition, the possession of conclusive reasons, on which knowledge depends. This is an external condition since one can have conclusive reasons for believing P and not know, not be able to know, one has them. One doesn't have to know that Barney wouldn't have said that P unless P were true. If, in point of fact, Barney would not have said P unless P was true, then a belief based on Barney's word is knowledge. You know because Barney told you, and Barney wouldn't have said so unless it was true. You don't have to know your honest well-informed teachers are honest and well informed in order to learn something from them.

(II) precludes knowledge when belief is based on (merely) high probabilities. If only one person is going to win the lottery, you can't know you are going to lose simply because, holding only a single ticket, the probability of your losing is 99.99 percent. You do, after all, have a ticket, and your ticket might win. You don't know it won't. You will probably lose, but you don't know you will. That, in fact, is why you bought a ticket in the first place. Being realistic, you thought you would probably lose but you weren't absolutely certain. You didn't know you would. Had you known it, or had you thought you knew it, you wouldn't have bought a ticket.

Years later, in *Knowledge and the Flow of Information* (1981), I described this external connection in informational terms: a belief that my wife is on the sofa qualifies as knowledge only if it is caused by the information that she is on the sofa. Given the way I conceive of information, the information-theoretic formulation came down to pretty much the same thing as the earlier conclusive reasons theory; an event, condition or signal that carries the information that P turns out (because it requires a contextualized probability of 1) to be a conclusive reason for believing P. Given its informational origin, a belief that P caused by the information that P can't be false. It eliminates all (relevant) possibilities of being wrong.

I have always thought that this condition on knowledge, assuming for the moment that something like it is both necessary and sufficient for knowledge, was strong enough to capture the distinction – obvious since at least Gettier (1963) – between justified true belief and knowledge. No matter how probable your evidence makes P, if this evidence leaves room for mistake, if your belief that P, given this evidence, might be false, then no matter how justified (reasonable) you are in accepting P you don't know that P. If (to take the lottery situation again), given your evidence, you might win, then you don't know you won't. The condition is also weak enough to be routinely satisfied and, thus, to constitute an answer to skepticism. I really can come to know it is 3:30 by asking Barney if Barney is the sort of chap who wouldn't have

said “3:30” in answer to my query about the time unless it was 3:30. That is not to say that the condition is weak enough for us to know it is routinely satisfied and, thus, weak enough to know that we know what time it is and, thus, weak enough to know that skepticism is false. No. It is weak enough to know what time it is by asking Barney or by looking at a clock and, thus, weak enough for skepticism to *be* false. Whether skepticism is false or not is another, an empirical, question. Would (not *could*, but *would*) Barney have said it was 3:30 if it wasn’t? Would (not *could*, but *would*) the hall clock have registered “3:30” if it weren’t 3:30? If the answers to these questions are “No”, and a skeptic (being a skeptic) will not purport to know what the answers are, then we have what seems to me a satisfactory answer to skepticism. We have a theory that tells us we might know. Whether or not we know is an empirical question. It depends on whether our reasons for belief are conclusive. It depends on whether we actually have information. Skepticism might be true; but then, again, it might not. I can live with this conclusion. I think, in fact, that there is much to be said for second-order skepticism – the view that we probably, at least we might, know a lot, but that we can’t know whether or not we do. We at least have an answer for the first-order skeptic, the skeptic who thought he could demonstrate that we don’t know.

It is important to recognize that this account of knowledge, though externalist in character, is a theory about the kind of evidence (reasons, justification) needed to know. In that respect it is a very traditional sort of theory. It belongs in the classical JTB (justified true belief) genre of epistemological theories. It must, therefore, to be distinguished from externalist accounts of knowledge that eschew evidence, reasons, and justification (for a belief) in favor of something like simple reliability of belief. Nozick (1981), for example, identifies knowledge with belief that tracks the facts where tracking is defined in terms of a dependency (expressed by subjunctive conditionals) between the belief that P and P. S knows P if his belief that P tracks P – if, in particular (I ignore part of his

requirement on tracking since I don’t see the need for it), S wouldn’t believe P unless P. Nothing is said here (in his account of tracking) about the evidence or reasons S might have for P. Indeed, perfectly reliable beliefs (ones that S wouldn’t have unless they were true) end up being knowledge even if S has nothing to show in the way of support or justification for them. It has always seemed to me that this takes externalism and the genuine insights it provides a step too far. It exposes reliability theories (I count my own theory as a reliability theory) to needless counterexamples. I mention only two.

Unreliable beliefs, beliefs that (in Nozick’s sense) fail to track the facts, can easily qualify as knowledge. An old and familiar example from Alvin Goldman illustrates the way this might happen. S has an isolated false belief about wolves. He thinks they are dogs. Otherwise his knowledge of canines is beyond reproach. S sees (in good light, ideal viewing conditions, etc.) a dachshund, a distinctive-looking and easily identified dog, and believes it to be a dog. Indeed, he knows it is a dog. There is nothing that looks like *that* that isn’t a dog. Yet S’s belief that it is a dog fails to track the fact that it is a dog since if the animal were not a dog – if it were a wolf, for instance – he would still believe it to be a dog (we may suppose that there are several wolves in the neighborhood, so its being a wolf is a relevant possibility). So S’s belief that it is a dog needn’t track the facts in order to be knowledge. What is necessary is that the evidence (on which that belief is based) track the facts. What is necessary is not that S wouldn’t believe it to be a dog if it weren’t, but that (in the case of perceptual knowledge) the animal wouldn’t look that way – in this case, like a dachshund – if it weren’t a dog. What we need are conclusive reasons, not reliable beliefs. I argued this point at some length, using a similar canine example, in Dretske 1969: 88–92.

There is also an example (unpublished as far as I know) due to Saul Kripke that favors a conclusive-reason (or information-based) account of knowledge as opposed to a simple reliability (tracking) theory. Suppose that in the relevant neighborhood there are fake barns but



no fake *red* barns, so that something might look like a barn without being a barn, but nothing would look like a red barn without being a (red) barn. Then it turns out one could track red barns without tracking barns, and on a simple reliability (or tracking) theory of knowledge one could know something was a red barn without knowing it was a barn. This result is an embarrassment even for someone (like Nozick) who is prepared to deny closure (the principle that you know all the things you know to be implied by what you know). The example, however, is not effective against a “conclusive reason” or information-theoretic style analysis since these theories are formulated not in terms of reliability of *belief*, but in terms of the reliability of evidence on which the belief is based. In Kripke’s example we need to ask: On what is the belief (that it is a barn) based? S knows it is a barn if that feature of the evidence causing S to believe it is a barn would not exist if it were not a barn. In the case of perception, if its looking like a red barn is what is causing S to believe it is a barn, then (assuming that its looking like a red barn is a conclusive reason for thinking it is a red barn) S has conclusive reasons to believe it is a barn: it would not look the way it looks – like a red barn – unless it was a barn. Hence, he knows it is a barn. If, on the other hand, it is merely the building looking like a barn that is responsible for S believing it is a barn (its color being causally irrelevant to S’s belief), then S fails to know it is a barn. Although S’s visual experience carries the information needed to know it is a barn (since it carries the information that it is a red barn), that isn’t the aspect of the experience that causes him to believe. Knowledge is *information-caused belief*, and in this second case the information didn’t cause the belief. (This theory clearly needs an account of what it means to say that information caused something. I tried to supply this in Dretske, 1981.)

This account of knowledge (as well as Nozick’s version) has a well-known and much-discussed consequence. It leads, quite naturally, to the denial of *closure*. Many philosophers have found this result unfortunate. That is putting it mildly. They have

thought it constituted a *reductio ad absurdum* of the theory (see, for example, Bonjour, 1987; DeRose, 1995; Feldman, 1999; and Fumerton, 1987). So I conclude by discussing the significance of this result (for details, see Dretske, 2005).

Closure is the principle that one knows, or at least one is automatically positioned to know (I hereafter ignore niceties in the expression of this principle), all the known logical consequences of things one knows. Surely, the defenders of this principle argue, if I know it is my wife on the sofa, and I know that if it is my wife on the sofa it can’t be an imposter, then I know, or at least I have all it takes to know, that the person on the sofa is not an imposter.

An externalist theory of the sort I have described, whether expressed in terms of conclusive reasons, information, or tracking, denies this. We sometimes know the logical consequences of what we know, and sometimes we know them, or are positioned to know them, *because* we know they are consequences of what we already know, but not *always*. Sometimes, it turns out, one can’t get knowledge that Q from knowledge that  $P \rightarrow Q$  and knowledge that P.

Why not? Why isn’t a conclusive reason for P automatically a conclusive reason for Q for anyone who knows that P implies Q? Why isn’t information that P necessarily information that Q for those who know P implies Q? Or, putting aside particular theories of what it takes to know, how can one not know that Q is true if one knows that P is true and knows that P implies Q? There can, after all, be no more secure grounds for belief. Given what you already know ( $P$  and  $P \rightarrow Q$ ), you can’t go wrong in accepting Q. What more could possibly be needed to know that Q is true?

What is needed is something that closure doesn’t necessarily give you: a *reason* to believe Q. I come to know my wife is on the sofa by a procedure – sense perception – that requires certain facts to be solidly in place, unquestioned, not in dispute. Certain facts are not relevant possibilities. If they are, if I can’t simply assume that deceptive demons and cunning hoaxes are not possibilities my evidence need rule out, possibilities I can ignore

for purposes of evaluating appearances, then I really can't see what I purport to see – that my wife is on the sofa. I can no longer see, just by looking, that she is there because these ways of her not being there would, by hypothesis, make things look exactly the same way to me. Confronted with these possibilities, my visual evidence is completely neutralized. I cannot *see* that these possibilities are not so. So if these possibilities are in play, if I can't simply assume, as I normally do, that these possibilities are not genuine, I can't see what I said (and thought) I could see – that my wife was on the sofa. So closure fails in these special (isolated) cases because the key premise, P, is not available to those who want to use it to infer Q. You don't get to know P if you are going to use it to infer the absence of those conditions, Q, you had to assume were absent in order to know P.

This is why one cannot use the rings in tree stumps to refute Russell's skeptical conjecture about the past. The rings in the stump indicate that the tree is 40 years old, yes, but you can't use this knowledge of the tree's age to know, by closure, that the past is real because the reality of the past was taken for granted in assessing what the rings indicate. This, too, is why you can't know you are not being tricked into falsely believing your wife is on the sofa from what you can plainly see to be the case – that she is on the sofa. When you know P and know  $P \rightarrow Q$ , you can certainly infer Q (e.g., that you are not being deceived) from what you know, yes, but if there is a question about whether or not you know that Q is true the answer is "No". At least you no longer have a reason (viz., P) to think it true. Closure is not an acceptable epistemological principle: you can know P is true and know  $P \rightarrow Q$  and still not be in a position to know, even with the help of deductive reasoning, that Q is true.

#### BIBLIOGRAPHY

- BonJour, L.: "Nozick, Externalism, and Skepticism," in *The Possibility of Knowledge*, ed. L. Luper-Foy (Lanham, MD: Rowman and Littlefield, 1987).
- DeRose, K.: "Solving the Skeptical Problem," *Philosophical Review* 104: 1 (1995), 1–52.
- Dretske, F.: *Seeing and Knowing* (Chicago, IL: University of Chicago Press, 1969).
- Dretske, F.: "Conclusive Reasons," *Australasian Journal of Philosophy* 49 (1971), 1–22.
- Dretske, F.: *Knowledge and the Flow of Information* (Cambridge, MA: MIT Press/Bradford, 1981).
- Dretske, F.: "The Case against Closure," in *Contemporary Debates in Epistemology*, ed. Ernest Sosa and Matthias Steup (Oxford: Blackwell, 2005).
- Feldman, R.: "Contextualism and Skepticism," in *Philosophical Perspectives*, 13, *Epistemology*, ed. James Tomberlin (Oxford: Blackwell, 1999), 91–114.
- Fumerton, R.: "Nozick's Epistemology," in *The Possibility of Knowledge*, ed. Steven Luper-Foy (Lanham, MD: Rowman and Littlefield, 1987).
- Gettier, E.: "Is Justified True Belief Knowledge?," *Analysis* 23 (1963), 121–3.
- Goldman, A.: "Discrimination and Perceptual Knowledge," *The Journal of Philosophy* 73 (1975) 771–91.
- Nozick, R.: *Philosophical Explanations* (Cambridge, MA: Harvard University Press, 1981).

#### Richard Foley

Two major movements are at play in my writings on epistemology, one of separation and the other of integration.

The separation movement consists in establishing the independence of the theory of justified belief from the theory of knowledge. Much of recent epistemology assumes that knowledge is justified true belief plus some condition to handle Gettier problems. On this assumption, the properties making a belief justified are by definition such that when a true belief has these properties it is a good candidate for knowledge. The theory of justified belief is thereby placed in service to the theory of knowledge and is consequently divorced from the everyday assessments we make of each other's opinions, which tend to emphasize whether one has been responsible in forming beliefs rather than whether one has satisfied the prerequisites of knowledge.

The remedy is to resist any quick assumption to the effect that knowledge is to be understood in terms of justified true belief. By the end of the epistemological enterprise,

interesting connections between knowledge and justified belief may well emerge, but it is needlessly constraining to stipulate from the start that there is a simple, necessary tie between the two.

The integration movement insists on treating issues of rational and justified belief in relation to other issues of rationality, the presupposition being that the way we understand the rationality of beliefs ought to be of a piece with the way we understand the rationality of plans, actions, decisions, etc. The movement has three main parts. The first involves articulating a general theory of rationality; the second shows how the notion of epistemically rational belief is embedded within this general theory; and the third introduces a notion of justified belief that is closely connected with the everyday notion of responsible believing.

A first step towards a general theory of rationality is to recognize that, although rationality is concerned with the effective pursuit of goals, being rational does not ensure success. Consider plans, for example. A plan can be rational even if it turns out badly, because it may be that no one could be reasonably expected to see that the plan was likely to have such consequences. Considerations such as these suggest a general schema of rationality:

It is rational for an individual S to X (where X can be a plan, action, intention, strategy, intention, decision, belief, etc.) if it is rational for S to believe that X would acceptably contribute to his or her goals.

To say that a plan would "acceptably contribute" to one's goals is to say its estimated desirability is sufficiently high, where estimated desirability is a function of what it is rational to believe about the probable effectiveness of X in promoting one's goals and the relative value of these goals.

Uninstantiated plans, decisions, actions, etc., can satisfy this schema. Even if I have not adopted a plan P, it may none the less be rational for me to believe that P would have acceptably contributed to my goals. If so, the unchosen P satisfies the above schema.

Similarly, a plan can conform to the schema even if it is adopted for the wrong reasons. Suppose it is rational for me to believe that plan Q would satisfy my ends, but I choose Q only because a Ouija board recommends it. Plan Q is rational in the relevant abstract sense (it is the plan that it is rational for me to prefer) even if its concrete realization (the process by which I came to choose Q) leaves much to be desired.

When assessing the rationality of a decision, strategy, action, etc., we can take into consideration all of the individual's goals or only a specific type of goal. For example, if it is rational for S to believe that action X is an effective way to promote his or her economic interests, we can say that it is rational in an economic sense for S to X, and we can say this even if, with respect to all of S's goals, X is not rational. Thus, the above schema can be further refined to reflect the fact that different senses of rationality can be defined in terms of different kinds of goals:

It is rational in sense Y for an individual S to X if it is rational for S to believe that X would acceptably contribute to his or her goals of type Y.

Distinguishing senses of rationality is especially critical when beliefs are at issue. Epistemologists are traditionally interested in a specific kind of rational belief, one that is concerned with what it is rational to believe in so far as the goal is now to have accurate and comprehensive beliefs. The everyday assessments we make about the reasonability of each other's opinions, by contrast, take into account that we have many goals, that these goals are often more important or at least more pressing than our intellectual goals, and that accordingly there are constraints on how much time and energy it is reasonable to devote to intellectual ends. Epistemically rational belief is thus an idealized notion, and as such it ordinarily is not directly relevant for the everyday assessments we make of each other's opinions. On the other hand, its idealized character is a theoretical advantage. Virtually all epistemologists agree that a condition of adequacy

of an account of epistemically rational belief is that it not make ineliminable use of any other notion of rationality, and it is the idealized character of epistemically rational belief that permits this to be a condition of adequacy. Moreover, just because epistemic rationality is to be understood without recourse to any other notion of rationality, it is a suitable theoretical ground for other notions of rationality.

The easiest way to see how this works is to notice that the above schema has an apparent drawback, in that it makes use of the notion of rational belief in order to understand the rationality of intentions, plans, decisions, etc. It thus seems to leave us within the circle of rationality-related notions we wish to understand. Epistemically rational belief provides the schema with an escape route from this circle. It provides, in effect, a theoretically respectable anchor for other notions of rationality. In particular, with an account of epistemically rational belief in hand, the above general schema of rationality can be further refined:

It is rational in sense Y for an individual S to X if it is epistemically rational for S to believe that X would acceptably contribute to his or her goals of type Y.

This schema is altogether general. It applies to all phenomena (plans, decisions, strategies, and so on) and to all forms of rationality (economic rationality, rationality all things considered, and so on). Indeed, epistemically rational belief is itself automatically an instance of this schema, as can be seen by inserting the epistemic goal into the general schema for "goals of type Y":

It is rational in an epistemic sense for an individual S to believe X if it is epistemically rational for S to believe that believing X would acceptably contribute to the goal of his or her now having accurate and comprehensive beliefs.

This, then, is the outline of a general theory of rationality, and it is one in which epistemically rational belief plays a grounding role. The general theory is neutral among rival theories of epistemically rational belief, but I also do

defend a particular theory, one that associates epistemic rationality with an idealized form of intellectual integrity.

Being epistemically rational, I propose, is a matter of having opinions and using faculties, methods, and practices capable of standing up to one's own most severe critical scrutiny in so far as one's goal is now to have accurate and comprehensive beliefs. For an opinion to pass this test, it must not be the case that one's other opinions can be used to mount what on reflection one would regard as a convincing critique of its truthfulness. Nor can it be the case that on reflection one has what one would regard as convincing considerations that undermine the case for the truthfulness of the opinion. Thus, not only must the opinion be in accord with one's other reflective first-order opinions; it must also be in accord with one's reflective second-order opinions about the kinds of faculties, methods, and practices that can be used to generate and revise opinions reliably.

I make no claim of exclusivity for this sense of epistemically rational belief. On the contrary, I insist that there are different senses of rationality, which being distinct are not competitors. I've already indicated that one way of distinguishing senses of rationality is in terms of different kinds of goals being sought, but in addition senses of rationality (and corresponding senses of reasons) can be distinguished in terms of the different perspectives from which an action, decision, plan, etc., can be assessed.

Sometimes we are interested in evaluating an action, decision, plan, etc., from an objective perspective, and it is appropriate to express these evaluations using the language of rationality. For example, if of the available options X is the one that in fact is likely to accomplish the decision-maker's goals, then there is an objective reason for her to prefer X, whether or not she or anyone else is aware of this. There are other contexts, however, in which we are interested in evaluating an action, decision, plan, etc., from the perspective of what a hypothetical reasonable person would do in the decision-maker's situation, where this hypothetical individual is not to be thought of as someone who knows with

certainly the consequences of the available options but rather as one who is in possession of widely available information about the probabilities of the options yielding various outcomes. In still other contexts, however, we are interested in evaluating a decision from the perspective of the very individual making the decision. We want to project ourselves into her situation and see the decision from her perspective. We thus bracket information and background knowledge that is available to us but was not available to her, and try to locate grounds for the decision that are appropriate given her perspective. If we succeed in discovering factors that identify her decision as satisfactory relative to information and background beliefs that she herself on reflection would endorse, we can express this finding by saying that we have shown why in at least one sense of reason she had reasons for her decision, and hence why in at least one sense of rationality it was a rational decision for her to make, even if it turns out to be mistaken. On the other hand, if we judge that, even given her limited information and background beliefs, she would have seen the decision's shortcomings had she been a little more cautious or reflective, we shall have shown in an especially convincing way why her decision was irrational. It was irrational not so much because it in fact failed to achieve her goals, and not so much because it was at odds with what a reasonably well informed person would have decided, but rather because it was at odds with what she herself would have thought appropriate had she been more thorough, more vigilant, or more discerning.

This notion of rational (and irrational) decision is the practical counterpart of the notion of epistemically rational (and irrational) belief that I defend, a notion that is to be understood in terms of the individual's own perspective on reflection on how best to satisfy the epistemic goal of now having accurate and comprehensive beliefs. Being rational in this sense is a matter of making oneself invulnerable to intellectual self-criticism to the extent possible, of living up to one's own deepest intellectual convictions and standards. Reliability is the goal but not a

prerequisite. Indeed, one can be rational in this sense even if one is a brain-in-a-vat whose opinions are massively unreliable. This notion of epistemically rational belief is thus distinct from reliabilist conceptions. Similarly, it is distinct from conceptions that understand epistemic rationality in terms of the intellectual practices of one's community. Although most people's deepest intellectual convictions and standards are at least loosely congruent with those of their social environment, this is not necessarily the case. And so it is possible for one to be rational in this sense even if one's opinions, methods, and assumptions are radically at odds with those of one's community, era, or tradition.

Epistemically rational belief is a matter of what it is rational to believe in so far as one's sole aim is to have accurate and comprehensive beliefs, whereas the everyday notion of responsible belief recognizes that non-intellectual goals place constraints on how much time and effort it is appropriate to devote to investigating and deliberating about an issue. Responsible believing can thus come apart from the idealized notion of epistemically rational believing, and yet it is the idealized notion that provides a theoretical anchor for the everyday notion.

Note that our everyday evaluations of each other's beliefs tend to be reason-saturated. We are interested, for example, in whether someone in forming her beliefs has been *reasonably* thorough in gathering evidence, and *reasonably* careful in assessing this evidence, and *reasonably* cautious about drawing conclusions from it. The standards of reasonability at work in these assessments are realistic ones. They reflect the fact that all of us have non-intellectual interests, goals, and needs, many of which are more important and pressing than our intellectual ends.

These kinds of considerations can be captured in a concept of justified belief that is closely associated with responsible belief rather than with knowledge. Following Alvin Plantinga, I reserve the term "warrant" for the property that turns true belief, absent Gettier problems, into knowledge. Justifiably believing a proposition, by contrast, is a matter of it being rational, all things considered,



for one to have acquired (and subsequently retained) the belief. More precisely, an individual justifiably believes *P* if it is epistemically rational for her to believe that the processes by which she has acquired and sustained the belief *P* have been acceptable, that is, acceptable given the limitations on her time and capacities, and given all of her goals. By “processes” I mean to include whatever methods, faculties, and skills are involved in the acquisition and sustenance of the belief. Thus, if an individual has an epistemically rational belief that she has spent an acceptable amount of time and energy in gathering and evaluating evidence about *P* and has also used acceptable methods, faculties, and skills in processing this evidence, her belief *P* is justified.

A related concept is that of non-negligent belief. Whereas justifiably believing a proposition *P* is a matter of it being rational to have acquired (and subsequently retained) the belief, non-negligently believing *P* is a matter of it not being irrational to have acquired (and subsequently retained) the belief. Often we are not in a good position to determine how we came to believe what we do. We may not remember, or perhaps we never knew. The beliefs in question are thus not justified, but there none the less may be something to recommend them; they may be non-negligent. In particular, an individual non-negligently believes *P* if it is not epistemically rational for her to believe that when all of her goals, needs, and constraints are taken into account she has spent an unacceptably small amount of time in gathering and evaluating evidence about *P* or that she has used unacceptable methods, faculties, and skills in processing this evidence.

A belief can be justified (or non-negligent) even if we have spent little or no time gathering evidence or deliberating about it. Many of our beliefs are acquired automatically (most perceptual beliefs, for example), and ordinarily it is epistemically rational for us to believe that this is an acceptable way to acquire such beliefs. Hence, the beliefs are justified. But, in addition, a belief can be justified (or non-negligent) even when we are in the possession of information that would have convinced us that what we believe is

incorrect, had we spent some time reflecting on this information. Given the topic’s insignificance, however, it may not have been appropriate to devote time to reflecting on it. If so, the belief does not meet the standards of epistemically rational belief, which are concerned with what it is rational to believe in so far as one’s only concern is to have accurate and comprehensive beliefs, even though it does meet the standards of justified or non-negligent belief, which are concerned with the degree of effort it is appropriate to devote to having accurate and comprehensive beliefs about a topic in light of its overall importance in one’s life.

The concepts of justified and non-negligent belief are likewise sensitive to the way in which intellectual standards can vary with social role. If it is your job but not mine to keep safety equipment in good working order, the intellectual demands upon you to have accurate beliefs about the equipment are more stringent than those upon me. My belief that the equipment is in good working order might be justified even if I have done little, if any, investigation of the matter. I need not have tested the equipment, for example. A cursory look might suffice for me, but this won’t do for you. The standards of justified belief are higher for you.

Social role can also be relevant when the issue at hand is primarily of theoretical interest. For example, my justifiably believing that the principle of conservation of energy is not violated in the beta decay of atomic nuclei is a very different matter from a theoretical physicist justifiably believing this. My familiarity with the issue largely derives from discussions of it in the *New York Times* science section. This kind of information may be enough for me to have responsible beliefs about the matter, because no more can be reasonably expected of me. On the other hand, more can be reasonably expected of the authorities themselves. They are part of a community of inquirers with special knowledge but also special responsibilities, and as a result they should be able to explain away the apparent violations in a detailed way.

These distinctions between epistemically rational belief on the one hand and justified

and non-negligent belief on the other hand flow out of a general theory of rationality. At the heart of the theory is a schema:

A decision, plan, strategy, etc., is rational in sense Y for an individual S if it is epistemically rational for S to believe that the decision, plan, strategy, etc., will do an acceptably good job of contributing to her goals of type Y.

This schema is altogether general; it can be used to distinguish different kinds of rationality and the rationality of different kinds of phenomena.

The result is an epistemology that is both theoretically respectable and that matters. It is theoretically respectable in that it is based on a general schema of rationality that uses epistemically rational belief, which can be explicated without recourse to any further notion of rationality, as a theoretical anchor. It is an epistemology that matters not only because epistemically rational belief becomes the core notion that grounds the notions of rational decisions, rational intentions, rational plans, and so on, but also because it grounds the notions of justified and non-negligent belief, which capture the everyday concerns we have in assessing each other's opinions. These concerns tend not to focus on whether we have met the prerequisites of knowledge but rather on whether we have been reasonably careful, reasonably cautious, and reasonably thorough in forming our opinions, where the standards of reasonability reflect the actual complexities of our lives.

#### BIBLIOGRAPHY

- Foley, Richard: *The Theory of Epistemic Rationality* (Cambridge, MA: Harvard University Press, 1987).  
 Foley, Richard: *Working without a Net* (Oxford: Oxford University Press, 1993).  
 Foley, Richard: *Intellectual Trust in Oneself and Others* (Cambridge: Cambridge University Press, 2001).  
 Foley, Richard: "The Foundational Role of Epistemology in a General Theory of Rationality," in *Virtue Epistemology*, ed. L. Zagzebski and A. Fairweather (Oxford: Oxford University Press, 2001).  
 Plantinga, Alvin: *Warrant: The Current Debate* (New York: Oxford University Press, 1993).

## Richard Fumerton

So, as it turns out, describing my philosophical identity isn't nearly as easy as I thought it would be. And the first conclusion I should probably draw from this is that, even if there are foundations to knowledge and justification, belief states are not the most uncontroversial candidates for the subject matter of foundational truths. I am, nevertheless, convinced that foundationalism is true; and, in the final analysis, it is my commitment to a foundationalism grounded in direct acquaintance with properties and facts that is the cornerstone of not only my epistemology but also my metaphilosophy – my very conception of how to go about arriving at philosophical truth.

#### EPISTEMOLOGICAL FOUNDATIONALISM

The arguments for epistemological foundationalism are too familiar to explore in any detail here. The primary concern is, of course, the threat of regress. I have argued (1996), however, that we should distinguish two importantly different sorts of regress – epistemological and conceptual. The concern over epistemological regress is the more familiar. It is overwhelmingly plausible to suppose that, if our putative evidence for believing that P consists in whole, or in part, in some other belief that E1, that other belief had better be justified if it is to confer justification on P. Much more controversially, I would argue that one had also better have justification for believing that E1 makes plausible P. These two claims constitute the two clauses of what I call the principle of inferential justification (2004, 2001, 1996). If the only way to justify a belief in some proposition is to justifiably infer it from some other proposition, and the principle of inferential justification is true, then to justifiably believe any proposition one would need to complete not one but infinitely many, infinitely long chains of reasoning. This we cannot do. The solution, the foundationalist argues, is to recognize what common sense itself dictates – we have justification for

believing some truths where that justification does not consist in the having of other different justified beliefs.

The conceptual regress argument for foundationalism focuses on the principle of inferential justification and seeks to understand the source of its apparent necessity. Why is it that one cannot justifiably believe P on the basis of E without justifiably believing E? One obvious suggestion is that the claim is analytic. It captures part of our very *understanding* of justification. The concept of an inferentially justified belief just is the concept of a belief that is constituted in part by the having of another different justified belief. But this definition of inferential justification appears to rely on a prior understanding of the concept of epistemic justification. The solution to the threatening *conceptual* regress once again involves appeal to the concept of non-inferential justification. Our understanding of justification is recursive, and the base clause of this recursive analysis requires an understanding of non-inferential justification.

It is one thing to have an abstract argument for the need to introduce non-inferential epistemic justification. It is another to develop its analysis. Like Russell, I'm convinced that the source of non-inferential justification resides in direct acquaintance with both facts and the correspondence between thoughts – the primary bearers of truth value – and the facts that make those thoughts true. One would, of course, like a definition of direct acquaintance; but, as almost all analytic philosophers explicitly recognize, there must be some concepts that are unanalyzable – the fundamental conceptual atoms out of which other concepts are built. Although I can't define direct acquaintance, I can, I believe, "point" to it through thought experiments. We are all familiar with the nagging pain that recedes from consciousness as we become engaged in interesting conversation. I don't believe that the pain ceases – I believe, instead, that it continues even as we no longer notice it. The awareness that was present, ceased, and began again as the conversation ended – *that* is the direct acquaintance of which I speak. To be sure, those suspicious of the very idea of direct acquaintance with properties or facts

might suppose that in the situation I just described the only change involved belief. When we no longer noticed the pain, we no longer believed that we were in pain. But that seems just wrong to me. There is a world of difference between merely believing that one is in pain and being conscious of the pain. An odd sort of hypochondriac might become convinced that he is constantly in pain, albeit a pain that is unconscious. That hypochondriac will have no difficulty distinguishing that belief state from awareness of pain when the latter occurs.

Even if we employ the concept of acquaintance in developing an account of non-inferential justification, an account upon which we build a recursive definition of epistemic justification, we still need to decide *what* we can be non-inferentially justified in believing. Given that direct acquaintance is a real relation (as opposed to an intentional "relation") that requires the existence of its relata, and given that the justification we have for believing commonplace truths about the past, our physical environment, other minds, and the future seems perfectly compatible with those beliefs, being neither true nor even "almost" true, it seems to me that we are left with the undeniably flimsy foundations recognized by the radical empiricist. Truths about the external world, the past, and the future really do lie beyond our "direct" apprehension, and the only epistemic road to those truths lies through legitimate inference (Fumerton, 1985, 1996). Because I think that the second clause of the principle of inferential justification is true, I think that the road to robust knowledge and justified belief will be blocked unless we have non-inferential justification for accepting both deductive and non-deductive principles of reasoning – principles that sanction the move from empirical foundations to the rest of what we believe (Fumerton, 2005).

While the proposition that we can directly apprehend relations of entailment was once the received view, it has never been easy for foundationalists to convince themselves that they are directly acquainted with the truth-makers for principles describing probabilistic connections between propositions. To reach

that conclusion one needs to convince oneself that Keynes was right about the existence of synthetic necessary truths describing *sui generis* “quasi-logical” relations of making probable holding between propositions. And here we reach another crossroad in epistemology. The question is of fundamental importance. To what extent can we model non-deductive reasoning on deductive reasoning? Are there non-deductive arguments that are non-deductively “valid” in virtue of their structure? Does it even make sense to suppose that there are propositions that make probable other propositions and that do so quite independently of contingent features of the world? When the dust settles, it seems to me that most externalists reject classical foundationalism in part because they answer these questions negatively. Committed to the rejection of skepticism, they have no alternative but to replace the inferential internalist’s *a priori* probabilistic connections with such contingent facts as the reliability of belief-forming processes or the effectiveness of beliefs at tracking the truth. In fact the Humean skeptic and the externalist have a great deal in common. Both are convinced that one cannot find a *a priori* reason to accept non-deductive principles of reasoning. Hume infers from the failure to find such principles that skepticism is true. The externalist infers that we must abandon the pre-suppositions of inferential internalism.

The concept of epistemic justification sketched above will strike many as hopelessly idealized. The classical foundationalist’s conditions for justification are simply too demanding. Skepticism inevitably awaits those who seek this sort of justification. This may be right, and I have often referred to the justification I am interested in as *ideal* justification – the sort of justification the philosopher wants to secure when engaged in the distinctive, and perhaps somewhat odd, activity of seeking philosophical assurance. There are, no doubt, derivative concepts of justification. And perhaps the externalists have plausible approaches to understanding the kind of justification in which we are sometimes interested, particularly, from the second-, and third-person perspectives. I have argued

elsewhere (2004), however, that even when developing less demanding standards for justification one will do better beginning with the classical foundationalist’s understanding of ideal justification and stripping away conditions. While reliabilism, for example, may seem initially like a plausible account of less demanding justification, I’m not sure it ever does succeed in accommodating our intuitions about certain hypothetical situations. Malicious demons, “Matrix-world” hallucination, and the like, can obviously deprive us of true beliefs (on a massive scale), and thus can deprive of us reliably produced beliefs, but I’m not sure they can deprive of us of epistemically justified beliefs.

#### PHILOSOPHICAL METHOD

I indicated that my commitment to direct acquaintance is the cornerstone of the foundationalism that defines my epistemology. But I also suggested that reliance on direct acquaintance goes beyond epistemology narrowly conceived. Philosophy without the much maligned given is blind. I simply wouldn’t know how to do philosophy without relying on phenomenological evidence, without relying, for example, on direct apprehension of similarities and differences between properties and thoughts. To be sure, both coherentists about truth and justification, and anti-realists about representation-independent facts often *seem* to embrace implicitly the idea that we should replace the radical empiricist’s veil of perception with an equally opaque veil of belief or representation. On these views, we cannot move beyond the circle of our beliefs. The most radical versions of such views claim that we can make no *sense* of a reality that exists independently of our representations of that reality. But it is hard to believe that proponents of the view really believe it. Both coherence theorists and anti-realists almost always talk as if we do have unproblematic access at least to what we *believe* and how we *represent* the world around us. The *consistent* coherentist/anti-realist is committed to the view that there are no facts which exist independently

of our representations of those facts. But that, of course, implies that our representations of reality are themselves somehow “constructed” out of meta-representations, meta-representations which are constructed out of meta-meta-representations, and so on, *ad infinitum*. To state the view is to see its absurdity. Even from a coherentist’s perspective we have to start somewhere, both ontologically and epistemically. We need to presuppose the existence of beliefs, and if we are to get anywhere we had better have an access to those beliefs, an access that is unmediated by metabeliefs (Fumerton, 2002).

I implied earlier that philosophical arguments for foundationalism simply buttress common sense. One can certainly wonder if the epistemic regress arguments are really that threatening. We are extraordinarily complex beings and can contain an infinite number of beliefs, perhaps enough to give us the *capacity* to come up with infinitely many arguments, one for each of the propositions we believe. But the obvious truth is that we don’t *need* arguments to recognize the severe pain that dominates our consciousness after we are struck with a hard blow. The pain itself, the exemplification of that property, is directly and immediately present to consciousness. It is precisely for that reason that we correctly treat eliminative materialism and logical behaviorism as a kind of joke. The “folk psychology” that recognizes the phenomenally given property of being in pain is more secure than any science of the past, the present, or the future. We can, perhaps, wonder whether that property is identical with the occurrence of neural events or functional states, but we can’t wonder whether or not that property exists.

So-called knowledge arguments have always been at the fore of the debate between dualists and physicalists. Descartes was convinced that he knew that he existed in a way that he couldn’t know of the existence of anything physical and for that reason (though not for that reason alone) he was convinced that he could not be identical with his body. More recently, some philosophers have thought that one could make some headway defending at least property dualism

by reflecting on the obvious change in Mary as she left that black and white room and became dazzled by the experience of colored objects. Knowledge arguments, however, have always been problematic. Indeed, the debate hasn’t progressed much since Smart (1959) developed the most promising physicalist response to such arguments. Cartesian knowledge arguments always implicitly rely on the first part of Leibniz’s Law:

If *x* is identical with *y*, then whatever is true of *x* is true of *y* and vice versa.

If I know that I exist in a way in which I don’t know that my body exists, then I am not identical with my body. If after leaving the room Mary knows that an object is red in a way in which she didn’t know before that the object was red, then the redness she discovers later isn’t identical with any “redness” she knew about before.

In response to such arguments, Smart made the obvious observation that any decent philosophy must come to grips with the fact that there are informative identity claims. One simply cannot infer from the fact that lots of people believe that there is lightning in the sky without believing that there is an electrical discharge in the sky that the former is not identical with the latter. The culprit is *not* Leibniz’s Law. Unlike superficially similar principles concerning the substitutions of co-referential expressions *salva veritate*, Leibniz’s Law simply has no exceptions. The problem is that the objects of knowledge and belief are propositions – *representations* of reality. And from the fact that we have two different representations of reality we simply cannot straightforwardly infer that the reality represented is different. The *proposition* that I exist is (as Descartes proved) different from the *proposition* that my body exists. There is a property of the one *proposition* that is not a property of the other. The *proposition* that Mary believes about the color of objects after leaving the room is a different *proposition* from the proposition about “color” that she believed while in the black and white room (as Jackson (1982), and countless others before, proved). But how do we get



from a distinction among propositions to a distinction in the reality represented by those propositions?

Again, the foundationalist's concept of direct acquaintance is ready to play a pivotal role in epistemological arguments for various forms of dualism. According to the classical foundationalist, acquaintance is a real relation that we bear to certain properties and certain facts (and perhaps certain individuals). We not only believe that we are in pain when we are; we are typically directly aware of the pain itself – the phenomenal property we exemplify. If we are not directly aware of neurons firing, functional states, dispositions to behave, or the other sundry candidates advanced by the physicalist for identification with pain, then we have found a property of pain that is not a property of any physical state and we are back in business employing Leibniz's Law to reach the conclusion that being in pain is not being in any of those physical states. The situation is quite different, according to the classical foundationalist, with the analogies upon which physicalists often rely. To be sure, we may discover through empirical investigation that lightning is an electrical discharge or that water has molecular structure  $H_2O$  but, according to the hardcore foundationalist, we are not directly acquainted with water, molecular structure, lightning or electrical discharges.

As Maxwell (1978), inspired by Russell (1948) claimed, the above argument does not leave the physicalist without resources. When Kripke (1980) used an argument similar to the above to argue for dualism, he failed to consider that the classical foundationalist denies that we have direct access to facts about the physical world – a physical world that includes the brain and its various states. As long as reference to brain states is "indirect" there remains the possibility of an a posteriori discovery that the phenomenal properties with which we are directly acquainted in introspection are in fact constituents of facts the existence of which we must infer. But, even if this is a form of physicalism, it is important to realize that we are not "getting rid" of the "problematic" phenomenally given properties by "reducing" them

to "unproblematic" physical properties. It is much more accurate to suggest that we are theorizing about the unproblematic phenomenally given being a constituent of the epistemically more problematic physical. In the language of reduction, it is much more accurate to describe such a view as reducing certain constituents of brain states to phenomenally given properties.

It is an understatement to suggest that I am unlikely to convince anyone of fundamental positions in the philosophy of mind with these few cryptic comments. But my concern is only to give the reader a feel for the way in which I think foundationalist epistemological commitments should dominate our thinking not just about epistemology but about a host of other fundamental issues in metaphysics. Again, I wouldn't know how to go about thinking about any issues if I really thought there was no access to facts that was independent of our representation of those facts. One cannot do philosophy without data upon which to reflect. And one cannot end a regress in the search for such data without direct awareness of properties and facts.

## BIBLIOGRAPHY

- Fumerton, Richard: *Metaphysical and Epistemological Problems of Perception* (Lincoln, NB: University of Nebraska Press, 1985).
- Fumerton, Richard: *Metaepistemology and Skepticism* (Lanham, MD: Rowman and Littlefield, 1996).
- Fumerton, Richard: "Classical Foundationalism," in *Resurrecting Old-Fashioned Foundationalism*, ed. Michael De Paul (Lanham, MD: Rowman and Littlefield, 2001).
- Fumerton, Richard: *Realism and the Correspondence Theory of Truth* (Boston, MA: Rowman and Littlefield, 2002).
- Fumerton, Richard: "Epistemic Probability," *Philosophical Issues* 14 (2004).
- Fumerton, Richard: "Achieving Epistemic Ascent," in *Sosa and His Critics*, ed. John Greco (Oxford: Blackwell, 2004).
- Fumerton, Richard: "The Challenge of Refuting Skepticism," in *Contemporary Debates in Epistemology*, ed. Mathias Steup and Ernest Sosa (Oxford: Blackwell, 2005).
- Jackson, Frank: "Epiphenomenal Qualia," *Philosophical Quarterly* (1982).

- Kripke, Saul: *Naming and Necessity* (Cambridge, MA: Harvard University Press, 1980).
- Maxwell, Grover: "Rigid Designators and Mind-brain Identity," in *Perception and Cognition*, ed. Wade Savage (Minneapolis, MN: University of Minnesota Press, 1978).
- Russell, Bertrand: *Human Knowledge: Its Scope and Limits* (New York: Simon and Schuster, 1948).
- Smart, J. J. C.: "Sensations and Brain Processes," *Philosophical Review* (1959).

## Alvin I. Goldman

### THE CAUSAL THEORY OF KNOWING

My first paper in epistemology was "A Causal Theory of Knowing" (Goldman, 1967). It aimed to counteract the then-prevailing dogma that the knowledge status of a belief state has nothing to do with its genesis (Reichenbach, 1938), while also accommodating Gettier's (1963) counter-examples to the justified true belief analysis. The causal theory was a version of what later came to be called an "anti-luck" theory. It was written against the background of Grice's (1961) causal theory of perception and Martin and Deutscher's (1966) causal theory of remembering. "A Causal Theory of Knowing" defended the strong thesis that (empirical) knowledge requires an appropriate causal connection between a true belief and its truth-maker. This strong thesis was later abandoned, but I continue to embrace the weaker thesis that a belief's knowledge status depends on how it is formed (or causally sustained).

### NO RELEVANT ALTERNATIVES THEORY

The causal theory ran into problems with the fake barns example (due to Carl Ginet), as pointed out in "Discrimination and Perceptual Knowledge" (Goldman, 1976). Suppose that Henry's belief that there's a barn before him is perceptually caused by a genuine barn. The causal theory considered perceptual causation an appropriate causal process, so Henry's belief ought to qualify as knowledge. But suppose the district is full of fake barns that

would be indistinguishable from the real one; the intuitive judgment is that Henry doesn't know. To address this problem, an etymological observation was made. One sense of "know" is "to distinguish one thing from another", and the propositional use of "know" seems to be related. To know is to exercise a discriminative ability with respect to the fact F in question, where a discriminative ability with respect to F was explicated in counterfactual terms: one wouldn't mistake a relevant non-F situation (e.g., the presence of a mere barn façade) for an F-situation. This is a version of the idea that knowledge is reliably, or non-accidentally, acquired true belief. Applied to perceptual knowledge, the subject must not only have a true perception-based F-belief, but also there must be no relevant non-F situation in which an equivalent perceptual experience would occur that would lead the subject to a (false) F-belief.

### PROCESS RELIABILISM

"What Is Justified Belief?" (Goldman, 1979) presented a reliabilist approach to justifiedness rather than knowledge. The theory focused on the truth-ratios of the process types used in belief formation, and generally goes by the name "process reliabilism". In its simplest form, it says that a belief's justificational status hinges on the psychological processes that produce it, e.g., perception, memory, introspection, or various inference patterns. Beliefs formed by highly reliable processes are justified; beliefs formed by insufficiently reliable processes are unjustified. The approach is motivated by examples. Intuitively, beliefs formed by unreliable processes like sheer guesswork or wishful thinking are unjustified even if their propositional contents stand in appropriate relations to evidence beliefs. Process reliabilism contrasts with traditional theories like foundationalism and coherentism in being a "historical" theory rather than a "current time-slice" theory. Foundationalism and coherentism imply that justificational status is fixed by mental states held at the time of believing. According to reliabilism, it's the mental history of a belief

that fixes its justificational status. The subject may not recall the processes responsible for the belief and may not know their reliabilities; but these are the factors that matter.

Process reliabilism is a species of externalism, the thesis that some of the facts relevant to justificational status (so-called “justifiers”) are external to the agent. Facts about the truth-ratios of belief-forming processes are external because truth-values in general are external to an agent. Reliabilism acknowledges that other justifiers of a belief are internal. For example, which mental processes generate a belief is relevant to its justifiedness, where mental processes are internal in the sense of being “under the skin”.

Internalists pose various objections to process reliabilism including the new evil demon problem and the reliable clairvoyance problem (BonJour, 1985; see also Lehrer’s 1990 Truetemp example). The new evil demon problem challenges the necessity of reliability for justifiedness, whereas the clairvoyance case challenges its sufficiency. Internalists like BonJour contend that the clairvoyance case shows that it isn’t sufficient for justifiedness to form a belief by a reliable process; one must also have, it appears, a justified belief in the process’s reliability. This higher-level condition, however, would violate a constraint proposed in “What Is Justified Belief?”, namely that an explanation of justifiedness should be non-circular, that is, should eventually bottom out in states of affairs that aren’t themselves epistemic. I pointed this out in a critique of BonJour (Goldman, 1989). A third main problem for process reliabilism is the generality problem, which is discussed in a separate entry (see RELIABILISM in Part III).

Although I emphasize the need to explain justifiedness in a non-epistemic fashion, I don’t dispute the intuitive force of the new evil demon or clairvoyance examples. I have therefore explored a variety of ways to modify process reliabilism to accommodate these and similar examples. In *Epistemology and Cognition* (Goldman, 1986) I floated a “normal worlds” version of reliabilism. This theory held that a belief is justified just in case its generating process is reliable in normal worlds, whether or not it’s reliable in the subject’s

own world. Perception-based belief-forming processes would not be reliable in a demon world but they are reliable in normal worlds, so someone situated in a demon world would have justified beliefs when using such processes. “Strong and Weak Justification” (Goldman, 1988) advanced a two-senses-of-justification approach, and “Epistemic Folkways and Scientific Epistemology” (Goldman, 1992a) argued for a “two-stage” form of reliabilism. The last approach is what I currently regard as the most promising version of reliabilism, and it remains under development (see RELIABILISM in Part III).

#### NATURALISTIC EPISTEMOLOGY

*Epistemology and Cognition* featured a more extended and refined treatment of reliabilism, dealing with both knowledge and justification. In analyzing knowledge it used the resources of both type reliabilism (“global” reliability) and token-directed modal reliabilism (“local” reliability, in this case the no relevant alternatives theory). Also, in addressing the epistemic properties of psychological processes, it added to reliability the properties of power and speed. A reliable belief-forming process is one that rarely breeds error, but it might be very cautious and sparse in producing true beliefs. Powerful belief-forming processes, by contrast, are ones that often produce true beliefs.

*Epistemology and Cognition* was anchored in a reliabilist account of justification and of knowledge, but its main agenda lay elsewhere. It argued that epistemology’s job should not be exhausted by analysis of epistemic concepts. Epistemologists should also consider how human agents can go about obtaining knowledge, justified belief, or other epistemological desiderata. Regulative epistemology should identify specific reliable processes and methods available to human cognizers. This implies that epistemology might need help from cognitive science, because an inventory of available cognitive processes and an assessment of their reliability, power, and speed are tasks belonging to cognitive science, an empirical discipline. Would

this be a break with the field's tradition? I answer in the negative. Epistemology has long been interested in the merits or demerits of putative sources of knowledge. Historical philosophers never drew a sharp distinction between epistemology and (what is today called) psychology; they did psychology on their own. In modern times, epistemology should take notice of what cognitive science says about the mechanisms or heuristics of belief formation, both their strengths and their infirmities (Goldman, 2002a). This way of naturalizing epistemology contrasts with Quine's (1969) way by preserving both the analytic and the normative aspects of epistemology.

#### RECENT WORK ON RELIABILISM AND INTERNALISM

In recent years I have made fresh attempts to display the strength of reliabilism by showing how it solves problems its competitors cannot solve, and by pinpointing weaknesses in opposing positions, e.g. internalism.

Many internalists are foundationalists. Since a core thesis of foundationalism is that some beliefs (so-called basic beliefs) are directly or immediately justified, it is incumbent on internalist foundationalists to explain immediate justifiedness. How can a belief be justified other than by standing in a suitable relation to other justified beliefs? In "Immediate Justification and Process Reliabilism" (Goldman, 2008a) I argue that internalism lacks the resources to solve this problem but process reliabilism has the requisite resources.

The article focuses on an example introduced by Feldman (2003). A novice birdwatcher and an expert both see a rare pink-spotted flycatcher in the woods and each spontaneously identifies it as such. The expert has a well-founded (justified) belief while the novice merely jumps to a lucky conclusion, and lacks a justified belief. Can internalists explain this? Feldman says that one but not the other birdwatcher makes a "proper response" to his experience.

But what makes a response "proper" must be spelled out. Feldman first proposes that response propriety depends on closeness of a belief's content to the content of the experience. But some beliefs about experience are immediately justified, although their content concerns the experience *type* rather than its content. So the contents of the belief and of the experience may not be close at all. Second, Feldman tries to explain the propriety of the expert's response in terms of his suitable training. How does training matter to justifiedness? I answer: training makes beliefs of the appropriate category come out true. But then the explanation is a thinly disguised way of saying that suitable training breeds *reliability*. Thus, process reliabilism is right on track to explain immediate justifiedness, whereas internalism lacks the necessary explanatory resources.

A more thoroughgoing critique of internalism, "Internalism Exposed" (Goldman, 1999a), appeared a decade earlier. This paper explained how a "guidance" conception of justifiedness might rationalize internalism. To guide doxastic decision-making so that one believes only when belief is justified, an agent must be able to tell (correctly) when belief is justified. If this is right, the "justifiers" of a prospective belief (i.e. the determiners of justification) must be accessible to the agent, which is what access internalism demands. But what exactly is accessibility, and which facts are accessible in the relevant sense?

Some epistemologists, including Chisholm (1989), restrict the justifiers to states of consciousness. This seems to imply that access consists in introspection. An introspectionist interpretation of internalism, however, breeds a worrisome brand of skepticism. The vast majority of one's beliefs are stored in memory. Because these beliefs aren't conscious or occurrent at most moments at which they are held, it's unlikely that they have adequate justifiers residing in consciousness at those times. According to this form of internalism, then, these beliefs aren't justified. Introspectionist internalism threatens a drastic diminution in the stock of beliefs ordinarily deemed justified. Even if internalism

expands the conception of “access” to include retrieval from memory, permitting stored beliefs to qualify as justifiers, there is still the problem of forgotten evidence, which also threatens skepticism.

Suppose “internal” is understood in the sense of “mental” rather than “accessible” (Conee and Feldman, 2001). This faces the problem of accommodating logical and probabilistic relations. Any variety of internalism must hold that many justified beliefs achieve justificational status in virtue of their (objective) logical or probabilistic relations to the contents of evidential states. But logical and probabilistic relations aren’t mental states, so the indicated definition of “internal” doesn’t permit these relations to qualify as justifiers, a bad result for internalism (cf. Comesaña, 2005).

#### EPISTEMIC VALUE

Epistemology is immersed in evaluative discourse, ostensibly invoking a distinctive set of “epistemic” values. What are these values? True belief, justifiedness, and rationality are commonly mentioned as among the intellectual values. Are they entirely disparate values, or is there is a unifying theme?

“The Unity of the Epistemic Virtues” (Goldman, 2001a) made an exploratory effort to substantiate value monism, where the fundamentally valuable state is something like true belief. This fits naturally with reliabilism inasmuch as reliability produces a high ratio of true belief, so justification construed reliabilistically rests on the value of true belief. One type of epistemic good, however, seems to resist this treatment. Proportioning one’s belief to the evidence, a desideratum stressed by evidentialists and probabilistic epistemologists of various stripes, is not straightforwardly explainable in terms of the pursuit of true belief. It seems to require a value of “fitting the evidence” that is independent of true belief. Do evidentialists therefore advocate a species of value monism in which the *only* epistemic value is proportionment? That would be inconsistent with another epistemic

value, that of gathering evidence well, a value clearly prized in experimental science. Perhaps evidentialists only mean to advocate a pluralism of values that include evidence proportionment. The challenge then facing true-belief monism is to show how evidence proportionment can be rationalized.

Isn’t this type of monism already undercut by the existence of error avoidance as a separate value? This worry is handled by the value framework of veritism in *Knowledge in a Social World* (1999b). According to veritism, higher degrees of credence (or subjective probability) in a true proposition have more veritistic value than lower degrees of credence in the same proposition. Veritism thereby provides a way to blend the two traditional epistemic values, true belief and error avoidance, into a single magnitude or quantity of veritistic value, yielding a single type of value (realizable in different degrees) in place of two types.

Veritism might be able to explain the value of evidence proportionment in terms of, say, maximizing expected veritistic value. Perhaps expected veritistic value is maximized if one assigns a degree of belief in proportion to the strength of the evidence. Unfortunately, this idea doesn’t seem to work, because expected veritistic value is maximized by extreme degrees of belief, not by intermediate ones (DePaul, 2004). So it remains to be seen whether the value of evidence proportionment can be grounded, ultimately, in something like veritistic value.

A recent challenge to reliabilism in the value domain is the “value of knowledge” (or *Meno*) problem (Zagzebski, 2003; Kvanvig, 2003). How does process reliabilism about knowledge explain the extra value of knowledge over mere true belief? Why would the property of being caused by a reliable process make a true belief more valuable than it is intrinsically, any more than having been brewed by a reliable espresso machine makes a good cup of espresso any better than it is in itself? A response to this problem is presented in Goldman and Olsson (forthcoming). We offer several ways for reliabilism to meet the challenge.



Reliabilism is often associated with epistemological naturalism, which in turn is associated with resistance to the a priori (Quine, 1969; Kitcher, 1981). What is my stance toward the a priori? In “A Priori Warrant and Naturalistic Epistemology” (Goldman, 1999c) I part company with most naturalists in finding a priori warrant perfectly acceptable. Much depends, of course, on how one construes this notion. It is clearly essential that an a priori source of belief be non-perceptual or non-experiential. This poses no conflict with reliabilism because there may well be reliable non-perceptual processes. But not all non-perceptual processes are clear cases of a priori processes. Memory encoding, memory retrieval, and introspection may all have instances that are neither sensory nor sense-derived, but that shouldn’t suffice to qualify them as sources of the a priori. How can a priori belief-forming processes be demarcated? Is there a distinctive phenomenology associated with the a priori? I doubt it. We should not expect important distinctions among cognitive processes to reside in phenomenologically detectable properties. Rather than rely on introspection, we should turn to cognitive science to identify the elementary processes involved in mathematical thought and logical reasoning, and cognitive science is indeed shedding light on these subjects. What we shouldn’t demand is that these processes have some tight resemblance to one another. If this is right, the a priori may not be a very “natural” epistemic category. For this reason, I question the importance of the a priori to epistemology, but I don’t disavow it altogether.

#### PHILOSOPHICAL ANALYSIS, INTUITIONS, AND NATURALISM

In describing the project of *Epistemology and Cognition*, I noted that part of the project involved fairly standard philosophical analysis that appeals to intuitions to support or refute proposed analyses. How “naturalistic” can that project be? Many naturalizing

epistemologists are highly skeptical, even dismissive, of reliance on intuition. Experimental philosophers have tried to demonstrate the unreliability of intuition by showing the absence of intersubjective agreement, instability, and the impact of irrelevant factors on intuitions (Weinberg et al., 2001; Nichols et al., 2003; Swain et al., forthcoming). In light of this, can a naturalistic philosophy have any truck with intuitions? A consistent naturalistic epistemologist should pay close attention to this kind of evidence. It must not be assumed uncritically that intuition in its philosophical uses is a reliable method for forming beliefs about the target questions, or that philosophical analysts can make reliable use of the intuitions of others.

As argued in “Philosophical Intuitions: Their Target, Their Source, and Their Epistemic Status” (Goldman, 2007), classification intuitions are best-interpreted as providing evidence about the content of a subject’s concept, in a psychological sense of “concept”. A concept in this sense is best-understood as the mental representation of a kind or a property. Examples would be a subject’s mental representation of knowledge, of justifiedness, of causation, or what have you (NB, not knowledge, justifiedness, or causation “themselves”). Under this approach it is hard to imagine how an intuition could fail to be evidence of some sort about this target, because an intuition is a fairly direct output of (1) the kind representation and (2) the representation of the instance, or example, that the subject has been asked about. There must be some systematic connection between the intuition and these other two representations. To be sure, representations of many other things – including previous examples the subject has heard, and so forth – can influence the resulting intuition. All this needs to be taken into account. But it is hard to think that intuitions couldn’t be helpful evidence at all concerning the contents of kind representations. Many philosophers, of course, hold that philosophical analysis should inquire into kinds and properties themselves, not into people’s mental representations of them (e.g. Kornblith, 2002). Ultimately we may indeed be interested in

kinds and properties, but raw intuitions do not offer the best insight into those things. Intuitions are defensible, however, as potentially helpful evidence – cautiously and critically handled – about individuals' concepts. The proper use of intuitions in philosophy would not make it the signature of armchair philosophizing, but a properly naturalistic – that is, scientific and psychologistic – aid to philosophizing. Intuitions can be collected both inside and outside the “laboratory”, but their interpretation should be assessed against the backdrop of a scientifically informed theory of concepts and their deployment.

#### SOCIAL EPISTEMOLOGY

In the last two decades, beginning in 1987, a large chunk of my epistemological energy has been devoted to social epistemology. I have sought to develop a systematic and comprehensive approach to social epistemology predicated on the idea that epistemology is an evaluative or normative subject. Traditional epistemology focused on how individuals can do well intellectually, whereas social epistemology should focus on how communities or groups can do well (or how individuals can do well via social means). “Do well” in what way? What criterion or set of criteria should be used to judge the quality of social-intellectual activity? In *Knowledge in a Social World* (1999b), four types of criteria were initially considered.

*Consensus consequentialism* would judge social practices by their tendencies to promote agreement. *Pragmatism*, or *utility consequentialism*, would evaluate social practices by the utility that would accrue. *Proceduralism* would fasten on intrinsic merits of social-intellectual practices. Finally, *truth consequentialism*, or *veritism*, would assess social practices by their contribution to the veritistic attainments of groups of individuals. I chose to focus on the fourth conception, because of the centrality of truth possession as an epistemic, or intellectual, value.

The kind of knowledge addressed in *Knowledge in a Social World* (henceforth KSW) was called “weak” knowledge, i.e. true belief,

without the usual addenda such as justification, anti-Gettierization, etc. Additional evidence for the existence of this weak sense of “knowledge” is now presented in Goldman and Olsson (forthcoming). The decision to frame KSW in terms of weak knowledge facilitated various technical developments not reviewed here. It also allowed me to bypass thorny issues about justification and the analysis of “strong” knowledge that fill other epistemological works. Because KSW’s agenda was already so large, this was convenient, but it wasn’t intended to suggest that justification isn’t of interest in social epistemology (see below).

Can any definite, substantive results be established in veritistic social epistemology? Here is an example of a substantive result involving the optimal deployment of expertise in forming individual judgments. Suppose you want to rely on the judgments or opinions of several people, e.g. weather forecasters, in forming a belief about tomorrow’s weather. How should you use their judgments? One method is to weight their opinions equally and go with the majority judgment. Another is to go with the opinion of the single most expert individual. A third is to use some sort of weighted voting scheme. Shapley and Grofman (1984) show that if the experts are mutually independent, and if the a priori likelihood of either outcome is the same (one half), then a weighted majority rule assigning the following weights maximizes your chance of being correct:

$$w_i \propto \log \left( \frac{p_i}{1 - p_i} \right)$$

For example, if there are five weather forecasters whose respective probabilities of being correct are 0.9, 0.9, 0.6, 0.6, 0.6, then an optimum weighting scheme assigns them, respectively, the weights 0.392, 0.392, 0.072, 0.072, 0.072. This method is the veritistically best way to amalgamate the expertise of the five forecasters.

*Knowledge in a Social World* examines a wide swath of social epistemological territory from a veritistic point of view. Rules of good argumentation are examined in terms

of their tendency to increase hearers' knowledge (ch. 5; also see Goldman, 1994). The technology and economics of communication are examined in terms of their impact on knowledge (ch. 6). Communication technology can expand the size of the "message infosphere" in which messages are available for receipt or retrieval by interested parties. The traditional mass media are subject to owner and sponsor influence, but digital media have their own liabilities. Today one would concentrate on the risks to public information from media concentration and the gradual replacement of newspapers by the blogosphere (Goldman, 2008b). Another chapter of KSW deals with the bearing of governmental speech policies on public information (ch. 7). Institutions of legal adjudication are assessed in terms of their effect on the truth-getting propensities of different trial and evidence systems (ch. 9; also see Goldman, 2005). The importance of information to a democracy is also assessed (ch. 10). When casting a vote in an election, a voter confronts what I call a "core voter question". It is shown that greater core voter knowledge among the electorate improves the chances that democracy's aims will be served.

In articles postdating KSW I tackle other questions in social epistemology, often from a non-veritistic standpoint. Questions of justification or reasonableness are prominent. A fundamental question in social epistemology is when a hearer is justified in accepting the testimony of a speaker. A more complex question is which speaker to believe when two or more deliver conflicting testimony. This is especially problematic when a novice tries to choose among putative experts who disagree with one another. Can a novice make a reasonable, justified choice among experts and, if so, on what basis? These issues are addressed in "Experts: Which Ones Should You Trust?" (Goldman, 2001b).

Another salient topic in social epistemology is epistemic relativism. Is there a unique system of epistemic norms or principles that is objectively correct, independent of culture or community? An affirmative answer would constitute strong epistemic objectivism. A

negative answer would constitute epistemic nihilism (at least, if it holds that there is *no* objectively correct system). In "Epistemic Relativism and Reasonable Disagreement" (Goldman, 2008c), a novel form of relativism is proposed that occupies a middle position between strong objectivism and nihilism. "Objectivity-based relativism" assumes the truth of strong objectivism but endorses another thesis friendly to relativism. People occupy different evidential positions vis-à-vis the correct system of norms. Hence, the objective justificational status of different people vis-à-vis different epistemic systems is varied rather than uniform. Some people are objectively justified in believing certain epistemic norms to be correct and others are objectively justified in believing rival epistemic norms to be correct. If this kind of relativism is right, there are interesting consequences for the possibility of reasonable disagreement among individuals with the same ("material") evidence on a given factual question.

In light of the fascinating theoretical issues in social epistemology as well as their potential applications to society, social epistemology deserves intensive development. I therefore devote considerable energy to this mission by editing the journal *Episteme: A Journal of Social Epistemology* ([www.episteme.us.com](http://www.episteme.us.com)).

#### BIBLIOGRAPHY

- BonJour, Laurence: *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1985).
- Chisholm, Roderick: *Theory of Knowledge*, 3rd edn (Englewood-Cliffs, NJ: Prentice Hall, 1989).
- Comesana, Juan: "We Are (Almost) All Externalists Now," *Philosophical Perspectives* 19 (2005), 59–76.
- Conee, Earl and Feldman, Richard: "Internalism Defended," in *Epistemology: Internalism and Externalism*, ed. H. Kornblith (Oxford: Blackwell, 2001).
- DePaul, Michael: "Truth Consequentialism, Withholding and Proportioning Belief to the Evidence," in *Epistemology: Philosophical Issues* 14, ed. E. Sosa and E. Villanueva (Boston, MA: Blackwell, 2004), 91–112.

- Feldman, Richard: *Epistemology* (Upper Saddle River, NJ: Prentice-Hall, 2003).
- Gettier, Edmund: "Is Justified True Belief Knowledge?," *Analysis* 23 (1963), 121–3.
- Goldman, Alvin: "A Causal Theory of Knowing," *Journal of Philosophy* 64 (1967), 357–72. Reprinted in Goldman (1992b).
- Goldman, Alvin: "Discrimination and Perceptual Knowledge," *Journal of Philosophy* 73 (1976), 771–91. Reprinted in Goldman (1992b).
- Goldman, Alvin: "What Is Justified Belief?," in *Justification and Knowledge*, ed. G. Pappas (Dordrecht: Reidel, 1979). Reprinted in Goldman (1992b).
- Goldman, Alvin: *Epistemology and Cognition* (Cambridge, MA: Harvard University Press, 1986).
- Goldman, Alvin: "Strong and Weak Justification," in *Philosophical Perspectives*, 2, *Epistemology*, ed. J. Tomberlin (Atascadero, CA: Ridgeview, 1988). Reprinted in Goldman (1992b).
- Goldman, Alvin: "BonJour's *The Structure of Empirical Knowledge*," in *The Current State of the Coherence Theory*, ed. J. Bender (Dordrecht: Kluwer, 1989).
- Goldman, Alvin: "Epistemic Folkways and Scientific Epistemology" (1992a), in Goldman (1992b).
- Goldman, Alvin: *Liaisons: Philosophy Meets the Cognitive and Social Sciences* (Cambridge, MA: MIT Press, 1992b).
- Goldman, Alvin: "Argumentation and Social Epistemology," *Journal of Philosophy* 91 (1994), 27–49.
- Goldman, Alvin: "Internalism Exposed," *Journal of Philosophy* 96 (1999a), 271–93. Reprinted in Goldman (2002b).
- Goldman, Alvin: *Knowledge in a Social World* (Oxford: Oxford University Press, 1999b).
- Goldman, Alvin: "A Priori Warrant and Naturalistic Epistemology" (1999c), in *Philosophical Perspectives*, 13, *Epistemology*, ed. J. Tomberlin (Malden, MA: Blackwell, 1999). Reprinted in Goldman (2002b).
- Goldman, Alvin: "The Unity of the Epistemic Virtues" (2001a), in *Virtue Epistemology*, ed. L. Zagzebski and A. Fairweather (New York: Oxford University Press, 2001). Reprinted in Goldman (2002b).
- Goldman, Alvin: "Experts: Which Ones Should You Trust?" (2001b), *Philosophy and Phenomenological Research* 63 (2001), 85–109. Reprinted in Goldman (2002b).
- Goldman, Alvin: "The Sciences and Epistemology" (2002a), in *The Oxford Handbook of Epistemology*, ed. P. Moser (New York: Oxford University Press, 2002), 144–76.
- Goldman, Alvin: *Pathways to Knowledge, Private and Public* (New York: Oxford University Press, 2002b).
- Goldman, Alvin: "Legal Evidence," in *The Blackwell Guide to the Philosophy of Law and Legal Theory*, ed. M. Golding and W. Amundson (Boston, MA: Blackwell, 2005), 163–75.
- Goldman, Alvin: "Philosophical Intuitions: Their Target, Their Source, and Their Epistemic Status," *Grazer Philosophische Studien* 74 (2007), 1–26.
- Goldman, Alvin: "Immediate Justification and Process Reliabilism" (2008a), in *Epistemology: New Essays*, ed. Q. Smith (New York: Oxford University Press, 2008).
- Goldman, Alvin: "The Social Epistemology of Blogging" (2008b), in *Information Technology and Moral Philosophy*, ed. J. van den Hoven and J. Weckert (Cambridge: Cambridge University Press, 2008).
- Goldman, Alvin: "Epistemic Relativism and Reasonable Disagreement" (2008c), in *Disagreement*, ed. R. Feldman and T. Warfield (New York: Oxford University Press, 2008).
- Goldman, Alvin and Olsson, Erik: "Reliabilism and the Value of Knowledge," in *Epistemic Value*, ed. A. Haddock, A. Millar and D. Pritchard (Oxford: Oxford University Press, forthcoming).
- Grice, H. P.: "The Causal Theory of Perception," *Proceedings of the Aristotelian Society*, supplementary vol. 35 (1961).
- Kitcher, Philip: "Why Kant Almost Wrote 'Two Dogmas of Empiricism' and Why He Didn't," *Philosophical Topics* 12 (1981), 217–50.
- Kornblith, Hilary: *Knowledge and Its Place in Nature* (Oxford: Oxford University Press, 2002).
- Kvanvig, Jonathan: *The Value of Knowledge and the Pursuit of Understanding* (Cambridge: Cambridge University Press, 2003).
- Lehrer, Keith: *Theory of Knowledge* (Boulder, Colo.: Westview Press, 1990).
- Martin, C. B. and Deutscher, Max: "Remembering," *Philosophical Review* 75 (1966): 161–96.
- Nichols, Shaun, Stich, Stephen and Weinberg, Jonathan: "Metaskepticism: Meditations in Ethno-epistemology," in *The Sceptics*, ed. S. Luper (Aldershot: Ashgate, 2003), 227–47.
- Quine, W. V.: "Epistemology Naturalized," in *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969).
- Reichenbach, Hans: *Experience and Prediction* (Chicago, IL: University of Chicago Press, 1938).
- Shapley, Lloyd and Grofman, Bernard: "Optimizing Group Judgmental Accuracy in the Presence of Interdependence," *Public Choice* 43 (1984), 329–43.



Swain, Stacey, Alexander, Joshua and Weinberg, Jonathan: "The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp," *Philosophy and Phenomenological Research* (forthcoming).

Weinberg, Jonathan, Nichols, Shaun and Stich, Stephen: "Normativity and Epistemic Intuitions," *Philosophical Topics* 19 (2001), 429–60.

Zagzebski, Linda: "The Search for the Source of Epistemic Good," *Metaphilosophy* 34 (2003), 12–28.

## Gilbert Harman

What is distinctive about my views in epistemology? One thing is that my concern with epistemology is a concern with methodology. Furthermore, I reject psychologism about logic and reject the idea that deductive rules like *modus ponens* are in any way rules of inference. I accept a kind of methodological conservatism and reject methodological theories that appeal to special foundations, analytic truth, or a priori justification. Although I believe that there are significant practical aspects of theoretical reasoning, I reject the suggestion that theoretical reasoning is a special case of practical reasoning applied to a special epistemic goal. I also believe that a methodological epistemology that is concerned with the reliability of inferential methods can benefit from an appreciation of important relevant concepts and results about reliability in statistical learning theory.

### 1. LOGIC AND EPISTEMOLOGY

Logic is a theory of implication and inconsistency. It has an abstract subject matter, being concerned with relations among propositions. It does not have a psychological subject matter, nor a normative subject matter. Logical principles are exceptionless necessary principles about what follows from what, not default principles about what can be inferred from what. From contradictory beliefs, anything and everything follows, but it is not the case that from contradictory beliefs it is rational or reasonable to infer anything and everything.

Sometimes, the fact that S follows from things you believe is part of what makes it reasonable for you to believe that S. But not always. It may be more reasonable to abandon one of the beliefs from which it follows that S. Furthermore, infinitely many things follow from your beliefs hardly any of which are worth adding to your beliefs.

Reasoning normally involves change in view, often giving up things previously accepted as well as coming to accept new things, so reasoning is not in general thinking through a proof or argument in logical order, with premises, intermediate steps, and final conclusion. Although in reasoning you may construct an argument, your construction of the argument may start with the conclusion and work backward to the premises, or you may start in the middle and work in both directions. If the argument is part of an explanation that you infer as the best explanation of some data, the conclusion of the argument may be the beginning of your reasoning and the premise of the argument may be the conclusion of your reasoning.

Deductive rules are rules of implication, not rules of inference or reasoning. They are perhaps rules about what makes something a proof or argument, but they are not rules that you could follow in constructing a proof or argument, except in the sense that they characterize what you are constructing.

Some theorists respond to these obvious points by saying that they show only that there is a harmless ambiguity in the phrase "rule of inference", one meaning of which is "rule of implication". But the "ambiguity" is not harmless. It has served to hide the difficult issue of saying how recognition of implications does or ought to figure in inference, it makes it much harder for good students to learn logic than it should, and it has been responsible for many confused attempts to develop various logics corresponding to aspects of the way people reason: inductive logic, because there is inductive reasoning; practical logic, because there is practical reasoning; relevance logic, because reasoning is sensitive to relevance; a logic of belief revision because reasoning is change in view; and so



on. (Similarly for attempts to state “closure principles” for knowledge.) These failed projects have been motivated entirely by the thought that rules of logic are “rules of inference”. This “terminology” has not been “harmless” and should be abandoned.

## 2. METHODOLOGICAL CONSERVATISM

In reasoning, you start where you are with your current beliefs, plans, and goals, and your current methods or procedures for modifying these plans and methods. It is not reasonable for you to make any changes in your starting points except to resolve tensions within them and to answer questions which you have reasons to answer. Any reasonable methodology has to be conservative in this sense.

In a certain (methodological) sense, your initial beliefs, plans, goals, and methods have an immediate default or *prima facie* “justification”. They are the “foundations” for your reasoning – foundations in the sense of starting points. Philosophers sometimes advocate more limited foundations, where the strictest theories allow as foundations only things you cannot doubt and where your non-foundational beliefs, goals, plans and methods are to be retained only to the extent that they have argument like justifications ultimately based on foundational beliefs, goals, plans, and methods. Such strict foundations theories do not seem plausible as methodological theories, since you do not normally associate argument like justifications with the beliefs, goals, plans, and methods you currently accept, which means these views imply you should abandon almost everything you believe.

Even apart from that consideration, it is obvious that the strictest methodological foundations theories imply that you ought to give up almost all your beliefs, goals, plans, and methods, as lacking sufficient justifications given the strictest standards. This result can be avoided only by adding to the foundations, making them less strict – for example, allowing into the foundations certain perceptual beliefs about the environment, and certain

memory beliefs about the past, as well as inferential methods like enumerative induction, inference to the best explanation, and reliance on testimony. These less strict foundations theories are closer to general conservatism. Inasmuch as the move to less strict theories is motivated by the initial implausibility of the skeptical implications of the stricter theories, the move appears to be based on a background acceptance of something like general conservatism.

General conservatism is a widely accepted methodology in philosophy outside of epistemology, for example in Rawls’s discussion of trying to put the principles of justice you accept into “reflective equilibrium” with your judgments about particular cases.

Rawls and others emphasize that the relevant reflective equilibrium should be “wide” rather than “narrow”, also taking into account what other views there are and what can be said to proponents of these other views. More generally, as Thagard has emphasized, you should consider how your current methods look in relation to the best scientific reasoning, what your goals in reasoning are, and how reliable various methods are. Furthermore, it is methodologically useful to maintain a Popperian questioning attitude towards the views you accept. These further methodological principles do not conflict with general methodological conservatism to the extent that they are themselves currently accepted or are principles there are reasons to accept given current beliefs, goals, plans, and methods.

People sometimes object to methodological conservatism by considering a case in which you irrationally come to accept a new belief, perhaps randomly. The objection is that it seems wrong to suppose your belief is justified as soon as it is accepted.

But methodological conservatism does not imply that your belief is justified as soon as it is accepted. It implies that the belief is *prima facie* of default justified. Whether it is fully justified (in the relevant methodological sense) depends on whether you are aware of how you came to accept that belief. If you are aware that you did not have good reason to accept it, then that awareness is in tension

with your continuing to accept the belief. Similarly, if your other beliefs and methods give you now reason to conclude that you did not have good reason to accept the belief in question, then there is a tension in your beliefs that might point to getting rid of that belief. On the other hand, if you are unaware of how you came to have the belief, that by itself is not enough to give you reason to abandon the belief, because almost all your beliefs have the property that you are unaware of how you came to have them.

Remember that general conservatism is a methodological principle, offering methodological advice of a sort a person can take. It is not a principle of justification in some other sense of "justification".

### 3. IRRELEVANCE OF APRIORITY AND ANALYTICITY

Philosophers sometimes suppose that a limited number of beliefs and methods (and maybe some goals and plans) are *a priori* justified, but I am unaware of any satisfactory accounts of what these suppositions might mean.

There are of course those beliefs and methods (and goals and plans) that are prior to a given inquiry, meaning those beliefs and methods (and goals and plans) that are taken for granted at the beginning of the inquiry. While I agree that all such beliefs and methods (and goals and plans) are (default) justified *a priori*, in the sense of prior to a given inquiry, this is not what is meant by the philosophers I have in mind, because they suppose that only a limited number of beliefs, etc., have the relevant sort of *a priori* justification.

There may also be beliefs and methods (and goals and plans) that are prior in the sense of prior to birth – innate or innately programmed to arise. But the fact that a given belief or method is innate does not make it *a priori* justified according to the philosophers I have in mind, because it is for them an "empirical question" whether a belief or method is innate and a further empirical question whether it is innate in a way that makes it likely to be true, where the fact that

these are empirical questions means that the beliefs and methods in question are not *a priori* justified in the relevant sense.

I know only two ways to try to explain the relevant sense of *a priori*. One presupposes a strict foundations theory of justification. The idea is that everything you are justified in believing is associated with your justification for believing it, and similarly for methods, goals, and plans, where there are strict standards for foundational beliefs, methods, goals, and plans, which must be directly justified either by your present sensory or perceptual experiences or by some sort of *a priori* insight. In this view, something is *a priori* justified if it is directly justified – not just justified by other beliefs, methods, goals, and plans for which one has justification – and not justified empirically by present sensory or perceptual experiences. One problem with this explanation is that the characterization of *a priori* justification is completely negative. A related problem is that strict foundations theories of this sort imply extreme skepticism and so, if interpreted methodologically, imply that you should give up almost all your current beliefs, methods, plans, and goals, which is really quite crazy!

The other way I know to explain a relevant sense of *a priori* supposes that some claims are (at least default) justified simply because of the meanings of the words used to make those claims, perhaps because one intends to use those words in such a way that those claims are justified. But it would seem that accepting any claims is in part accepting that the claims are justified, so it would seem that for any such claim there is a sense in which one intends to be using words in such a way that the claim is justified. Furthermore, given general conservatism, any claim one accepts is at least default justified, so it is unclear what methodological difference analyticity is supposed to be making.

Those who have been brought up to believe in an epistemically important analytic–synthetic distinction may well point out that these brief remarks do not refute the importance of the distinction. At best they show a need for an account of the distinction and its epistemological significance.

However, I have never seen an account that isn't obviously defective. So I have my doubts.

#### 4. PRAGMATISM

There is an important distinction between epistemic reasons to believe something and non-epistemic reasons to believe something. A study showing that smoking cigarettes is not statistically associated with higher risk of cancer might provide an epistemic reason to believe that cigarette smoking does not cause cancer. Learning that the company you would like to work for only hires people who believe that cigarette smoking does not cause cancer might provide you with a non-epistemic reason to believe it. Although this non-epistemic reason to have a belief is a practical reason, it is not obvious that a practical reason to believe something is always a non-epistemic reason. For example, considerations of conservatism and simplicity may be both practical and epistemic.

Some theorists suppose that epistemic reasons are actually practical reasons associated with an "epistemic" goal, such as the alleged goal of believing something if and only if it is true. In this view, theoretical reasoning, which affects beliefs, is properly seen as a special case of practical reasoning. This is actually quite implausible, however, given certain obvious differences between theoretical and practical reasoning, the most serious being that good practical reasoning almost always involves arbitrary decisions in a way that good theoretical reasoning cannot do so. In deciding where to go after work it is not enough to decide to go home, because there may be several equally good possible ways to get there, so an arbitrary choice must be made and it would be irrational not to make such choices. But in trying to decide which route home someone else took, where there are several possible routes with nothing favoring one rather than another, it would be irrational arbitrarily to believe he took a certain particular route.

Furthermore, as Hempel observed in his classic paper "Inductive Inconsistencies", the

alleged goal of believing something if and only if it is true might be interpreted as implying that you should believe nothing whose probability on your evidence is less than .5 and everything whose probability is more than .5. And, since this is practical reasoning, you can arbitrarily choose whether to believe those propositions with probability of exactly .5. Various other suggestions have been made about what supposed epistemic goals there might be, none very plausible. So I have trouble believing that theoretical reasoning can be reduced to practical reasoning. On the other hand, I do think that certain practical considerations are directly relevant to theoretical reasoning and epistemic reasons. I have already mentioned conservatism and simplicity. Furthermore, what it is reasonable to reason about can depend on practical considerations. And whether it is reasonable to end inquiry and accept certain conclusion can depend on practical considerations concerning how urgent it is to get an answer, what resources are available for further inquiry, and what other questions you want to answer.

#### 5. RELIABILITY

You use inferential methods that in some sense you take to be reliable. The thought that a belief of yours is the result of using an unreliable method would be in tension with the belief in question, providing some reason to abandon either the thought or the belief. Furthermore, in trying to reach an appropriately wide reflective equilibrium, you need to take into account how reliable are the various methods you have used in arriving at your various beliefs.

The reliability of certain methods has been studied empirically and compared with human performance, with such results as that simple linear models often do better than experts at predicting the success of medical interventions, predicting criminal recidivism, predicting tomorrow's weather, predicting academic performance, predicting loan and credit risk, or predicting the quality of a French wine vintage.

Statistical learning theory provides a more theoretical and philosophically deeper approach to the study of inferential methods. It provides an account of reliability in terms of a (usually unknown) background probability distribution. It provides a clear account of why inferential methods must have some sort of inductive bias favoring some hypotheses over others. It distinguishes different sorts of inductive inference with importantly different properties. It provides useful worst-case results concerning various inductive methods.

In one method of inferring a hypothesis from data, you do not simply accept that hypothesis that best fits the data but you balance that consideration against another consideration sometimes called “simplicity”. Popper argued that simplicity is connected with the “falsifiability” of the claim that the true hypothesis is a member of a certain class of hypotheses. Popper claims that the class of linear functions of one variable is simpler than the class of quadratic hypotheses, because the number of data points needed to get a result that does not perfectly fit some linear hypothesis is less than the number needed to get a result that does not perfectly fit some quadratic hypothesis. In statistical learning theory, this idea is captured more precisely by the important notion of “VC dimension”.

Popper thinks that what he calls the falsifiability of a class of hypotheses is connected with the number of parameters that have to be determined in order to specify a particular member of the class. Statistical learning theory rejects that part of Popper’s theory, noting for example that sine curves can be specified using two or three parameters even though the “falsifiability” or VC dimension of the class is very high. One thing this means is that, if sine curves are examples of simple curves, simplicity is not what should be balanced against fit to data.

Statistical learning theory has many other aspects, including discussions of learning in feed-forward neural networks, with which philosophers have some familiarity, and the much more effective support vector machines, about which philosophers tend to be ignorant. Statistical learning theory also discusses

“transductive methods” that do better than inductive methods that use labeled data to come up with general rules for classifying new unlabeled cases, where the transductive methods use information about what new unlabeled cases have come up to be classified. The theory of transduction has implications for discussions of moral particularism and for the question whether all inductive inference involves some sort of inference to the best explanation. But I do not have space here to elaborate.

## Peter Klein

### INTRODUCTION<sup>1</sup>

In my sophomore year in college, I read the *Meno*, *Meditations*, and *Posterior Analytics*. The topics raised in those books provide the basis for the three issues that I have written about: What is knowledge? How ought one respond to various forms of skepticism? What is the structure of justification?

Those questions are interrelated. Without first determining what knowledge is, you cannot reasonably discuss its extent; and I take it that having the right kind of justification for our beliefs is the hallmark of what is the most highly prized form of true belief (*Meno*, 97d). I think the “tether” referred to by Plato (*Meno*, 98a) that ties belief to the truth and distinguishes knowledge from the mere conjunction of truth with belief can be construed as reasoning that is sufficiently resistant to defeat.

### THE DEFEASIBILITY THEORY<sup>2</sup>

Gettier characterized the traditional definition of propositional knowledge as true, justified belief.<sup>3</sup> He presented clear counterexamples to the sufficiency of those conditions. There will be cases in which S has a justified, false belief and comes to justifiably believe a truth by deduction. In some of those cases, such a deduced belief will not be knowledge. The tether is either missing or not sufficiently strong.

Some epistemologists took the moral to be that false beliefs cannot play an essential role in the production of knowledge. I do not think that solution is correct and will return to that in a moment.

Gettier showed that it is relatively easy to construct cases of true, justified beliefs that fall short of knowledge because they are luckily held. But some kinds of luck are OK. If, when looking for my lost penny, an unusual rock catches my eye and I happen to find my penny lying next to the rock, there is a sense in which I come to know my penny's whereabouts by luck. That kind of luck does not prevent knowledge. By contrast, Gettier Cases are ones in which the kind of luck involved obstructs us from obtaining knowledge.

The fundamental disagreement between the reliabilists and defeasibility theorists concerns the characterization of such obstructive luck. Roughly, the reliabilists think that obstructive luck obtains when the *type* of process that resulted in the true belief is not sufficiently likely to produce true beliefs in similar circumstances. By contrast, defeasibility theorists think that obstructive luck obtains when, given all the relevant evidence for and against *p*, it is a lucky break that *S*'s justification of *p* results in *S* arriving at the truth. Roughly, this occurs when there is some true proposition, *d*, such that the conjunction of *d* with the propositions that provided the justification of *p* for *S* fails to justify *p*.

There are well-known problems with each attempt at resolving the Gettier Problem. Reliabilism must solve the generality problem: Since each *token* of a process that produces a true belief is an instance of both a reliable and unreliable *type* of process, what, in general, makes it the case that a token is properly classified as an instance of a reliable type? The defeasibility theory must solve the problem of misleading defeaters: Since there are some true propositions that defeat the justification but they do not preclude the belief from being certifiable as knowledge, what, in general, distinguishes genuine from misleading defeaters?

I will return to the problem for the defeasibility theory below; but, first, a comment

about an advantage the defeasibility theory has over reliabilism with regard to the generality problem. Let us grant that there is some description, *f*, of the facts which makes it appropriate to classify a process as an unreliable one. For example, in the Phony Barn Case, *f* is *there are lots of phony barns around*. Note that *f* is a defeater. To generalize: the proposition describing the *particular feature* of the circumstance in virtue of which the *token* process is unreliable is a defeater. Thus, because there will always be a defeater representing the token process when the belief is unreliably produced, the defeasibility theory can capture the intuitions underlying reliabilism without incurring the liability of the generality problem. Score one for the defeasibility theory!

Now to the problem for the defeasibility theory. Consider the example by which Lehrer and Paxson first introduced the problem.<sup>4</sup> *S* has good visual evidence for the claim that Tom Grabit stole a book from the library. But demented Mrs Grabit, Tom's mother, sincerely says that

i *it wasn't Tom, rather it was his identical twin, Buck, who is a kleptomaniac.*

So there is the true proposition

g *Mrs Grabit said that i.*

Note that *g* is a defeater. The problem is that, although *g* is true and defeats the justification, it doesn't block *S* from having knowledge because Mrs Grabit is just plain nuts. There is no twin – he's merely a figment of her imagination. It is a misleading defeater.

The problem is to provide a general characterization of misleading defeaters. My suggestion was/is this: a defeater is misleading *iff* it defeats only by rendering plausible a false proposition. For my purposes here and elsewhere in this essay, we can say that *x renders y plausible to S* just in case, given *S*'s other beliefs, *x* makes *y* such that *S* must take *y* into consideration when evaluating the truth of her other beliefs. In this case, it is only because *g* renders *i* plausible that *S* would



no longer be justified in believing that Tom stole the book. The power of *g* to defeat the justification is parasitic on a falsehood. Put another way, when the truth alone doesn't defeat but does so *only* indirectly by making a falsehood plausible, which directly defeats, the defeater is misleading.

One crucial point to notice is that if there is no genuine defeater of the justification for a proposition – say, *p* – then *p* is true because, if *p* were false,  $\sim p$  would be a genuine defeater. Thus, there cannot be an accidental connection between a knowledge-producing justification and the truth condition since such a justification must lead to the truth. That solves the Gettier Problem.

Or – so I thought. But I now think that won't quite do because of the problem of useful falsehoods. Sometimes we gain knowledge by inference from a believed, false proposition even though the denial of the falsehood is a genuine defeater. Here's a case I have used before.<sup>5</sup> Suppose, on Christmas Eve, Virginia's parents tell her that

*c* Santa Claus will put presents under the tree tonight.

She cheerfully goes to bed, knowing that

*p* there will be presents for her in the morning.

She reasons that since her parents told her that *c*, thus *c* is true, and since *c*, *p*. But her belief, *c*, is false, and if the truth,  $\sim c$ , is conjoined with *c*, the result is a proposition (*c* &  $\sim c$ ) that does not justify *p*.<sup>6</sup> Thus, Virginia knows that *p* even though she inferred *p* from a falsehood and the justification is defeated.

If  $\sim c$  were a misleading defeater, the defeasibility theory as formulated above would be OK. But it isn't a misleading defeater. It doesn't defeat *only* by rendering a false proposition plausible that, in turn, is the effective defeater. It is debatable, I suppose, whether it does render a false proposition plausible, namely, *no one will put presents under the tree*, but it clearly does not defeat *only* because it does so. It also defeats because it is the negation of a crucial proposition in the evidence path that justifies Virginia in believing that

there will be presents. Thus, because  $\sim c$  is a genuine defeater, the defeasibility theory as usually formulated will incorrectly classify Virginia's belief that *p* as ignorance.

How should the defeasibility theory be modified? A key to the answer depends upon noting that there is a non-defeated route from Virginia's evidence for *c*, namely *my parents told me that c*, to *p* that she could have taken but didn't. She *could have* reasoned this way: Since my parents tell me that Santa will put presents under the tree, then it is true that

*c\** someone will put presents under the tree

and since *c\**, *p*. Virginia's evidence for *c* is also evidence for *c\**, and  $\sim c$  is not a defeater of *c\**'s justification for *p*. Thus, even if the justifying proposition, *c*, for *p* is false, whenever there is a genuinely undefeated alternate justification path available for *S* to *p* from whatever justifies *c*, *S* gains knowledge that *p*. Call any false proposition that is so related to a genuinely undefeated justification path a "useful falsehood".

Hence, my suggestion for an account of *inferential* knowledge is that *S* knows that *p* on the basis of *e*, just in case:

K1: *p* is true (although this condition is not needed as pointed out above).

K2: *S* believes *p* on the basis of an adequate justification, *e*.

K3: Either there is no genuine defeater of *e*'s justification of *p* or *e* is a useful falsehood.<sup>7</sup>

I have limited the scope of the definition above to *inferential* knowledge. I think all knowledge that is the most highly prized form of true belief is inferential, so I think this is actually a perfectly general account of knowledge. I'll return to this when I discuss infinitism.

#### ACADEMIC/CARTESIAN SKEPTICISM<sup>8</sup>

Contemporary treatments of academic skepticism have focused on the Evil Genius and

Dream Arguments. But, at least in the “First Meditation”, Descartes explicitly rejects the dream argument as a basis for global skepticism about all external world propositions and he makes clear that the evil genius hypothesis is mere a heuristic device for recalling the outcome of the argument for global skepticism. The evidential basis for skeptical doubts arises from *within* his own metaphysical presuppositions – namely that since he has fallen into error it is plausible (in the sense defined above) to believe, at this point in the *Meditations*, that whatever is responsible for the nature of his epistemic equipment is less than perfect and, therefore, none of his beliefs is sufficiently worthy of his confidence (until he can show that the creator of his epistemic equipment is perfect or perfect enough so as to have provided him with knowledge-producing epistemic equipment).

By contrast, contemporary switched-world skepticism begins by merely imagining a far-fetched situation for which we have no evidence at all, either in a local form with mules being painted to look like zebras or in a global form with brains-in-vats that seemingly experience a virtual reality indistinguishable from our own. Unlike Descartes’ plausible hypothesis that his “Author” is less than perfect, there is no evidence whatsoever for thinking that the zebra-looking animals are really mules or that I am floating in a vat hooked up to a computer.

The contemporary argument for academic skepticism is this:

CAAS1: If I am justified in believing that I have hands, then I am justified in denying that I am in a switched-world having all the same evidence as I do now.<sup>9</sup>

CAAS2: It is not the case that I am justified in denying that I am in such a world.

Therefore, I am not justified in believing that I have hands.

There are only four epistemically responsible ways of responding: (1) Reject a premise; (2) claim the argument is not valid; (3) find

some other problem with the argument; (4) or accept the conclusion.

I have argued that the third option is the correct one. There are at least two ways in which an argument could be valid and have true premises but fail to provide a basis for accepting the conclusion: it could beg the question or it could have an unmotivated essential premise. If a premise employs the conclusion or if there is no reason for accepting an essential premise (i.e. one without which the conclusion is not entailed by the other premises), the argument adds no credibility to the conclusion. In this particular case, I think it can be shown that the argument is such that, given what *must* be true if the first premise is true, either every argument for the second premise would beg the question or there is no good motivation for accepting the first premise.

Let me begin by noting that the first premise is an instance of closure. So the question becomes this: Why would the skeptic think that if S is justified in believing an entailing proposition, x, then S is justified in believing any entailed proposition, y? It seems to me that the answer must be that there are only three possible evidential relations between any instantiation of x and any instantiation of y such that if any one of the relations held for each instantiated pair {x,y}, then closure would be true:

- (1) x is in the evidential ancestry of y (and x is sufficient to justify y);
- (2) y is in the evidential ancestry of x;
- (3) there is some common proposition, e, in the evidential ancestry of both x and y, such that e makes S justified in believing x *iff* e makes S justified in believing y.

So the skeptic would be claiming that CAAS1 is true because for every pair of propositions {x,y} one of those three evidential relationships obtained.

This seems plausible to me. There seem to be good examples of each:

- (1) In Euclidean Geometry, theorems prior to theorem 15 entail theorem 15; and, if S is justified in believing theorems

- 1–14, S is justified in believing theorem 15 precisely because theorems 1–14 are adequate evidence for theorem 15.
- (2) Suppose I am testing whether a certain hypothesis, H, is correct and know that necessarily, if H, then r, where “r” stands for some experimental result. In other words,  $H \rightarrow r$ . One could (partially) confirm H by appealing to r. In this case, r is in the evidential ancestry of H.
  - (3) Suppose only two types of animals frequent a field – a zebra and a mule cleverly disguised to look like a zebra. It would seem that what justifies me in believing the animal is a zebra does so *iff* it also justifies me in believing that the animal is not a cleverly disguised mule.

Now, whether those examples are convincing does not matter for my argument. What matters is that if I am right that these are the only three evidential ancestry relationships that can obtain between entailing and entailed propositions such that CAAS1 is true, then the contemporary academic skeptic must maintain that one of those three relationships holds between the “I have a hand” and “I am not in a switched world” in order to defend CAAS1. Now, if that is correct, then I think it can be shown that the argument for the second premise must beg the question or the first premise must be unmotivated.

Suppose the skeptic says that what makes premise CAAS1 true with regard to the relevant pair of propositions is that “I have a hand” is in the evidential ancestry of “I am not in a switched world” and is sufficient to justify it, then the only way the skeptic could argue for CAAS2 is by first showing that S is not justified in believing “I have a hand”. But, then, the argument for CAAS2 would employ the conclusion – clearly an instance of begging the question.

Suppose the academic skeptic thinks premise CAAS1 holds because there is some common evidence in the evidential ancestry of both “I have a hand” and “I am not in a switched world”. That is, suppose the skeptic in arguing for CAAS2 would have to show that there is no proposition that provides adequate evidence for either “I have a hand”

or “I am not in a switched world”. If so, the skeptic would have already established the conclusion in the argument for CAAS2. Again, the conclusion is reached before CAAS2 begins and the argument adds no credibility to the conclusion because all the work is done in the sub-argument for CAAS2.<sup>10</sup>

Finally, suppose the academic skeptic thinks that CAAS1 is true because “I am not in a switched world” must be in the evidential ancestry of “I have a hand”. The skeptic can try to argue for that in two ways:

W1: The skeptic could claim that the denials of all contraries to h, *I have a hand*, are in the proper evidential ancestry of h. Now, not only does that not seem to capture our epistemic practices and rules, but it has the consequence that if some evidence, e, is to justify h, e must entail h. To see that, let “a” stand for any proposition that neither entails nor is entailed by h. If the skeptic were correct that all contraries of h must be in e, then, because  $(\sim h \ \& \ a)$  and  $(\sim h \ \& \ \sim a)$  are both contraries of h, their denials must be in e. Thus,  $e = \{ \sim(\sim h \ \& \ a), \sim(\sim h \ \& \ \sim a), \dots \}$  It is easy to see that e entails h. Thus, the skeptic would be requiring that the evidence for h must entail h. That is surely to ramp up the requirements for empirical knowledge too high. Hence, the first premise would be unmotivated.

W2: The skeptic could make a more modest claim, namely that, although it is not necessary to have the denial of all contraries of h in the evidential ancestry of h, it is necessary to have “I am not in a switched world” in the evidential ancestry. It is, of course, logically possible that we are in a switched world. But the mere logical possibility can’t be what makes it a contrary the denial of which has to be in my evidence for h because that simply converts W2 into W1.

Indeed, here is the point where Cartesian skepticism is motivated in a way that the

contemporary switched-world argument is not. The Cartesian skeptical hypothesis is plausible (in the sense defined earlier), given the metaphysical presuppositions concerning perfection that were accepted by the “mediator” in the “First Meditation”. Thus, there is an obligation placed on anyone who shares the same presuppositions to remove that basis for doubt before becoming justified in believing that he/she has hands. Without doing that, there would always be the nagging doubt that perhaps his/her epistemic equipment is seriously flawed. But note that the contemporary switched-world hypothesis is deliberately constructed so that there can be no evidence that it is true. There are NO traces left by the switched-world makers.<sup>11</sup> There is no plausibility whatsoever attached to the contemporary skeptical hypothesis. Thus, since it is not plausible, there is no obligation to consider it, much less eliminate it, before I am entitled to believe that I have hands. So it strikes me that W2 cannot provide a motivation for the first premise.<sup>12</sup>

In sum: I think the contemporary switched-world argument for skepticism cannot provide a basis for accepting the conclusion that I am not justified in believing that I have hands. Perhaps there is some other argument that provides a basis for accepting academic skepticism. My point is only that the most commonly employed contemporary argument cannot do that.

#### INFINITISM<sup>13</sup>

I have developed and defended infinitism as the solution to the epistemic regress problem. Infinitism is the view that there is no last member in the set of justificatory reasons for our beliefs. Of course, in any actual reasoning session, we do stop giving reasons. So the claim is not that we don’t stop. Of course we do. The claim is that no reason has the privileged status of being basic – i.e. no reason is such that it is justified (in the sense that contributes to a true belief being the most highly prized kind) without a further reason. In other words, infinitism holds that all propositional knowledge is inferential.

Such a view will strike many as so highly implausible that it isn’t worth investigating. My aim has been to make it plausible (in the sense defined earlier) so that others will investigate it as thoroughly as foundationalism and coherentism have been investigated.

The epistemic regress problem can be stated this way: Some beliefs require inferential justification. No belief justifies itself. So either the justificatory path is infinite, or circular, or ends with a belief that is justified but not on the basis of another belief, or it ends with an unjustified belief.

Now, *if justification were either a property inherent in a belief or a property acquired through transfer from one belief to another by inference*, then the only plausible solution to the regress is foundationalism. Infinitism would be conceptually opaque because it could not explain how justification arises in the first place. Circular reasoning could not produce a justified belief. Finally, if the foundational belief were not justified, there would be no way for inference to transfer justification from the foundational belief to non-foundational ones. Hence, there must be foundational beliefs that are justified, but not in virtue of acquiring their justification from another belief.

So goes the argument for foundationalism. I think it does not succeed. To see why, let us say that a belief is foundational just in case it has a certain feature or set of features, F. F could be any property you wish. For example, a foundationalist could hold that a belief, e.g. “there is a table before me”, is foundational just in case it arises when I am having a table-like experience and I am aware of nothing odd in the “external” circumstances or with my perceptual equipment.

Now, I grant that there is some positive epistemic property that a belief has in virtue of its being produced by a reliable process. But, as I mentioned at the outset, I do not take that to be sufficient to certify it as the most highly prized form of belief.

Suppose Fred, the foundationalist, arrives at a basic belief, b, where a belief is basic just in case it has F. Now, Fred is asked (or wonders) whether he is a good F-detector and whether b’s having F makes it likely that b is true. Fred has five alternative responses: “Yes”,

"No", "I withhold", "there is something false presupposed by your question"; or he could say he's not interested in answering that question.

I take it that the last response is an abdication of the epistemic project – seeking to have the highest form of true belief. The penultimate response is appropriate to some questions, e.g. "when did you stop robbing banks?" – but the question posed to Fred doesn't seem to be like that.

So only the first three responses seem appropriate. Now, if Fred answers either "no" or "I withhold", he could stubbornly cling to believing that *b*, but the skeptical worry that has just been implanted is likely to gnaw at him until he sees that he had better answer "yes". Of course, once he does that, the obvious next question is: Why do you believe that you are a good *F*-detector and that *b*'s having *F* is truth-conducive?

Fred can only arbitrarily stop the regress at *b*. Hence, foundationalism is not a plausible solution to the regress problem. If that is correct, then we are forced to rethink the assumption that *justification is either a property inherent in a belief or a property acquired through transfer from one belief to another by inference*.

Contemporary "holistic" coherentism rejects that assumption; but, as I and others have pointed out, coherentism is nothing but a form of foundationalism in which the *F* property is coherence. Hence, the coherentist faces the same problem that Fred faces when asked whether she is a good coherence-detector and whether coherence is truth-conducive. She needs an answer to that and so must add another belief to her system of beliefs. Hence, the regress continues for her as well.

What is the infinitist's alternative to the transference assumption? It seems to me that at least part of what makes a true belief so highly prized is that we have justified it; and we justify a belief by providing reasons for it. But justification comes in degrees, and the infinitist will suggest that the further along the path of reasons we have traveled, the more justified the proposition becomes.<sup>14</sup> An infinitist can argue that on some occasions

we shall have traced the path of reasons far enough to reach the threshold of the strength of justification required to satisfy *K2* in the definition of knowledge mentioned earlier.

Though knowledge is obtainable, my brand of infinitism eschews the possibility of obtaining absolute certainty – where that means that some proposition is absolutely certain only when it is as justified as it could possibly be. There is always a further step on the path of reasons. Nevertheless, such infinitists remain skeptics in the original sense of the term (i.e. inquirers) because nothing is ever finally settled, and the last offered belief in the path of reasons can legitimately be questioned.

#### NOTES

- 1 I want to thank Matthias Steup for his very useful suggestions that went well beyond those normally supplied by an editor.
- 2 For more detailed discussions, see Klein, 1971, 1976, 1979, 1980, and 2004a.
- 3 Gettier, 1963.
- 4 Lehrer and Paxson, 1969.
- 5 See Klein, 2008. For an alternate account of useful falsehoods, see Warfield, 2005.
- 6 I take it that a contradiction cannot justify anything.
- 7 For sake of brevity I have condensed the justification condition by supposing that "e" stands for the entire justification path. With more space, I could spell out the various steps along the path in *K2*; and in *K3* indicate that if one of them is genuinely defeated, then it employs a useful falsehood.
- 8 For more detailed discussions, see Klein, 1995, 2000a,b, and 2003.
- 9 Here "justified in believing" means that *S* is justified in believing *p* just in case *S* has adequate reasons available for believing *p*. *S* might not actually believe *p* or might believe *p* for the "wrong" reasons.
- 10 Whether this is a case of begging the question or something closely akin to begging the question because it shares the same basic flaw that is exhibited in begging the question strikes me as merely a matter of taste. For simplicity, I think it can be taken to be a case of begging the question.
- 11 Similarly, there is no reason to believe that the zebras, made famous by Dretske, are painted mules.



- 12 A stronger claim would be that premise CAAS1 is false because it requires more than what can be legitimately required in order to be justified in believing that I have hands. But I have not claimed that CAAS1 is false. Unmotivated premises might be true.
- 13 For more detailed discussions, see Klein, 1998, 1999, 2000c, 2004b, 2004c, 2005, 2006a, and 2006b.
- 14 See Fantl, 2003.

## BIBLIOGRAPHY

- Fantl, Jeremy: "Modest Infitism," *Canadian Journal of Philosophy* 33 (2003), 537–62.
- Gettier, Edmund: "Is Justified True Belief Knowledge?," *Analysis* 23 (1963), 121–3.
- Klein, Peter: *Certainty: A Refutation of Scepticism* (Minneapolis, MN: University of Minnesota Press, 1981).
- Klein, Peter: "A Proposed Definition of Propositional Knowledge," *Journal of Philosophy* 67: 16 (1971), 471–82.
- Klein, Peter: "Knowledge, Causality and Defeasibility," *Journal of Philosophy* 73: 20 (1976), 792–812.
- Klein, Peter: "Misleading 'Misleading Defeaters,'" *Journal of Philosophy* 76: 7 (1979), 382–6.
- Klein, Peter: "Misleading Evidence and the Restoration of Justification," *Philosophical Studies* 37 (1980), 81–9.
- Klein, Peter: "Skepticism and Closure: Why the Evil Genius Argument Fails," *Philosophical Topics* 23: 1 (1995), 213–36.
- Klein, Peter: "Foundationalism and the Infinite Regress of Reasons," *Philosophy and Phenomenological Research* 58: 4 (1998), 919–25.
- Klein, Peter: "Human Knowledge and the Infinite Regress of Reasons," *Philosophical Perspectives* 13, *Epistemology*, ed. J. Tomberlin (1999), 297–325.
- Klein, Peter: "Contextualism and the Real Nature of Academic Skepticism," *Philosophical Issues* 10 (2000a), 108–16.
- Klein, Peter: "The Failures of Dogmatism and a New Pyrrhonism," *Acta Analytica* 15: 24 (2000b), 7–24.
- Klein, Peter: "Why Not Infitism?," in *Epistemology: Proceedings of the Twentieth World Congress in Philosophy*, ed. Richard Cobb-Stevens, vol. 5 (2000c), 199–208.
- Klein, Peter: "How a Pyrrhonian Skeptic Might Respond to Academic Skepticism," in *The Sceptics: Contemporary Essays*, ed. Steven Luper (Aldershot: Ashgate, 2003), 75–94.
- Klein, Peter: "Knowledge Is True, Non-defeated Justified Belief," in *Essential Knowledge*, ed. Steven Luper (Harlow: Longman, 2004a), 124–35.
- Klein, Peter: "What IS Wrong with Foundationalism Is That It Cannot Solve the Epistemic Regress Problem," *Philosophy and Phenomenological Research* 68: 1 (2004b), 166–71.
- Klein, Peter: "Closure Matters: Skepticism and Easy Knowledge," *Philosophical Issues* 14 (2004c), 165–84.
- Klein, Peter: "Infitism Is the Solution to the Epistemic Regress Problem," in *Contemporary Debates in Philosophy*, ed. Matthias Steup and Ernest Sosa (Oxford: Blackwell, 2005), 131–40.
- Klein, Peter: "Human Knowledge and the Infinite Progress of Reasoning," *Philosophical Studies* 134: 1 (2006a), 1–17.
- Klein, Peter: "How to Be an Infitist about Doxastic Justification," *Philosophical Studies* 134.1 (2006b), 25–9.
- Klein, Peter: "Useful False Beliefs," in *Epistemology: New Essays*, ed. Quentin Smith (Oxford: Oxford University Press, 2008).
- Lehrer, Keith and Paxson, Thomas: "Knowledge: Undefeated Justified True Belief," *Journal of Philosophy* (1969), 225–37.
- Warfield, Ted: "Knowledge from Falsehood," *Noûs* (Supplement) 19 (2005), 405–16.

## Hilary Kornblith

My work has been devoted to the articulation and defense of a thoroughgoing naturalistic epistemology. As I see it, traditional epistemological questions are most productively addressed by methods which are continuous with those of the empirical sciences. Such an approach has broad implications for many topics of long-standing interest within epistemology.

## FOUNDATIONALISM AND THE COHERENCE THEORY

A good deal of epistemological theorizing has been occupied with the debate between foundationalist and coherentist theories of justified belief, and for good reason. On the one hand, each of these views has features which are undeniably appealing; on the other, there are problems for each view which are arguably insurmountable. To make matters

worse, many theorists assume, and some argue for, a taxonomy of views about justification which make the choice between foundationalism and the coherence theory exhaustive of the reasonable options. I have long believed that these two views have a great deal more in common than is often recognized, and that the taxonomy which forces all views about justification into one of these two categories itself presupposes a mistaken view about the nature of justified belief. In particular, both of these views are committed to what I call (1980) the Arguments-on-paper View of Justification: to be justified in holding a belief is to be in possession of a good argument for it. The differences between the two views arises in their characterization of what constitutes a good argument: foundationalists require that a justified belief be derived from a privileged subset of one's beliefs by way of a favored set of principles of inference; coherentists allow all of one's beliefs to serve as premises of a complex non-monotonic argument. Either way, justification is identified with possession of argument of an appropriate sort.

The mistake which is common to these two views is best-illuminated by way of the reliability account of justification championed by Alvin Goldman (1979). No one should deny that possession of an appropriate argument can serve to justify a belief. On the reliability account, however, which identifies justified belief with belief which is the product of a reliable process, forming a belief as a result of being in possession of a good argument is just one way among many to arrive at justified belief; it is, in addition, highly atypical. Ordinary subjects will often arrive at accurate judgments about the nature of their environment without being in possession of good arguments for those beliefs; more than this, the accuracy of these beliefs is no accident, for they are often the product of highly reliable psychological processes. If we insist that these beliefs do not count as justified because they are not backed by appropriate arguments, we shall be committed to the view that, at best, very few of the beliefs which we ordinarily form are genuinely justified. There is an important distinction we wish

to draw between those beliefs which are justified and those which are not. Reliability theories make sense of the importance of this distinction, while foundationalist and coherentist theories do not.

#### INTERNALISM AND EXTERNALISM

The issues involved in the debate between foundationalists and coherentists, on the one hand, and reliability theorists, on the other, form the constellation of issues now known as the internalism/externalism debate. My sympathies here are very firmly on the side of externalism. The arguments in favor of externalism which I find most persuasive are straightforwardly empirical in character.

Consider, for example, the coherentist version of internalism defended by Laurence Bonjour in his (1986). Bonjour argues there that, in order for an agent's belief to be justified, it must not only cohere with the agent's other beliefs; it must also be held in virtue of that agent's recognition that it coheres. Bonjour, in the light of this account, offers a bit of epistemic advice: if one wants to arrive at one's beliefs in a responsible way, one should introspect in order to determine the content of one's beliefs, and then try to determine whether the candidate belief actually coheres with one's other beliefs. Candidate beliefs which are found to cohere should be accepted; all others should not be. Similarly, one might defend a foundationalist version of the same strategy, as Bonjour does in his more recent work (2002). One should reflect on the contents of and logical relations among one's beliefs in order to determine whether candidate beliefs meet appropriate foundationalist standards: all and only those candidates which are found to meet appropriate standards should be accepted. The internalist idea here is that introspection and reflection are put to work on a certain problem: the problem of determining which candidate beliefs are worthy of being held. Any such view takes for granted that our intellectual faculties will perform well in this task, for the point of reflection, for internalists such as Bonjour, is to make our beliefs more accurate. And this

just raises the empirical question of what the likely effect would be were we to reflect on the content and logical relations among our beliefs in the way that such internalists advocate.

#### LIMITATIONS ON INTROSPECTION AND REFLECTION

Descartes and other rationalists were extraordinarily optimistic about the native powers of the human mind. Many contemporary foundationalists and coherentists are similarly committed to this unfounded optimism: the assumption that introspection and reflection may be put to work as effective agents of epistemic quality control embodies this optimistic view.

There is, however, a good deal of experimental evidence about the likely effects of the kind of reflection which internalists advocate (*see* Nisbett and Wilson, 1977; Wilson, 2002; Kornblith, 2002, ch. 4). The processes by which our beliefs are formed are not directly available to introspection, although individuals typically have the impression that they are able to tell how their beliefs arise. Judgments we make about these processes are influenced by a wide range of biasing factors, factors unavailable to the reflective agent. Subjects asked about the source of their beliefs will, in a wide range of cases, sincerely cite factors which play no role in producing or sustaining their beliefs, while simultaneously demonstrating an ignorance of some of the most important causal factors in the processes of belief acquisition and maintenance. By and large, the sources from which agents take their beliefs to arise, and which they sincerely believe they can directly introspect, are processes which would justify their beliefs, if only they were the processes by which their beliefs were produced. So, when individuals turn reflective in the way many internalists encourage, they are likely to come to the conclusion that their beliefs are justified, whether those beliefs were arrived at in a reasonable manner or not. Introspection and reflection, in a very wide range of cases, do not make the reflective agent more accurate; they simply make agents, including

those whose beliefs are not only mistaken but deeply irrational, more self-confident.

There is nothing wrong with the internalist idea that we ought to be concerned about the ways in which we arrive at our beliefs; indeed, there is everything right about it. When the processes of belief acquisition which we employ are reliable, we should continue to make use of them; when they are unreliable, we need to make changes. The problem with internalism is that this very important task of epistemic self-evaluation is given over to introspection and reflection, which are singularly ill-suited to the task.

The cognitive sciences allow us to understand how it is that our minds actually operate, and allow us to see where it is that our native inferential tendencies are in need of correction. Since the mind is not transparent to reflection, and since the process of reflection so badly misleads us about the relevant features of belief acquisition, the process of reflection cannot be substituted for a scientific examination of the mind.

#### OUR NATIVE INFERENTIAL TENDENCIES

There is a traditional view about how we ought to reason, and it presents an ideal of human inference which includes deductive logic, as well as a variety of inductive principles and the probability calculus. When psychologists look at human inference, however, the principles which actually guide our inferences often do not even approximate to these ideals. People have a striking tendency to draw conclusions about a population based on very small samples – indeed, often on the basis of a single instance, contrary to what the probability calculus would counsel. Individuals will often rate the probabilities of conjunctions to be higher than the probabilities of either of their conjuncts. Base-rate information is routinely ignored. Vividly presented information is a powerful engine driving human inductive inference, even when far more telling, but less evocative, information is also available (Gilovitch, Griffin and Kahneman, 2002). Some psychologists have drawn the conclusion that human beings are

natively inclined to reason very badly: the gap between proper reasoning and the sad facts of human inferential competence, they argue, leave no room for any other conclusion.

There is, however, another way to view matters, and I have argued (1993) that some, even if not all, of the fault lies with the proffered ideals, rather than with the inferential tendencies themselves. Traditional philosophical ideals for human reasoning attempt to offer principles of inference which would work well in any logically possible environment. But this is not an ideal to which we should aspire, and when our reasoning falls short of it we should not suppose that there is a problem with the way in which we reason. Neither should we assume that the alternative to this sort of ideal is one of inferential tendencies which just happen to work out well. The environments in which human beings are found have deep constancies determined by the laws of nature; inferential tendencies which tap into those constancies will be reliable, and that reliability itself will not merely be a matter of good fortune.

A single example will have to serve to illustrate this point. Consider our tendency to draw conclusions about a population based on extremely small samples. While such a tendency violates the Law of Large Numbers, its reliability is dependent on the sorts of environments in which it is employed and the kinds of properties to which one applies it. In an environment with natural kinds – kinds which are largely uniform with respect to clusters of properties nomologically bound together – and a tendency to be sensitive to and to project properties which are among those nomologically clustered, inductive inferences based on small samples will be highly reliable. I have argued (1993) that this is in fact how some simple inductive inferences work, and our assessment of the tendency to draw conclusions about a population based on small samples should be made in light of this. We should aspire to make use of principles of inference which dovetail in appropriate ways with the causal structure of the world, rather than principles which would deliver accurate results regardless of the environment in which they are employed.

Such an approach does not by any means dictate that we should approve of any principles of inference with which we happen to find ourselves natively endowed. There is ample evidence that some of our native endowment is not even roughly reliable. The kind of requirement of ecological validity which I favor, however, sets a standard which is both possible to achieve and worth achieving. This is the sort of ideal to which, it seems to me, we ought to aspire.

#### SOCIAL DIMENSIONS OF COGNITION

Traditional epistemology vastly underestimated the extent to which cognition is a social phenomenon. A good deal of science studies, to my mind, now make the opposite mistake by underestimating, and sometimes simply ignoring, the extent to which a good deal of cognition is not social. I have attempted to strike some balance between these extremes (1994). Once one rejects internalism, the idea that the epistemic status of our beliefs should be dependent on various features of our social environment should be recognized as no more controversial than that the epistemic status of our beliefs is dependent on various features of the non-social environment. In both cases, the reliability of our processes of belief acquisition and retention are determined, in part, by things external to the individual.

Consider the social practice of giving and asking for reasons. When individuals express their opinions, they are often asked why it is that they believe as they do. Reasons are often presented, and these reasons are evaluated, at times, by comparing them with reasons offered in favor of contrary opinions. The way in which reasons are characteristically offered and evaluated has a large effect on the manner in which opinions are transmitted within a social group. The reliability of the process is influenced in important ways by a variety of social factors.

Consider the social distribution of bias. Individuals will, at times, form beliefs on the basis of biasing factors, i.e. factors which decrease the reliability of their processes of belief acquisition. Some of these biasing factors

are idiosyncratic: Jack may tend to overestimate the importance of one kind of data, while Mary may tend to overestimate the importance of another. When biasing factors are widely varied, the public discussion of reasons is likely to compensate for such idiosyncrasies, since Mary's mistakes are likely to be especially salient to Jack, and Jack's mistakes are likely to be especially salient to Mary. When a single source of bias is, however, extremely widespread, it is far more likely to go undetected, for those evaluating the biased presentation of reasons are likely to be influenced by the very same bias. The effectiveness of the public discussion of reason in getting at the truth is thus deeply dependent on the social distribution of bias (*see my 1999*).

This view of the social practice of reasoning does not make widespread bias undetectable, but it does make sense of the idea that there might be biasing factors at work behind the scenes, as it were, even when reasons are evaluated in an apparently impartial way. This does not suggest any sort of scepticism about reason, any more than the fallibility of perception inevitably leads to scepticism about the senses. There is, however, interesting experimental work in psychology which examines the social distribution of bias, and this work is relevant to understanding the workings of our practice of giving and asking for reasons. On my view, a certain distribution of biasing factors is a social prerequisite for the proper function of reason.

A proper defense of reason can thus be given, as I see it, not by way of some sort of transcendental argument, but by way of empirical investigation (*see also my 2000*).

#### THE SOURCES OF NORMATIVITY

Naturalistic epistemologists are often pressed to explain how it is that they can make sense of epistemic norms. If epistemology is merely descriptive, it is said, there is a special problem about epistemic prescription. While I agree that naturalistic epistemologists must explain how it is that they may account for epistemic normativity, I do not believe that the problem here is unique to naturalism. Those

who favor a more traditional approach to epistemology, seeing it as an a priori discipline, are not relieved of the task of accounting for epistemic norms. More than this, a license to appeal to the a priori does not make the problem of normativity go away.

I have argued (2002, ch. 5) that epistemic norms are a kind of hypothetical imperative. The hypothetical imperative approach I favor, however, does not make the imperative to form beliefs in certain ways dependent on a desire for true belief or a desire to be reasonable. Any such account would make epistemic norms irrelevant to those who do not share the motivating desires, and it is, I think, reasonable to suppose that epistemic norms are not so easily eluded. Rather, my view is that anyone who has desires at all is bound by epistemic norms – for the simple reason that, whatever one's desires, one will be more likely to succeed in satisfying them if one has true beliefs. Since action informed by true beliefs is more likely to succeed in satisfying one's desires, whatever one's desires may be, everyone who has desires at all – and that is simply everyone – ought to form their beliefs in ways which are likely to arrive at the truth. Epistemic norms, on this view, are hypothetical – since they are contingent on having desires – but they are, at the same time, universal – since they are not contingent on having any particular desires, but contingent only on the having of some desires or other.

#### THE NATURE OF EPISTEMOLOGICAL INVESTIGATION

Epistemologists have often begun their investigations by way of an analysis of our epistemic concepts. Proper method for conducting such an investigation, as many see it, is an examination of our intuitions about actual and hypothetical cases. My own view (2002, ch. 1) is that an examination of our concepts is irrelevant to epistemology. Epistemologists should be no more interested in our pre-theoretical concepts of knowledge and justification than chemists should be interested in our pre-theoretical concepts of atoms or acids.



Chemists have no interest in our pre-theoretical concepts of chemical kinds because chemical kinds are natural kinds, and our pre-theoretical views about natural kinds may be influenced by ignorance and error. Prior to serious empirical investigation, our concepts of atoms and acids were quite wide of the mark. The same, I believe, may be true of epistemic concepts, and for a similar reason. As I see it, knowledge is a natural kind.

Consider work in cognitive ethology. Cognitive ethologists attribute beliefs and desires to many non-human animals, and for much the same reason that we attribute such states to human beings: the complexity of much animal behavior can only be predicted and explained by viewing that behavior as a product of the interaction of complex representational states. But, when it comes to explaining the behavior of individual animals, we need to appeal to nothing more than belief: whether an animal knows that *p* or not is irrelevant to predicting its behavior; so long as it believes that *p*, it will act as if *p* were true.

Nevertheless, cognitive ethologists speak of animal knowledge, and this talk is not merely a *façon de parler*. An evolutionary approach to animal behavior sees the environment as placing certain informational demands on animals, and their cognitive capacities are adapted to serve those needs. Thus, when we wish to explain the cognitive capacities of a species, we shall recognize that certain reliable processes of belief acquisition are the product of evolutionary pressures; the ability of the species to survive in its environment is thus explained, in part, by the presence of reliable capacities to form true beliefs. When such capacities produce the very kind of belief which they were adapted to produce – namely, true beliefs – the result is knowledge. On this view, knowledge is a category in cognitive ethology; talk of knowledge does causal and explanatory work within a scientific theory. It is this very category that we have been talking about all along when we spoke of knowledge, just as human beings spoke of water, a certain chemical kind, long before the real nature of water was properly understood.

Thus, on my view, epistemology is a thoroughly empirical investigation. While traditional philosophical problems remain important, they are, as I see it, best-addressed by straightforwardly empirical means.

#### BIBLIOGRAPHY

- BonJour, L.: *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1986).
- BonJour, L.: *Epistemology: Classic Problems and Contemporary Responses* (Lanham, MD: Rowman and Littlefield, 2002).
- Gilovitch, T., Griffin, D. and Kahneman, D., eds: *Heuristics and Biases: The Psychology of Human Judgment* (New York: Cambridge University Press, 2002).
- Goldman, A.: "What Is Justified Belief?," in *Justification and Knowledge*, ed. G. Pappas (Dordrecht: Reidel, 1979), 1–23.
- Kornblith, H.: "Beyond Foundationalism and the Coherence Theory," *Journal of Philosophy* 77 (1980), 597–612.
- Kornblith, H.: *Inductive Inference and Its Natural Ground* (Cambridge, MA: MIT Press, 1993).
- Kornblith, H.: "A Conservative Approach to Social Epistemology," in *Socializing Epistemology*, ed. F. Schmitt (Lanham, MD: Rowman and Littlefield, 1994), 93–110.
- Kornblith, H.: "Distrusting Reason," *Midwest Studies in Philosophy* 23 (1999), 181–96.
- Kornblith, H.: "The Impurity of Reason," *Pacific Philosophical Quarterly* 81 (2000), 67–89.
- Kornblith, H.: *Knowledge and Its Place in Nature* (Oxford: Oxford University Press, 2002).
- Nisbett, R. and Wilson, T. D.: "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review*, 84 (1977), 231–59.
- Wilson, T. D.: *Strangers to Ourselves* (Cambridge, MA: Harvard University Press, 2002).

#### Keith Lehrer

I was introduced to epistemology in an undergraduate course at the University of Minnesota taught by John Hospers, which focused on texts by G. E. Moore, B. Russell, W. T. Stace, C. I. Lewis and J. Wisdom. Of all the texts I read, the one by Lewis, *Mind and World Order*, impressed me the most. I then went on to study Wittgenstein's *Tractatus*

and *Philosophical Investigations* with W. S. Sellars, who was generous enough to give me a private reading course on his recently published *Empiricism and the Philosophy of Mind*. These early studies were to prove powerfully influential in my own work, particularly Sellars's important insistence that a belief's status of being justified or being an instance of knowledge depends on its relation to the subject's background system.

R. M. Chisholm at Brown focused my study of epistemology on the analysis of evidence, justification and knowledge. Seeking to analyze knowledge in terms of evidence, truth and acceptance, he impressed upon me the need for uncovering principles or criteria of evidence. Chisholm had argued against analytic phenomenalism on the grounds that no pure physical object statement entailed anything about pure appearances. Assuming the truth of phenomenalism, it seemed plausible to argue that the evidence appearances provide for our beliefs about physical objects could be explained in terms of the hypothetical-deductive model of justification. But if physical object statements entail nothing about appearances, then the phenomenalist model of confirmation cannot explain why appearances supply us with evidence.

So what does justify our beliefs about the physical world and our perception of it? Chisholm argued that a certain subset of perceptual beliefs about physical objects as well as beliefs about the ways one is appeared to are self-justified. More worthy of belief than their denials, the justificational status of such beliefs, according to Chisholm, does not result from reasoning but rather is foundational – a view indebted to Reid, who argued that the justification of such beliefs is their birthright. Chisholm held that there are epistemic principles telling us under which conditions a belief is self-justified and thus foundational, and he suggested that such principles are synthetic a priori and, indeed, necessary truths.

As Chisholm developed this account of criteriological foundationalism, his reasoning on behalf of foundationalism convinced me that the coherence theory of justification explains better than foundationalism why our physical object beliefs are justified.

Chisholm was careful to limit the self-justified beliefs to beliefs concerning appearances and very cautious perceptual beliefs about sensible qualities – that one sees something red, for example. He denied that the perceptual belief that one sees a jaguar, of either the animal or mechanical variety, is a belief that is justified without the support of background information.

Chisholm's distinction between foundationally and non-foundationally justified perceptual beliefs led me to suggest a non-foundationalist alternative: the reason why beliefs about appearances and perceptual beliefs about sensible qualities are justified without the support of reasoning, while other perceptual beliefs are not, is a consequence of the risk of error. Consider the following two beliefs: "I see something red" and "I see a jaguar". Whereas justification for the latter depends on reasoning, my justification for the former does not. This discrepancy is a direct consequence of different risks of error associated with each of these beliefs. I am less likely to be in error about whether an object I see is red, or appears red, than about whether what I see is a jaguar.

In response to my claim about the relationship between the risk of error and probability or truth frequency, Chisholm replied that the appeal to the risk of error and the truth frequency of such beliefs would lead to arguing in a circle. I could see the point. One would have to know the truth of particular appearing statements and perceptual statements to arrive at justified claims about truth frequencies of beliefs concerning such matters. Chisholm thought the circularity involved committed one to a coherence theory of justification, and so did I.

Thus, I was pushed in the direction of the coherence theory. Studying under Sellars, I had come to think of the coherence theory of justification as aiming at maximizing explanatory coherence. I was struck with the power of the coherence theory to deal with traditional epistemological problems derived from Hume. Skeptical doubts about memory, the external world and induction seemed to dissolve in the solvent of maximizing explanatory coherence.

Sellars's coherence theory permits us to affirm both general principles and particular instances falling under them for the resulting gain in explanatory coherence. The theory allows us to make particular claims about the perception of sensible qualities and to justify them by appealing to general principles about the truth-frequency of these claims. Moreover, since scientific claims seem to derive their justification from their explanatory role, it seemed to me that Sellars's theory promised to combine common-sense epistemology with the philosophy of science. Finally, in *Empiricism and the Philosophy of Mind*, Sellars already noted the reason for rejecting externalist theories such as reliabilism. Knowledge plays a role in the justification game, which means that knowledge claims are embedded in the discourse of reasoning about truth and error, and reliability of which one is ignorant cannot fulfill the needed function in the discourse.

The solvent of explanatory coherence seemed to me, however, to invite several objections: (i) false hypotheses might maximize explanation; (ii) known truths might be inexplicable; (iii) general correlations might defy explanation and moreover resist any use for explanatory purposes. Nevertheless, I favored the coherence theory. According to the coherentist theory of justification I articulated in *Knowledge*, a target belief is justified just in case challenges to it, which I called *competitors*, can be beaten or neutralized in terms of a background system of belief, which I called a *doxastic* system. Roughly, the idea is the following: A competitor to a belief, whether a local doubt or a general skeptical one, conflicts with what is believed. Conflict is incoherence, and coherence is the absence of conflict. So, if a competitor can be met by the subject, incoherence is reduced. If all competitors to the belief can be met, then conflict and incoherence are resolved and the belief coheres with the background system.

This early work was motivated by an attempt to solve two problems in epistemology that occupied my attention in the 1960s. One was the lottery paradox and the other the Gettier problem. The lottery paradox results when it is assumed that any probability less

than one is sufficient for the justification of a knowledge claim. Consider the claim of a pessimistic ticket-holder to know that his ticket in a finite lottery will fail to win. Such a claim remains unjustified, no matter how large the lottery and, therefore, no matter how high the probability less than unity. For, if the claim were justified, then, by symmetry of reasoning, we would be justified in claiming to know that each ticket will fail to win, when at the same time we know that one of the tickets will win.

My analysis of the paradox was that the individual claims to the effect that each of the tickets fails to win compete with each other. The assumption that any one of these claims is true reduces, and in this way is negatively relevant to, the probability of any other claim. I concluded that negative probability relevance generated conflict or competition, thereby creating incoherence. Each of the claims that an individual ticket fails to win competes with such claims about the other tickets. The competitors to a specified claim cannot be beaten or neutralized, and, therefore, the specified claim is not justified. That solves the paradox. This solution involves two important aspects: first, an implicit appeal to a background system as the basis of the probability assignments (which I thought of as personal estimates of truth frequencies), and second some idealization resulting from the assumption that the probability assignment is coherent. The account was as naturalistic as probability itself.

At the same time, I became interested in the social constraints on probability and justification, and developed with Carl Wagner *Rational Consensus in Science and Society*. It seemed to me that probability assignments constitute one of the constraints on personal probability assignments. Moreover, setting forth a consensual account allowed us to include the special weight given to experts and, in general, the differential weight assigned to individuals depending on their authority and expertise. However, as my views developed further, I began to think of consensual views as informing and shaping individual views as the former are also informed and shaped by the latter. This work is an early and technically

articulated account of social epistemology that I continue to defend.

The other puzzle on which I focused in my early work, the Gettier problem, arises for any fallibilistic account of justification and knowledge. If we are fallible, as we in fact are, then justification cannot guarantee truth, and we can be completely justified in believing something false. Based on such fallibilism, I offered a defense of skepticism in an often reprinted essay, "Why Not Skepticism?". If knowledge requires that we possess a justification that guarantees truth, then we lack knowledge. My conclusion was that to defend or, perhaps, renovate knowledge so as to avoid skepticism we need a fallibilistic conception of justification. Exploiting the assumption of fallibilism, Gettier showed – using examples which I happily developed and multiplied with Thomas Paxson Jr in "Knowledge: Undefeated Justified True Belief" – that one can deduce a truth from a false belief that one was completely justified in believing. Assuming that justification can be extended by the deduction of a conclusion from a justified premise, one can wind up with a completely justified true belief that is not knowledge because the justifying premises are false.

The Gettier problem is of crucial importance to the coherentist because it helps us see why a standard and allegedly devastating objection to coherentism fails. According to the objection, the coherence theory implies that a belief can be completely justified even though it is false. Gettier cases reveal that this is a consequence of *any* theory of fallible justification. To convert justification to knowledge, any such theory requires a truth constraint: a fourth condition of knowledge the purpose of which is to ensure that the materials needed for justification are true.

To meet the challenge of articulating a fourth condition, the coherence theorist must impose some sort of truth constraint on the background beliefs with which the target belief must cohere to yield justification and, ultimately, knowledge. Prior to the Gettier problem, it was unclear what sort of truth constraint was necessary. Clearly, it is unrealistic to require that all background beliefs be true. Attention to the Gettier problem

clarified what was needed, namely, a fourth condition beyond justified true belief requiring that the target belief's justification not depend essentially on any false beliefs. Put in the positive mode, there must be sufficient true beliefs in the background system to avoid the Gettier problem and variations thereof.

So my early account of the coherence theory combined justification based on personal probability estimates with a truth constraint needed to avoid the Gettier problem by overcoming the fallible character of justification. This account differs from versions of coherentism that construe coherence as a property of the background system yielding justification. In contrast, my early account construes coherence as a relationship between a target belief and a background system. It is, therefore, a local and relational theory of coherence rather than a global and holistic one.

There were many valuable objections raised to the theory when *Knowledge* was published. I became convinced that there was a difficulty of substance that needed correction, which motivated the writing of *Theory of Knowledge*. Put simply, a probability assignment can be unreasonable yet consistent – indeed, consistent with the axioms of probability. Nothing in my early account ensured that a belief that the subject's background system rendered probable was also reasonable for the subject, given that the subject's probability assignments themselves might be unreasonable. Moreover, I became convinced that scientific hypotheses, even the best of them, are reasonable to accept, at least in part, because of the systematic advantages of believing them, and not just because of their probability. As a remedy for these difficulties, I replaced the idea that a belief is justified because it is more probable than its competitors with the idea that a belief is justified because it is more reasonable than the objections to it. On the revised account, a target claim can be defended against an objection either because it is more reasonable than the objection or because the objection can be neutralized. An objection to a claim is neutralized just in case the claim conjoined with a neutralizing addition is as reasonable to believe as the objection.

From this account of personal justification, it is easy to see what is needed to convert justification to knowledge and solve the Gettier problem. Though the detailed account is technically difficult, the basic idea is simple enough. Justification or coherence is based on a background system (or, as I came to call it, the *evaluation system*), the importance of which arises from the need to meet all objections to a target claim. The evaluation system is to be distinguished from what I called the *ultrasystem*, which results from eliminating in the evaluation system everything that is false leaving only truths. To avoid the Gettier problem, I proposed the following constraint: a target belief is an instance of knowledge only if the subject's ultrasystem is sufficient for the target belief's justification. So knowledge results when justification is sustained on the basis of the ultrasystem, that is, when the original justification based on coherence with the evaluation system is undefeated by errors in the evaluation system.

To put it in the first person, a belief of mine enjoys personal justification if my evaluation system enables me to meet all objections, including skeptical hypotheses to the effect that I am unreliable, untrustworthy, or deceived in general or in the particular instance. And a belief of mine qualifies as knowledge if my ability to meet objections to it is sustained in the ultrasystem. Coherentism, thus construed, makes the external conditions of truth, reliability, and counterfactual dependence a function of internal coherence relations within the subject's evaluation system.

In the second edition of *The Theory of Knowledge*, I expressed the requirement of being undefeated by error by saying that a belief's justification must be irrefutable by error in the evaluation system of the subject. Satisfaction of this condition ensures that there is the right match between internal justification and external conditions to convert a true belief to knowledge. External conditions necessary for knowledge appear internally in the evaluation system as answers to objections – for example, objections affirming that the external conditions fail to hold. So the idea is that internal coherence with the evaluation system that matches systematic truth

in the ultrasystem is sufficient for knowledge. Whether, in the end, one thinks of coherence as justification resulting from internal relations within the evaluation system or as the combination of that justification with the requisite systematic truth of the ultrasystem seems to me a merely terminological matter. Internal coherence plus the truth required to render it undefeated and irrefutable is what is needed for knowledge.

It is uncontroversial that knowledge is incompatible with luck. Some might object that the condition I take to be sufficient for knowledge – the target belief's coherence with the subject's ultrasystem – is a matter of luck. However, the appearance of luck is an illusion. The crux of the issue comprises two essential points: first, the subject must be trustworthy in what she accepts, in general and in the particular instance; second, the subject's trustworthiness must be successfully truth-connected in general and in the particular instance. Notice, however, that the subject's acceptance of her trustworthiness and the successful truth-connectedness of her trustworthiness is explained, and indeed justified, by the existence of her successfully truth-connected trustworthiness. The external success is internally justified just as the internal justification is externally explained. The internal and the external are tied together in a loop of justification and explanation that is not mere luck.

In my later work, especially in *Self-Trust: A Study of Reason, Knowledge and Autonomy*, I have insisted on a distinction between belief and acceptance. I proposed to call the positive evaluation of belief *acceptance* and affirmed that acceptance is at the heart of knowledge. Evaluation takes us up to a metalevel. The cognitive mind is, therefore, a metamind. However, as we have noted in discussing trustworthiness of acceptance, there is a loop in the account of trustworthiness. There is a loop of self-trust from mind to metamind and back again tying them up, down and together in a knot of knowledge. It is a knot that ties together science, art, and autonomy, holding them fast against the pounding wave of skepticism. The strength is in the structure of the circle.



## BIBLIOGRAPHY

- Lehrer, Keith: *Knowledge* (Oxford: Clarendon Press, 1974).
- Lehrer, Keith: "Why Not Skepticism?," *Philosophical Forum* 2, 283–98.
- Lehrer, Keith: *Theory of Knowledge* (Boulder, Colo.: Westview Press and London: Routledge, 1990).
- Lehrer, Keith: *Theory of Knowledge*, 2nd edn (Boulder, Colo.: Westview Press, 2000).
- Lehrer, Keith: *Self-Trust: A Study of Reason, Knowledge and Autonomy* (Oxford: Clarendon Press, 1997).
- Lehrer, Keith and Paxson Jr, T.: "Knowledge: Undeclared Justified True Belief," *Journal of Philosophy* 66 (1969), 225–37.
- Lehrer, Keith and Wagner, C.: *Rational Consensus in Science and Society* (Dordrecht: Reidel, 1981).

## Alvin Plantinga

As everyone knows, current epistemology displays a deep cleft between those who embrace internalism and those who embrace externalism. It isn't entirely easy to make a precise (and accurate) distinction between the two, and this isn't the place to try to do so.<sup>1</sup> Suffice it to say that among paradigm internalists are Richard Fumerton, Richard Feldman, and Laurence Bonjour; among paradigm externalists are Michael Bergmann, Alvin Goldman, and Ernest Sosa. Now, internalists often accuse externalists of refusing to take skepticism seriously, or perhaps of having no answer to the skeptic, or perhaps of replying in a way that is wholly unsatisfying from a philosophical point of view.<sup>2</sup> The skeptic declares: "You don't really know that you are not a brain in a vat." The externalist replies: "Yes, I do; my belief that I am not a brain in a vat has been reliably formed (or produced by properly functioning cognitive faculties, or produced in an epistemically virtuous way)." The skeptic asks: "Well, how do you know it's been reliably formed (or properly produced, or virtuously formed)?" The answer: "The belief that it's been reliably formed was itself reliably formed (properly produced, virtuously produced)." "How do you know *that*?" "That belief, too. . . ." Not a pretty picture, and one that will not sit well with the internalist. Regarding these answers

as miserable evasions, she will be wholly unmollified.

On the other side, the requirements internalists impose on knowledge often seem to externalists to be a bit on the demanding side, to put it mildly. For example, there is the suggestion that I know *p* only if I am able to give a good argument for the conclusion that my cognitive faculties are reliable, without relying on those faculties in giving the argument. But not even God himself, essentially and indeed necessarily omniscient as he is, can meet that standard. With requirements like that, the internalist may as well throw in the skeptical towel right at the outset; no defense against skepticism is possible.

Now, I'm strongly attracted to externalism. Furthermore, I have never seen much of a need to reply to the skeptic: why (apart from residual attachment to the project of classical foundationalism) should we think that the or a main concern of epistemology is replying to the skeptic? True, I may not be able to prove to the skeptic's satisfaction that I know anything (my name, where I live, whether I am married, etc.), but why should I have to do that? The skeptic tells me I don't have any knowledge; it's up to him, then, to give an argument for that conclusion; but the arguments actually produced are always, as far as I can see, at best questionable.

Externalism goes nicely with another thought to which I am strongly attracted: the thought that there is a deep connection between metaphysics (more exactly philosophical anthropology) and epistemology. Thus, I've argued that the theist ought to suppose, *prima facie*, anyway, that belief in God can and often does meet the conditions for knowledge. The atheist will sensibly demur; she will suppose that theism arises out of an illusion of some sort, and involves cognitive malfunction (perhaps she will follow Marx with his proposal that religion is the opiate of the people). Alternatively she may think theism arises out of cognitive proper function, but proper function of processes not directed towards the production of true belief (perhaps she will follow Freud and his thought that theistic belief is produced by

wish-fulfillment). More lately, however, I've come to believe that there is a much broader and much deeper connection here, a connection that crucially involves skepticism: I believe that the ontological naturalist, the person who thinks there is no such person as God or anything like God, is committed to a deep and pervasive skepticism. I've made this argument in several places;<sup>3</sup> here I want to make it again in briefer compass and with a revised argument for the crucial premise.

Note first that many thinkers going back at least to Nietzsche and possibly William Whewell have pointed to a potentially worrisome implication of orthodox Darwinian evolutionary theory. The worry can be put as follows. According to orthodox Darwinism, the process of evolution is driven mainly by two mechanisms: random genetic mutation, and natural selection. The former is the chief source of genetic variability; by virtue of the latter, a mutation resulting in a heritable, fitness-enhancing trait is likely to spread through a population and be preserved as part of the genome. It is fitness-enhancing behavior and traits that get rewarded by natural selection; what get penalized are maladaptive traits and behaviors. In crafting our cognitive faculties, natural selection will favor cognitive faculties and processes that result in adaptive behavior; it cares not a whit about true belief (as such) or about cognitive faculties that reliably give rise to true belief. As evolutionary psychologist Donald Sloan Wilson puts it, "the well-adapted mind is ultimately an organ of survival and reproduction" (Wilson, 2002: 228). What our minds are *for* (if anything) is not the production of true beliefs, but the production of adaptive behavior.

But then the fact that our species has survived and evolved guarantees, at most, that our behavior is adaptive (to the ancestral environment); it does not guarantee or even make probable that our belief-producing processes are for the most part reliable, or that our beliefs are for the most part true. That is because our behavior could perfectly well be adaptive, but our beliefs false as often as true. Darwin himself apparently worried about this question: "With me," says Darwin,

the horrid doubt always arises whether the convictions of man's mind, which has been developed from the mind of the lower animals, are of any value or at all trustworthy. Would any one trust in the convictions of a monkey's mind, if there are any convictions in such a mind?<sup>4</sup>

Call this "Darwin's Doubt". We can briefly state the doubt as follows. Let *R* be the proposition that our cognitive faculties are reliable, *N* the proposition that naturalism is true, and *E* the proposition that we and our cognitive faculties have come to be by way of the processes to which contemporary evolutionary theory points us: what is the conditional probability of *R* on *N*&*E*? i.e. what is  $P(R/N\&E)$ ? Darwin fears it may be rather low.

Of course it is only *unguided* natural selection that prompts the worry. If natural selection were guided and orchestrated by the God of theism, for example, the worry would disappear; God would presumably use the whole process to create creatures of the sort he wanted, creatures in his own image, creatures with (overall, and with various qualifications and caveats) reliable cognitive faculties. So it is unguided evolution, and metaphysical beliefs that entail unguided evolution, that prompts this worry about the reliability of our cognitive faculties. Now, naturalism entails that evolution, if it occurs, is indeed unguided. But, then, so the suggestion goes, it is unlikely that our cognitive faculties are reliable, given the conjunction of *N* and *E*. If so, however, one who believes that conjunction will have a defeater for the proposition that our faculties are reliable – and, if that's true, she will also have a defeater for any belief produced by her cognitive faculties – all of them, naturally enough; she is therefore committed to a deep and pervasive skepticism. We can state the argument schematically as follows:

- (1)  $P(R/N\&E)$  is low.
- (2) Anyone who accepts *N*&*E* and sees that (1) is true has a defeater for *R*.
- (3) Anyone who has a defeater for *R* has a defeater for any other belief she holds.

Therefore

- (4) Anyone who accepts N&E and sees that (1) is true is committed to skepticism.

Each of these premises, naturally enough, warrants a good deal of discussion, discussion I don't have space for here. I'd like instead to give an argument for (1), an argument somewhat different from (and, I hope, superior to) the arguments I've given for it elsewhere. Then I'll give brief and abbreviated versions of the arguments for premises (2) and (3) (directing the reader to the fuller versions to be found elsewhere).

This argument for (1) goes as follows. First, in order to avoid undue influence from the natural assumption that our own cognitive faculties are reliable, think not about us but about hypothetical creatures a lot like us (they hold beliefs, make inferences, change beliefs, etc.), perhaps existing in some other part of the universe; and suppose N and E are true with respect to them. Next, note that all or most naturalists are materialists (about human beings); they hold that human beings are material objects, and neither are nor contain as parts an immaterial soul or mind or self. So take naturalism to include materialism.<sup>5</sup> Now, what would a belief *be*, from this point of view? Presumably something like a long-term event or structure in the nervous system – perhaps a structured group of neurons connected and related in certain ways. Such a neural structure will have *neurophysiological* properties ("NP properties"): properties specifying the number of neurons involved, the way in which those neurons are connected with each other and with other structures (with muscles, glands, sense organs, other neuronal events, etc.), the average rate and intensity of neuronal firing in various parts of this structure, and the ways in which these rates of fire change over time and in response to input from other areas. If this event is really a *belief*, however, then it will also have *content*; it will be the belief that *p*, for some proposition *p* – perhaps the proposition *naturalism is all the rage these days*.

What is the relation between the NP properties, on the one hand, and content proper-

ties – such properties as *having the proposition that naturalism is all the rage these days as content* – on the other? Here nearly all materialists would invoke one or the other of two possibilities: reductive materialism (RM) or non-reductive materialism (NRM). Perhaps NRM is the more popular position here. According to NRM, content properties are distinct from but *supervene on* (see the entry on supervenience in the *Stanford Online Encyclopedia of Philosophy*) NP properties;<sup>6</sup> a belief (a neural structure of that sort) will have NP properties, but will also have a content property, which supervenes on but is not identical with or reducible to NP properties. Supervenience can be either broadly logical or nomic. In the latter case, there would be psychophysical laws relating NP properties to content properties: laws of the sort *any structure with such and such NP properties will have such and such content*. These laws will presumably be contingent (in the broadly logical or metaphysical sense). In the former case, there will also be such laws, but they will be necessary rather than contingent.

Suppose we start with the possibility according to which the supervenience in question is causal or nomic. Take any belief *B* you like on the part of a member of that hypothetical population: what is the probability that *B* is true, given N&E and non-reductive materialism – what is  $P(B/N\&E\&NRM)$ ? What we know is that *B* has as content a certain proposition *q*, and (we may assume or concede) having *B* is adaptive in the circumstances in which that creature finds itself. What, then, is the probability that *q*, the content of *B*, is *true*? Well, what is the probability that the relevant psychophysical law *L* connecting NP properties and content properties yields a *true* proposition as content in this instance? Having *B* is adaptive, in the circumstances in which the creature finds itself; its displaying the NP properties on which *C* supervenes causes adaptive behavior. But why think the content connected with those NP properties by *L* will be true in this creature's circumstances? What counts for adaptivity are the NP properties and the behavior they cause; it doesn't matter what that supervening content is, or whether it is true. The NP properties

are indeed adaptive; but that provides no reason, so far, for thinking the supervening content is true. Having B is adaptive by virtue of its causing adaptive behavior, not by virtue of having true content. Of course, if theism is true, then human beings (as opposed to those hypothetical creatures, for whom naturalism is true) are made in the divine image, which includes the capacity for knowledge; so God would presumably have chosen the psychophysical laws in such a way that, in the relevant circumstances, the neurophysiology yields true content. But nothing like that is true given naturalism; to suppose that the content properties that are adaptive, for the most part also lead to true content, would be touching, but naïve and wholly unjustified.

So what is  $P(B/N\&E\&NRM)$ ? Well, since the truth of  $q$  needn't make a difference to the adaptivity of B, B could indeed be true, but is equally likely to be false; we'd have to estimate the probability that it is true as about the same as the probability that it is false. But that means that it is improbable that the believer in question has reliable cognitive faculties, i.e. faculties that produce a sufficient preponderance of true over false beliefs. For if the believer in question has a thousand independent beliefs, each as likely to be false as true, the probability that, say, three-quarters of them are true (and this would be a modest requirement for reliability) will be very low – less than  $10^{-58}$ . So  $P(B/N\&E\&NRM)$  specified to these creatures will be low.

That's how things go for NRM, given that the supervenience in question is causal or nomic. But a little thought shows that the same holds if the supervenience of belief content on NP properties is by virtue of metaphysical or broadly logical necessity. In this case, as in the former case, there will be psychophysical laws connecting NP properties with content properties; but in this case those laws will be necessarily true. Still, that gives us not the slightest reason to think the content assigned a given belief structure B by the laws in question will be a true proposition. We can see no connection at all between any given NP property or conjuncts of NP properties and true content. Consider, once more, any

given belief on the part of one of those hypothetical creatures: what is the probability that it is true, given N&E and NRM, the proposition that content supervenes logically on NP properties? The NP properties, we may once more suppose, are adaptive; but that gives us not even a hint of a reason for thinking the proposition constituting the content of that structure is true. In this case, as in the former, the probability of any given belief's being true would have to be estimated as about .5; and that means that the probability of these creatures' having reliable cognitive faculties is low.  $P(B/N\&E\&NRM)$ , specified to those creatures, is low. But, of course, the same would hold for us, if naturalism is true:  $P(B/N\&E\&NRM)$  specified to us is equally low.

That's how things go for NRM. What about RM, reductive materialism? For present purposes take RM to be the proposition that content properties *just are* (no doubt complex) NP properties; the property of having as content *Naturalism is all the rage these days* just is some complicated NP property of a belief, which is itself a certain kind of neurological structure. Many, following Leibniz, will think it impossible that this should be so; after all, couldn't a person grasp, apprehend, be acquainted with the property of having *Naturalism is all the rage these days* as content without, for any NP property P, having any grasp at all of P? But let's concede, for present purposes, that the idea is coherent. A belief B will be a neuronal structure of some kind. B will cause behavior and, as we can once again suppose, B will cause adaptive behavior. B will have NP properties. Among them will be a certain property P which is both an NP property and a content property: the property of having some proposition  $q$  – for definiteness, the proposition that naturalism is all the rage these days – as its content. P is the property of having  $q$  as content; it is also an NP property. So P will itself have at least two properties of present interest. On the one hand, it will have a neurophysiological side: it will have (for example) the property of being such that any structure that exemplifies it has  $n$  neurons, which are arranged in way  $w$ , and cause behavior  $b$  in circumstances  $c$ . Let "P" name this property. On the

other hand, P will have a content side: it will be such that any structure that exemplifies it, has q as its content. Call this property P\*. Although P\* is a second-order property, a property of properties, it can be thought of as a content property; and, although P is also a second-order property, it can be thought of as an NP property. P and P\* are presumably equivalent in the broadly logical sense, but they are clearly distinct properties.

Now, I argued above that, even if content properties supervene on NP properties, we can see no connection at all between a given NP property and any particular content property. But something similar goes here. If content properties just are NP properties, then for any second-order content property P\* there is a second-order NP property P that entails and is entailed by it; but we can see no connection at all between any given P and any given P\*. So once more consider any belief B on the part of one of those hypothetical creatures. B will have some proposition q as its content; and the property of having q as content will be identical with some NP property P. That NP property P will itself have a content property P\* and an NP property P. But once more there will be no connection we can see between P and P\*. Having P (and hence P\*), we may assume, is adaptive; but that gives us no reason for supposing that P\* confers true content on P and hence on that belief B. Nor can we see any other reason for thinking that the content conferred on P by P\* is a true proposition. But that means that in this case, too, we must estimate the probability that B is true as in the neighborhood of .5. And *that* means that P(R/N&E&RM), the probability of their faculties being reliable on naturalism, evolution and reductive materialism, will be low. As in the previous case, furthermore, the same will be for us: P(R/N&E&RM), specified not to them but to us, will also be low. So both P(R/N&E&NRM) and P(R/N&E&RM) are low; but then P(R/N&E) is low.

That's the argument for the first premise. As for premises (2) and (3), I'll briefly sketch the arguments for them, referring the reader elsewhere for fuller statements.<sup>7</sup> According to the second premise, one who sees that (1) is true and also accepts N&E has a *defeater* for

R, a reason to give it up, to cease believing it. The support offered for this premise is by way of analogy from clear cases. Suppose I believe there is a drug – call it XX – that destroys cognitive reliability; I believe that 95 percent of those who ingest XX become cognitively unreliable. Suppose further that I now believe both that I've ingested XX and that P(R/I've ingested XX) is low; taken together, these two beliefs give me a defeater for my initial belief or assumption that my cognitive faculties are reliable. Furthermore, I can't appeal to any of my *other* beliefs to show or argue that my cognitive faculties are still reliable; any such other belief is also now suspect or compromised, just as R is. Any such other belief B is a product of my cognitive faculties; but, then, in recognizing this and having a defeater for R, I also have a defeater for B. Of course, there will be many other examples: I'll get the same result if I believe that I am a brain in a vat and that P(R/I'm a brain in a vat) is low; the same goes for the classic Cartesian version of the same idea (namely that I've been created by a being who delights in deception) and for other more homely scenarios – for example, the belief that I've gone insane (by way of contracting mad cow disease?). In all of these cases I get a defeater for R.

Now, according to the third premise, one who has a defeater for R has a defeater for any belief she takes to be a product of her cognitive faculties – which is, of course, *all* of her beliefs. She therefore has a defeater for N&E itself; so one who *accepts* N&E (and sees that P(R/N&E) is low) has a *defeater* for N&E, a reason to doubt or reject or be agnostic with respect to it. Nor could she get independent evidence for R; the process of doing so would of course *presuppose* that her faculties are reliable. She'd be relying on the accuracy of her faculties in believing that the alleged evidence is in fact present and that it is in fact evidence for R. Thomas Reid put it like this:

If a man's honesty were called into question, it would be ridiculous to refer to the man's own word, whether he be honest or not. The same absurdity there is in attempting to prove, by any kind of reasoning, probable or demonstrative, that our reason is not fallacious, since the very



point in question is, whether reasoning may be trusted.<sup>8</sup>

The argument concludes that the conjunction of naturalism with the theory of evolution cannot rationally be accepted – at any rate by someone who is apprised of this argument and sees the connection between N&E and R. But, of course, the premises of the argument also imply that anyone who accepts N&E – indeed, anyone who accepts the proposition that we and our cognitive faculties have come to be by way of unguided evolution (evolution unguided or orchestrated or directed by God or any other person) is committed to a deep and pervasive skepticism. For any such person is committed to the thought that she has a defeater for all of her beliefs – including, of course, that one.

## NOTES

- 1 The best account I know of is to be found in Bergmann, 2007.
- 2 See, e.g., Fumerton, 1995, ch. 6.
- 3 See, e.g., Plantinga, 2000 and 2008.
- 4 Darwin, 1881.
- 5 If you feel inclined to complain that naturalism doesn't entail materialism, then take my argument as directed against the conjunction of naturalism and evolution with materialism.
- 6 Or, to accommodate content externalism ("meaning ain't in the head"), on NP properties together with certain environmental properties. This qualification will be presupposed but not mentioned in what follows.
- 7 See Plantinga, 2008.
- 8 Reid, 1983, p. 276.

## BIBLIOGRAPHY

- Bergman, Michael: "Externalist Responses to Skepticism," in *The Oxford Handbook on Skepticism* (Oxford: Oxford University Press, 2007).
- Darwin, Charles: "Letter to William Graham, Down, July 3<sup>rd</sup> 1881" (1881), in *The Life and Letters of Charles Darwin Including an Autobiographical Chapter*, ed. Francis Darwin (London: John Murray, 1887), vol. 1, pp. 315–16.
- Fumerton, Richard: *Metaepistemology and Skepticism* (Lanham, MD: Rowman and Littlefield, 1995).

Plantinga, Alvin: *Warranted Christian Belief* (New York: Oxford University Press, 2000).

Plantinga, Alvin: "Opening Statement," in Alvin Plantinga and Michael Tooley, *Knowledge of God* (Oxford: Blackwell, 2008).

Reid, Thomas: *Essays on the Intellectual Powers of Man*, in *Thomas Reid's Inquiry and Essays*, ed. Ronald Beanblossom and Keith Lehrer (Indianapolis, IN: Hackett, 1983).

Wilson, Donald Sloan: *Darwin's Cathedral: Evolution, Religion and the Nature of Society* (Chicago, IL: University of Chicago Press, 2002).

## John Pollock

### 1. AN ARCHITECTURE FOR RATIONAL COGNITION

#### 1.1 Setting Aside the Gettier Problem

Since Gettier, much of epistemology has focused on analyzing "S knows that P", but that is not my interest. My general interest is in rational cognition – both in what it is to be rational, and in how rational cognition works. The traditional epistemological question, "How do you know?", can be taken as addressing part of the more general problem of producing a theory of rational cognition. It is about specifically epistemic rationality. I interpret this question literally, as a question about *how* we should proceed in our epistemic endeavors. Epistemological theories that try to answer this question are theories of *procedural epistemology* (see my 1998), and when, from this perspective, we assess beliefs in terms of their justifiedness, the concept of justification is one of *procedural epistemic justification*. Whether this has anything to do with the analysis of knowledge is an open question, and not one that I have much interest in addressing.

#### 1.2 Interest-driven Epistemic Cognition

I think it is helpful to approach epistemology from the design stance, and ask what role rational epistemic cognition has in the broader cognitive architecture of a cognitive agent. We can make a rough division of rational cognition into epistemic cognition, which is about what to believe, and practical

cognition, which is about what to do. Epistemologists normally assume that we can study epistemic cognition without paying any attention to practical cognition. But approaching epistemology from the broader perspective of designing a rational agent quickly gives the lie to this assumption. The main point of a system of cognition is to direct an agent's interaction with the world. So its main purpose is to direct action. The intelligent direction of action requires the agent to have information both about its environment and about itself. It is the function of epistemic cognition to provide that information. Viewed in this way, epistemic cognition is subservient to practical cognition. Its role is to provide the information needed for rational decision-making.

It might seem that perception constitutes an exception to the rule that epistemic cognition is driven by practical cognition. Philosophers are sometimes attracted by the simplistic view that in perception we just take in information as it is presented to our senses. However, this is readily shown to be false. Consider vision. Visual perception produces a very rich visual image, and it requires attention to retrieve information from it for further processing.<sup>1</sup> You must attend to the information before you can form beliefs about it. Some mechanisms for attention are automatic. For example, sudden motions in an otherwise still scene, or flashing lights, grab your attention automatically. But many mechanisms of attention are interest-driven.

### 1.3 Empirical Investigation

That epistemic cognition is interest-driven is the simplest way in which it is influenced by practical cognition. Another fairly obvious connection that has none the less been ignored in most epistemological investigations is that epistemic queries cannot usually be answered simply by reasoning from information the agent already has. The acquisition of new information sought for practical purposes typically involves some degree of "empirical investigation", ranging from simply directing your attention, to redirecting your eyes, to looking up information in a book or

online, to engaging in scientific experiments. These are all actions that you perform, and as such are driven by practical cognition. You often have to engage in difficult problem-solving (practical cognition) in deciding how to pursue a desired piece of information.

### 1.4 Reflexive Cognition

These interactions between practical cognition and epistemic cognition are all fairly obvious, and the traditional epistemologist may retort that, of course, there are interactions, but his interest is in the purely epistemic cognition that transpires once practical cognition has posed its queries. However, this is still a simplistic view of epistemic cognition. The difficulty is that human beings are *reflexive cognizers*. We do not just engage in cognition about the world. We also engage in cognition about cognition, and that often gives us the power to redirect the course of our own cognition.<sup>2</sup>

Typically, we are faced with more cognitive tasks than we can immediately undertake, so we need some way of prioritizing them. We can regard these prioritized tasks as stored in a "cognitive task queue", and retrieved in order of priority. There has to be a way of prioritizing them automatically, without thinking about it, because otherwise we would be led into an infinite regress. Plausibly, our default ordering of problem-solving tasks would order the more important ones before the less important ones.

We can decide to refrain from accepting the conclusion of some bit of reasoning even when we are firmly convinced that the premises are true and we cannot see anything wrong with the argument. An obvious example of this occurs when we are presented with logical paradoxes and arguments whose conclusions we reject.

Finally, we often accept conclusions on the basis of incomplete argument sketches rather than fully worked out arguments. Consider mathematical theorem-proving. Philosophers often have the fantasy that that produces the most certain of knowledge. In fact, mathematical reasoning is one of the most error-prone forms of reasoning, and professional mathematicians adopt a healthy skepticism towards

new arguments purporting to establish novel conclusions. No one produces fully worked out arguments of the sort we teach students to produce in introductory logic courses. Such arguments would be too long to comprehend – we would lose track of the forest for the trees. Instead, mathematicians produce argument sketches, with the presumption that the details could all be filled in as necessary. Their basis for believing this is probably some sort of reasoning by analogy from previous cases. This is reflexive cognition, because we are reasoning about our reasoning and concluding that we could fill in the details if we had to. It should be emphasized, though, that the use of argument sketches is not the exclusive province of mathematics. Everyone forms beliefs on the basis of argument sketches.

The upshot of the preceding observations is that epistemic cognition and practical cognition are not separable modules. A complete theory of epistemic rationality is inextricably interwoven with a theory of practical rationality. Thus, an architecture for epistemic cognition cannot be evaluated independently of its interactions with practical cognition. They jointly form a cognitive architecture, and what makes the epistemic parts of it good or bad is how they contribute to the functioning of the whole architecture.

### 1.5 A Two-factor Architecture

A very important feature of the human cognitive architecture, and probably an essential feature of any cognitive architecture able to function efficiently in a complex and rapidly changing environment, is that beliefs and decisions need not be the product of explicit reasoning. Suppose I toss an apple to you, and you catch it. How did you do that? You certainly did not do it by measuring distances and velocities and computing parabolic trajectories. Perhaps you *could* have done it that way, but it would have been much too slow and you would not have caught the apple. Instead, humans and most higher animals have a built-in cognitive module whose purpose is rapidly to produce predictions of trajectories. We rely upon that in forming beliefs about where the apple is going to be when

we try to catch it. I call such modules *Q&I modules* (“quick and inflexible modules”, from Pollock, 1989).<sup>3</sup>

I have long emphasized the importance of Q&I modules in epistemic cognition. It seems that most inductive and probabilistic beliefs are produced by Q&I modules. The general problem is that explicit reasoning is slow and computationally difficult. Q&I modules produce beliefs quickly, but they do so by making assumptions that may fail. Giving priority to explicit reasoning over Q&I modules is an essential feature of our cognitive architecture.

Q&I modules are important for epistemic cognition, but they may be even more important for understanding human practical cognition. In recent years I have become more and more a confirmed nativist. I suspect that most human decision-making, particularly in social contexts, is carried out by Q&I modules. Inborn personality traits reflect individual differences in these practical Q&I modules. Again, when we have the relevant information to make informed decisions on the basis of reasonably held beliefs about values and probabilities, rationality dictates that we should override our Q&I modules and engage in some form of decision-theoretic reasoning (see my 2006 for a detailed account of this).

The result is a two-factor cognitive architecture, in which most beliefs and decisions are produced quickly and fairly automatically by Q&I modules. Explicit reasoning sits above this bundle of Q&I modules and (1) attempts to modulate it, overriding conclusions that can be expected to be wrong, and (2) tries to fill in the gaps when our Q&I modules are unable to produce automatic solutions to problems.

### 1.6 Rules of Rationality

Epistemologists often suppose that they are in the business of producing rules for rational cognition. Simple examples of such rules might be “Do not hold a belief for which you have no good reason”, “Accept the conclusion of a good argument”, and more complicated rules would tell us when to accept the conclusions of inductive arguments, how to handle probabilities, when to draw conclusions

on the basis of perceptual input, etc. But the preceding conclusions demonstrate that none of these rules can be right. Reflexive cognition can lead us to refrain from accepting the conclusion of an argument even when we cannot see anything wrong with it, or to accept a conclusion on the basis of an argument sketch, without having a complete argument. Such cognitive behavior is not irrational.

And the same point applies to the more complex principles describing inductive or probabilistic reasoning and perception. These are at best default rules for how to proceed in the absence of reflexive cognition.<sup>4</sup> True rules of rationality must describe how these default rules interact with reflexive cognition, and stating such rules probably requires formulating a complete theory of rational cognition. It is unlikely that there are piecemeal rules that have any standing on their own, independently of being embedded in the larger system.

## 2. HUMAN RATIONALITY

Investigating rationality from the design stance can take either of two directions. We can ask how to build a well-functioning cognitive agent, without caring much whether it works the way humans work. This is one strand of research in artificial intelligence. But we can also ask specifically how human rational cognition works, and employ the design stance in an attempt to understand why it works in the way it does. The design stance imposes important constraints on theories of human rationality generally, and on epistemological theories specifically, because they are theories about *how we work*. If such a theory is to be correct, it must be possible for an agent actually to work that way. Thus, an important test of such a theory is to consider whether, if we built an agent that worked in the way envisioned, it would perform in ways that we regard as rational. Would it accord with our standards of rationality? If not, the theory must be wrong.

Epistemologists have traditionally tried to meet this requirement by engaging in thought experiments run from their armchairs, and

that must inevitably be the first test of an epistemological theory. But we are very limited as to how far we can go in this way. The problem is that, as in all philosophy, the devil is in the details. It is easy to sketch theories that look good in the abstract, but they usually break down because it is impossible to fill in the details in any reasonable way. In epistemology, the only way to surmount this problem is to work out the details. Philosophers are unaccustomed to working this hard, partly because, once the details have been supplied, the theory tends to become too complex to evaluate from the armchair. This is for two reasons: (1) It can be difficult to be sure what consequences a complex theory has when applied to a complicated case: (2) It can be surprisingly difficult to be sure that you have even supplied all the details needed for the theory to have any implications at all. Both of these problems are familiar to computer programmers. No matter how skillful the programmer, a complex computer program *never ever* works properly the first time. All complex programs are initially buggy, and the only way to find the bugs is to run the program on test cases and see what it does. Epistemological theories are much like computer programs. They are theories about *how something works* – namely, rational cognition – and the only way to get such a theory right is to run it on complex cases. If the theory is complex and the cases are complicated, we cannot do this by just sitting in our armchairs and thinking. As far as I can see, the only way to do it is actually to build an agent that works in the way described by the theory and see what it does. In other words, we must test theories of rational cognition by building AI systems that model them. This conviction gave rise to the OSCAR project in 1985.<sup>5</sup> The objective of the OSCAR project is to construct a general theory of rationality and test it by implementing it as an AI system. We can then apply the working system to complex scenarios and see what it does. As with any computer program, I can assure you that no implemented theory of rational cognition will do what we expect it to do the first time around. Assuming that the program faithfully captures the theory, designing and refining the

theory must proceed as programming always does – by systematic testing and debugging. In this case it is the theory itself that we are debugging.

In epistemology, however, there is another twist to the problem. In the hard sciences, we test our theories by looking at the world and seeing whether it behaves in the manner portrayed by the theory. In philosophy, the criterion for correctness of a theory of rational cognition is that a system performing as described will actually behave rationally. But, to test a theory in this way, we must be able to tell whether a concrete bit of behavior does accord with our standards of rationality. How do we do that? Epistemologists have traditionally tested their theories by appealing to their “philosophical intuitions”. But what are these, and why should we trust them?

Although theories of privileged access have fallen into disrepute in the philosophy of mind, I claim that we do have a kind of privileged access to the rationality of our judgments. This reflects an important feature of our cognitive architecture. I remarked above that we often form beliefs on the basis of argument sketches rather than complete arguments. But, for this to work satisfactorily, we must be able to criticize argument sketches on the grounds that there is no way to fill them out into complete arguments, or alternatively we must be able to confirm an argument sketch by showing that it can be filled out. To do that, we must be able to inspect candidate expansions of argument sketches and evaluate them as good or bad arguments. But that just amounts to judging whether, if we reasoned in that way, we would be conforming to the dictates of rationality. Thus, an essential feature of rational cognition must be the built-in ability to judge whether particular bits of cognitive behavior conform to the dictates of rationality. This kind of self-monitoring of cognition is perfectly analogous to our similar ability to monitor physical behavior and judge whether we are performing our actions in conformance with our procedural knowledge for how to do whatever it is that we are doing.<sup>6</sup> This is a built-in feature of our cognitive architecture, and it is this that underlies our so-called “philosophical intuitions”, at

least regarding rational cognition, and in particular regarding epistemic cognition.

The preceding explains how we can judge the rationality of particular bits of cognitive behavior, but what is it *to be rational*? By virtue of what are the built-in standards to which our philosophical intuitions appeal correct? I doubt that this has an answer, at least in the form of a logical analysis of rationality. Judging rationality is a built-in feature of reflexive cognition, and thus just one more aspect of our cognitive architecture. The ability to make these judgments is included in the architecture because it makes the complete architecture work better. In saying this, we are judging the architecture from an external perspective, and there are numerous external perspectives from which we might make such an evaluation: e.g., we might adopt an evolutionary perspective and ask how well the architecture contributes to the propagation of the human genome. This is a different kind of judgment from our internal judgments of rationality. They are part and parcel of the system itself, and proceed as they do because that is the way we are built. To be rational just is to conform to our reflective judgments of rationality. I doubt that there is more to be said on the matter. From an external perspective, we may be able to show why a system whose rational standards satisfy some broad general principles will tend to work better than one that does not. For instance, I think it can be argued convincingly that any sophisticated cognitive agent needs a system of defeasible reasoning.<sup>7</sup> But there may be many different systems of reasoning that will work equally well from an external perspective. Only one of these will be endorsed by the agent’s internal judgments of rationality, but which one that is may be largely an accident.

### 3. HUMAN IRRATIONALITY

Epistemologists sometimes regard their task as that of discovering (or discovering how to discover) rules for avoiding irrationality. But they rarely stop to consider a deeper question. Why is it possible for human beings to be irrational? If evolution has deemed it



desirable for us to behave according to certain standards of rationality, why didn't it just build us so that we work that way? For example, consider my artificial agent OSCAR. OSCAR is able to engage in some quite sophisticated cognition, but (unless it is broken) OSCAR cannot be irrational. OSCAR is built to work in accordance with rules motivated by studies of human rationality, but OSCAR cannot violate those rules. That is just the way OSCAR works. Humans, on the other hand, can behave irrationally without being broken. Why is this possible?

What makes human irrationality possible is that humans are reflexive cognizers. Unlike OSCAR, we can reason about our own cognition and decide to redirect its course in various ways. When we do this we are engaging in practical (decision-theoretic) reasoning about how to proceed. Although OSCAR can do this in principle, at its current stage of development OSCAR performs no reflexive cognition. It is this ability to deviate from our default rules of cognition that enables us to behave irrationally. For example, I noted that we can decide what to think about. This makes us more effective problem-solvers, but it can also lead to irrational behavior. For instance, every researcher has had the experience of having occur to him a possible difficulty for a favored theory, and felt the temptation to ignore it, i.e. to reorder his cognitive task queue so that he never thinks about it again. That would be irrational, and what makes it possible is reflexive cognition.

In my (2007a), I surveyed cases of human irrationality, and argued that they can all be traced to a single source. I remarked that a large proportion of our judgments are the result of applying Q&I modules. However, rationality dictates that when we have reason for being suspicious of the outcome of a Q&I module we should override it by reasoning explicitly about the matters at hand. Unfortunately, we often have difficulty overriding Q&I modules. I am unsure whether this difficulty manifests itself in epistemic cognition, but it obviously does in practical cognition. Explicit decision-theoretic reasoning is difficult for us, because we often lack knowledge of the requisite probabilities and utilities. We get

around this by employing a wide range of Q&I modules for practical decision-making. For example, we take a desire to do something to be a reason for doing it. But sometimes fulfilling a desire can have long-term negative effects. Consider the conditioned desire to smoke cigarettes. On the basis of that desire, many people smoke. Most of them know that smoking is bad for them, and that they should not smoke, but they do it anyway. In other words, their explicit decision-theoretic reasoning does not have the power to override their conditioned desire. I discussed these matters at length in my (2006). In so far as a person fails to override a Q&I module by explicit reasoning leading to contrary conclusions, he is being irrational. In my (2007a), I suggested that this is the sole source of human irrationality.

#### 4. THE ROLE OF RATIONALITY IN A THEORY OF COGNITION

We can distinguish between two kinds of cognition. Much of our cognition, like the construction of the visual image on the basis of perceptual input, proceeds automatically, and we cannot voluntarily alter its course (except indirectly, e.g. by closing our eyes). But some aspects of our cognition are "voluntary", in the sense that we can engage in reflexive reasoning about them and decide whether to follow our default rules, or do something else. Rationality only pertains to the latter. The production of the visual image can be erroneous, in the sense of misrepresenting our surroundings, but it cannot be irrational. But both epistemic and practical cognition can be overridden by reflexive cognition, and as such they can be irrational.

The traditional view of philosophical theories of practical and epistemic cognition is that they are normative. They are about how we *ought to cognize*, and as such they are orthogonal to psychological theories of cognition, which are about how we *do cognize*. But I think this view is wrong. I believe that both philosophical and psychological theories of rationality are empirical theories about the contingent structure of our cognitive

architecture. I argued that philosophical theories are based on our philosophical intuitions about how we ought to cognize, but these in turn are a contingent aspect of our cognitive architecture. The theories of rational cognition that we construct on the basis of these philosophical intuitions are theories about the structure of our cognitive architecture. In particular, they are often theories about the default rules that we follow when we are not engaging in any reflexive cognition. However, reflexive cognition itself is not different in kind from other cognition. It is just a matter of turning our general-purpose reasoning procedures on a different subject – itself.

The conclusions we draw in this way about the structure of our cognitive architecture could in principle be discovered by the psychologist using non-philosophical methods. We could arrive at the same theories in either way. However, philosophical theorizing trades upon the fact that our cognitive architecture affords us privileged access to certain features of our cognitive architecture, via our philosophical intuitions. That it does this is just one more contingent feature of that architecture. This privileged access does not apprise us directly of how our cognitive architecture works. Rather, it gives us particular instances of rational or irrational cognition, and we can take that as our data for constructing general theories of cognition. The construction and confirmation of theories on the basis of this data works the same as theory confirmation does anywhere in science. There is nothing uniquely philosophical about it.<sup>8</sup>

The preceding remarks suggest that philosophical theories of rational cognition are perfectly ordinary theories about certain aspects of human cognition, not essentially different from psychological theories. However, it cannot be denied that there is something normative about judgments of rationality. If I become convinced that I have cognized irrationally, that moves me to try to correct my cognition. Judgments of rationality are normative in the sense that they provide assessments of value that plug into practical cognition and effect our subsequent behavior. We value being rational, and that moves us

to try to achieve it. However, this is just one more aspect of our cognitive architecture. It builds in loops whereby judgments about rationality affect the subsequent operation of the architecture. This is a sense in which philosophical theories are normative, because they issue in judgments of rationality, but all of this is again a contingent aspect of our cognitive architecture, and it is the sort of thing that psychologists must study just as much as philosophers if they are to obtain a complete theory of human cognition.

The upshot is that philosophical theories of rationality are not different in kind from psychological theories of cognition. They focus on a restricted subclass of cognition – voluntary cognition – and they are based on a different methodology. However, the theories produced are about the contingent structure of the human cognitive architecture. The theories have normative import, because our cognitive architecture involves loops whereby judgments of rationality affect the course of cognition. But this would be no less true if the theories were produced by psychologists rather than by philosophers. Ultimately, the study of rational cognition belongs more generally to cognitive science – not just to philosophy or psychology. Their research methods are complimentary rather than in competition. And this makes it easier to acknowledge that traditional philosophical methodology, with its emphasis on armchair thought experiments, is too limited even for philosophy. To get theories of rational cognition right, we must test our theories by engaging in computer modeling.

#### NOTES

- 1 For illustration, see a well-known example by Simons and Chabris at <http://viscog.beckman.uiuc.edu/media/ig.html>.
- 2 See my (2007a) for a more extensive discussion of this. See also my (1995) and Pollock and Cruz (1999).
- 3 See my (2006) for a more extensive discussion of Q&I modules and their role in both epistemic and practical cognition.
- 4 See my (2007a) for a more careful discussion of this.

- 5 For a continually updated report on the status of OSCAR, see <http://oscarhome.soc-sci.arizona.edu/ftp/OSCAR-web-page/oscar.html>.
- 6 This proposal derives originally from my (1986).
- 7 I was one of the first philosophers to write about defeasible reasoning, beginning in my 1965 PhD dissertation, and then in my (1967), (1970), (1971), (1974), (1986). Defeasible reasoning formed the cornerstone of my epistemology in my (1974) and (1986), and its implementation has been one of my main interests in AI – see my (1995) and (2002). See my (2007b) for a sketch of my general theory of defeasible reasoning.
- 8 These observations derive from my (1986) and (2007b).

## BIBLIOGRAPHY

- Cheng, P. W. and Holyoak, K. J.: "Pragmatic Reasoning Schemas," *Cognitive Psychology* 17 (1985), 391–406.
- Pollock, John: "Criteria and Our Knowledge of the Material World," *The Philosophical Review* 76 (1967), 28–60.
- Pollock, John: "The Structure of Epistemic Justification," *American Philosophical Quarterly*, monograph series 4 (1970), 62–78.
- Pollock, John: "Perceptual Knowledge," *Philosophical Review* 80 (1971), 287–319.
- Pollock, John: *Knowledge and Justification* (Princeton, NJ: Princeton University Press, 1974).
- Pollock, John: *Contemporary Theories of Knowledge* (Lanham, MD: Rowman and Littlefield, 1986).
- Pollock, John: *How to Build a Person* (Cambridge, MA: MIT Press/Bradford, 1989).
- Pollock, John: *Cognitive Carpentry* (Cambridge, MA: MIT Press, 1995).
- Pollock, John: "Procedural Epistemology – at the Interface of Philosophy and AI," in *The Blackwell Guide to Epistemology*, ed. John Greco and Ernest Sosa (Oxford: Blackwell, 1998a).
- Pollock, John: "Perceiving and Reasoning about a Changing World", *Computational Intelligence* 14 (1998b), 498–562.
- Pollock, John: "Defeasible Reasoning with Variable Degrees of Justification," *Artificial Intelligence* 133 (2002), 233–82.
- Pollock, John: *Thinking about Acting: Logical Foundations for Rational Decision Making* (New York: Oxford University Press, 2006).
- Pollock, John: "Irrationality and Cognition," in *Epistemology: New Philosophical Essays*, ed. Quentin Smith (New York: Oxford University Press, 2007a).
- Pollock, John: "Defeasible Reasoning," in *Reasoning: Studies of Human Inference and Its Foundations*, ed. Jonathan Adler and Lance Rips (Cambridge: Cambridge University Press, 2007b).
- Pollock, John, and Joseph Cruz: *Contemporary Theories of Knowledge*, 2nd ed (Lanham, MD: Rowman and Littlefield, 1999).
- Simons, D. J. and Chabris, C. F.: "Gorillas in Our Midst: Sustained Inattention Blindness for Dynamic Events," *Perception* 28 (1999), 1059–74.
- Wason, P.: "Reasoning," in *New Horizons in Psychology*, ed. B. Foss (Harmondsworth: Penguin).

## Ernest Sosa

I distinguish between "animal" and "reflective" knowledge and justification, without suggesting that the former is restricted to lower animals and the latter to human beings. An athlete's animal, athletic prowess can be distinguished from his more cerebral strategic excellence.

Animal, unreflective knowledge is largely dependent on cognitive modules and their deliverances. The visual deliverances of someone with 20/20 eyesight will differ in quality from those of someone nearly blind. Reflective knowledge manifests not just modular deliverances blindly accepted, but also the assignment of proper weights to conflicting deliverances, and the balance struck among them.

Reflective knowledge is, again, a *sort* of knowledge, to be listed alongside others, such as perceptual knowledge, or consultative knowledge acquired through social interaction. Among these it does traditionally have a special standing, given its prominence as an intellectual desideratum, and its perennial attraction for the reflective. When we rely on dialogue or seminar discussion, as did Socrates and Plato, we also put a premium on consultative knowledge. Why not give this equal billing? Why the pride of place for reflective knowledge?

One answer is to be found in the special bearing of reflective knowledge on the understanding and coherence dear to intellectuals, and on the intellectual agency that we honor. But we do prize discussion and

interpersonal dialectic as well, so why high-light reflection above consultation? Here is a further reason: No matter how much we value consultation, we are unwilling to yield our intellectual autonomy, which requires us to assess the place of consultation, in any particular instance, in the light of all our other relevant information and recognized desiderata. For to assess it thus is to evaluate it in the light of reflection. Of course, reflection itself might benefit from consultative evaluation. So why put reflection above consultation? Partly, it seems to me, because the deliverances of consultation need assessing in the light of reflection, in a way different from how reflection is assessed through consultation. In the end, reflection deserves a closer, more finally determinative influence on the beliefs we form, and the deliverances of consultation bear properly only through reflection's sifting and balancing. *Conscious* on the spot is not required, however, since a second-order perspective can work subliminally.

There is, moreover, a further way in which reflection is particularly important. It is the ideal of reflective knowledge that best explains the traditional attraction and importance of skepticism. The skeptic sees vicious circularity as an insurmountable obstacle to our pursuit of reflective knowledge. Reflective knowledge requires us to assess our basic epistemic sources as reliable (enough), and this must, of course, be done with rational justification. But how can we attain such justification except by trusting those very sources, which seems inevitably, and viciously, circular. Nothing like this affects consultative knowledge (or justification).

How now can we know our cognitive faculties to be as reliable as we ordinarily take them to be? Would it be by appeal to the foundational inputs delivered by these very faculties? Consider the intuitive, perceptual, introspective, and mnemonic inputs that they deliver. Suppose that, based on such inputs, we form a picture of ourselves and the world around us, and of how the two are systematically related. Could such a picture be used to underwrite the reliability of the very faculties that deliver those inputs?

In accepting those deliverances, do we not already manifest a commitment to the reliability of those very faculties? *Arbitrarily* inputs will yield no knowledge to underwrite the relevant picture. Faculties of perception in particular involve commitment to accept inputs of certain general sorts. Consider such a faculty or subfaculty: vision, say, or color vision, beyond just perception in general. By entailing the delivery of certain sorts of beliefs in certain correlated general conditions, it may bundle built-in implicit commitments to accept certain beliefs based on awareness of certain circumstances. That causes a problem for any foundationalism that would infer a belief in the reliability of a faculty such as perception from the deliverances of that very faculty (among others). After all, a commitment to the reliability of such deliverances is already required for the proper operation of the faculty, since the faculty, as I conceive of it, amounts to or at least requires an associated bundle of commitments, and a commitment to the reliability of such a bundle is as closely associated with the bundle itself as is the belief that a belief is true with the belief itself. So the commitment must be there already and, on pain of vicious circularity, cannot be supposed to gather epistemic status from any such inductive inference. How, then, could such commitments gain their required status? Absent such status, the deliverances of the correlated "faculty" would be worth little, and could not provide inputs inductively yielding conclusions with derived epistemic worth.

One important concept of justification involves evaluation of the subject as someone separable from his contingently given environment. Is it only agents and subjects that are assessed as justified or not? No, actions and beliefs are assessable, too. Indeed, when one is assessed as justified, one is assessed as justified *in acting or believing a certain way*. Nevertheless, the evaluation of particular acts might imply an indirect evaluation of the agent or subject. A tennis shot may count as accurate or not, which will imply only a minimal comment on the shotmaker. That same shot may also be assessed as skillful or not, however, involving some evaluation of the agent, if only indirectly.

One's commitment to a view that *when things look a certain way, one can expect a certain outcome* may be revealed by repeated expectation of the outcome in situation after situation when things in fact do look that way. Our accepting that "When things look F they are likely to turn G" may be manifest not through conscious articulation, but only through a pattern of "inferences". Thus, we come to the question "What might justify commitment to such an inference pattern?". Sundry things could do so: for one thing, the fact that the pattern fits one's experience and is accepted because of that. That's one way to block the regress. Another way is simpler: namely, that we be innately hard-wired for that inference pattern, by God or Nature, so that it might appropriately guide our reasoning. And this it may do even independently of whether the environment jibes with our specific cognitive makeup when we make that inference. The inference is still "justified" in any case, since it is the sort of inference that it is appropriate for us to make downstream from experience, regardless of the aetiology of that experience or the nature of the environment, thus regardless of how "apt" the belief may be relative to the environment within which it is acquired or sustained. So long as in our world it is a normally successful pattern, and it is no accident that we host the commitment to its use (a gift perhaps of God, or Mother Nature) with sensitivity to its validity, we may evaluate ourselves individually as "justified" in the beliefs acquired or sustained by means of that pattern, and evaluate the beliefs (the believings) themselves as justified on that basis.<sup>1</sup>

Having grasped how it is around us through connections involving the perception/memory/reasoning required for animal knowledge, further reasoning on that basis may enhance our conscious integration and explanatory coherence, lending epistemic virtue to our beliefs. There is no more vice in this circular procedure than in a case of visually-apparent-sprinkling/circles-in-puddles/pitter-patter-on-the-window-panes/car-wipers-wiping/umbrellas-up/felt-drops-on-one's-bare-arms/recalled-forecast, etc. There is nothing wrong with accepting various sub-arguments

in such a case concurrently, believing the conclusion of each partly on the basis of the other beliefs used as premises. It *would* of course be bad to hold those sub-arguments concurrently *absent any connection with the relevant externalia*. Even when one holds the lot of them concurrently interlocked, however, *this does not imply that one holds them so detached*. A special, rational form of viciousness spoils attempts to reach conclusions about the contingent world around us through reasoning detached not only from the world beyond but even from those states and beliefs required as peripheral intermediaries for the desired connection with the world. Prominent among these states are the experiential states whose job it is precisely to mediate in that way between our contingent beliefs about the world around us and the world that they are about. Much purely circular reasoning would be defective in just that way.

Here is a kind of Moorean argument:

1. Datum: I know 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 explanation 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 were it not a reliable faculty.
5. So, finally, my perception must be a reliable faculty.

About this one might reasonably wonder how it could possibly help to provide us with knowledge of the conclusion, if the only way we could come to know the initial datum, premise 1, would require that we already trust our faculty of perception. And this in turn would require that we host commitments of perceptual belief formation that, bundled together, if true, *amount* to the reliability of our perceptual faculty.

Given our earlier reflections, we can now distinguish the implicit commitments that, bundled together, constitute our faculties of perception from the much richer perspective



through which one might explain, by appeal to the workings of such faculties of perception, how it is that one knows oneself to see a hand. This richer, explicit perspective is attained through reasoning that does involve a kind of circularity, but it is hard to deplore the circularity as vicious when we compare it with all the conceivable alternatives, namely these:

- (a) Abandoning any attempt to gain such perspectival knowledge of our own cognition, dismissing or ignoring all such questions;
- (b) Attempting but *failing* to attain much of a perspective at all;
- (c) Attempting to gain such a perspective and reaching only one that undermines, rather than endorsing, the faculties by which it was attained, entailing that they are unreliable, and yield mostly error.

By comparison with these outcomes, how can the following be vicious?

- (d) Attempting to gain such a perspective and reaching a rich perspective that endorses the faculties by which it was attained.

To think this vicious is to think that there is something bad or wrong about it, but how can that be if it is at least as good as any conceivable alternative?

We come into the world outfitted with sundry default commitments, our brains wired for cognitive and other success in the environment in which a human is normally born. We can turn to Evolution or to Divine Providence hoping for a further explanation, but even in the absence of any such meta-narrative it is plausible that, innate to us, or soon naturally acquired, are default commitments that serve us well. So long as we exercise these cognitive or epistemic virtues or competences, we do well epistemically, we gain non-accidental access to truths of interest and importance to us. This is not to say that the default justification thereby acquired cannot be defeated. It would be defeated if we gained further, specific information about the

particular situation at hand, in the sorts of ways in which we might gain the information that the lights were indeed red, and hence unsuitable for taking color appearances at face value. Absent such special reason, however, we have epistemic standing for our normal perceptual belief in a normal situation.

How can those basic implicit commitments be justified, if their contents are general, contingent propositions, at least about tendencies or the like? How can such a proposition be justified foundationally or immediately as a “self-evident” universal or probabilistic truth? How can we sensibly allow ourselves justification for believing such a truth “a priori”, absent proper empirical inquiry into our actual contingent surroundings? At this fundamental (and foundational?) level, we proceed with epistemic adroitness, in an epistemically appropriate and desirable way, if we satisfy the distinctively epistemic value of systematic acquisition and retention of truth, and if we do so not by accident, but in a way that derives from our nature and the nature of things, which makes us non-accidentally sensitive precisely to the “validity” of the inferential patterns constitutive of those faculties and their bundled implicit commitments. It is not easy to specify the kinds of accident that will spoil epistemic justification. We are, in any case, a certain way by nature, a way that, given our normal environment, enables us to attain truth and understanding on questions of interest, and with non-accidental responsiveness to the truth, including the conditional truths constitutive of the validity of our inferential patterns. Our inbuilt mechanisms may still operate correctly even if, unfortunately, we are in an abnormal environment relative to which those very mechanisms distance us from both truth and understanding. The mechanisms include taking our sense experience at face value, and coming to know what others think or feel when prompted by the external, behavioral signs of their states of mind.

That is a way in which we can become justified by using our basic faculties, through implicit commitments to the inference patterns that constitute such faculties. Even victims of a Cartesian evil demon would

retain such justification. By reasoning from such implicit commitments we may eventually gain conscious awareness of our faculties and sub-faculties. We may thus gain conscious knowledge of their nature and of how they enable our cognitive success. And this can help provide a perspective that underwrites, with coherent understanding, our use of those faculties. Such conscious awareness of our intellectual makeup may also enable its gradual improvement, as when we are no longer taken in by familiar perceptual illusions. Is that vicious? No more so here than it was for Descartes, or so I have argued.

In previous publications I have defended the following principle:

*Exclusion*

*If one knows full well that p and considers whether one does, and one then (at that very moment) justifiably believes that for one to know full well that p it must also be the case that q, then one must also be justified in believing that q.*

It may be suggested that if we can find a way to rebut this principle, then we can dispense with the requirement of epistemic perspective, and can more easily resist skepticism. I agree, but two problems remain. First, I see no way to reject Exclusion, not at least for knowledge at the reflective level perennially attractive to philosophers, and to the reflective more generally. Second, if we reject this principle, we may succeed too well, making it a mystery how skepticism could have had the attraction that it has had for so long. It is by way of the principle of exclusion that one can best explain the impressive power of skeptical reasoning.

Finally, our principle need not just stand on its own. It gains support from two other principles, that of Ascent and that of Transfer, which I will now state. Aided by plausible auxiliary hypotheses, Ascent and Transfer together entail Exclusion.

*Ascent*

*If one knows full well that p and one considers whether one does, then one must be justified in thinking that one does.*

*Transfer*

*If one is justified in believing X and in believing that for X to be the case Y must also be the case, then one must also be justified in believing Y.*

*Exclusion*

*If one knows full well that p and considers whether one does, and one then justifiably believes that for one to know full well that p it must also be the case that q, then one must also be justified in believing that q.*

I can see no way to rebut this reasoning, whose conclusion, the exclusion principle, supports a skeptical position. While agreeing that, unless we can reject the exclusion principle, skepticism is attractive indeed, I doubt that the principle could ever be rejected plausibly enough, and would myself seek a response to skepticism along another avenue. In defense of the ascent principle, more specifically, I must stress that my use of "full well" in that principle is meant to suggest that the sort of knowledge involved is of a higher level than the animal knowledge requiring only that one's belief track the truth, or be formed reliably, whether or not one appreciates any of this. Ascent therefore postulates a sort of knowledge requiring one's belief to be on a higher epistemic level, to which one is denied access if unable to affirm that one does know, not even when one consciously ponders the question.

I defend this stance by comparing the three pairs of attitudes that open up once one consciously ponders whether one knows in believing that p, provided one retains one's belief that p. The three are: affirmation, denial, and conscious suspension of belief. Can there be any doubt that it is the first of these that is most rationally coherent?

Finally, I propose a distinctive view of epistemic normativity. I deny that the epistemic value of a true, apt belief is exhausted by its contribution to the life of the believer. For such a token belief would retain epistemic justification even if it were part of a life that was overall quite unhappy, and even if its contribution to the value of that life were overall minimal or even negative. Of course,

it may be that – having fundamental epistemic value – such apt, virtuous believings do constitutively contribute to the value of the hosting life, even if instrumentally they are deleterious to its overall worth.

Although I cannot here go into the matter in much depth, I would like to record some skepticism about that idea, limited and modest as it is. The skepticism sets in when I consider the sort of epistemic value that is constitutive of knowledge, the sort that is constitutive of the safety, aptness, and virtue of an instance of believing, and when I ask myself whether that sort of value is the sort of value that in ethics or value theory would be regarded as intrinsic. I must say I rather doubt it. When we evaluate a belief in respect of whether it is a good candidate for knowledge, we consider whether it satisfies certain conditions such that it is better that a belief satisfy those conditions than that it not do so, other things being equal. But from this it does not follow that it is a good thing that the belief exist, whether satisfying the conditions or not, nor does it even follow that it is good that there be such a belief satisfying the conditions. From the fact that it is better that a belief that *p* satisfy the conditions than that it violate the conditions, it does not follow that it is better that it satisfy the conditions than that it fail to exist altogether. And the latter would seem necessary for a conditions-satisfying belief that *p* to have intrinsic value *tout court*, as opposed to simply having fundamental *epistemic*, i.e. epistemic value not wholly constituted by the relations of such a believing to other things with their own epistemic value.<sup>2</sup>

## NOTES

1 Swampman has “human justification” in an extended sense, because of his isomorphism to a normal human.

2 This self-profile gives a partial account of my work in epistemology prior to my Locke Lectures, published as *Apt Belief and Reflective Knowledge*, in which I further develop my form of virtue epistemology and apply it to the problem of skepticism, especially dream skepticism, to the epistemic standing of intuitions, and to the nature of epistemic normativity. For a fuller

profile, see the introduction to *Ernest Sosa and His Critics*, and the entries on me by John Greco in the *Dictionary of Modern American Philosophers* and by Matthias Steup in the second edition of the *Macmillan Encyclopedia of Philosophy*.

## BIBLIOGRAPHY

- Greco, J., ed.: *Ernest Sosa and His Critics* (Oxford: Blackwell, 2004).
- Sosa, E.: “The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge,” in *Midwest Studies in Philosophy* 5, ed. Peter French, Theodore Uehling, Jr. and Howard Wettstein (Minneapolis, MN: University of Minnesota Press, 1980), 3–25.
- Sosa, E.: “Beyond Skepticism, to the Best of Our Knowledge,” *Mind* 97 (1988), 153–89.
- Sosa, E.: *Knowledge in Perspective: Selected Essays in Epistemology* (Cambridge: Cambridge University Press, 1991).
- Sosa, E.: “Philosophical Scepticism and Epistemic Circularity,” *Proceedings of the Aristotelian Society* 68 (1994), 263–90.
- Sosa, E.: “Reflective Knowledge in the Best Circles,” *Journal of Philosophy* 94 (1997), 410–30.
- Sosa, E.: *Apt Belief and Reflective Knowledge* (Oxford: Oxford University Press, 2006).
- Sosa, E. and BonJour, L.: *Epistemic Justification: Internalism vs. Externalism; Foundations vs. Virtues* (Oxford: Blackwell, 2003).

## ENTRIES ON ERNEST SOSA

- Oxford Companion to Philosophy*, ed. Ted Honderich (Oxford: Oxford University Press, 1995).
- Biographical Dictionary of Twentieth-century Philosophers*, ed. S. Brown, D. Collinson and R. Wilkinson (London: Routledge, 1995).
- Dictionary of Modern American Philosophers* (London: Thoemmes Press, 2005).
- Macmillan Encyclopedia of Philosophy*, 2nd edn (London: Macmillan, 2005).

## Barry Stroud

My interests in epistemology were generated at first more by questions about necessary truth and our knowledge of it than by “empirical” knowledge of the world around us. A widely shared conception of perception and of the methods of confirmation thought

to be sufficient to yield knowledge of the world made it look as if necessary truths could be known only in some other “non-empirical” or “a priori” way, and perhaps were not really true of the world at all. Despite disagreements of detail about exactly what can be perceived and exactly how those perceptual data support the conclusions drawn from them, this generally hierarchical or foundational picture of the basic structure of “empirical” knowledge dominated twentieth-century epistemology and gave determinate shape to the philosophical problem of our knowledge of the external world.

The problem was to explain how knowledge of the world is possible in general, not simply how human beings come to know certain things about it on the basis of certain other things they already know about it. The goal was to explain, at least in outline, how anyone ever comes to know anything at all about a public objective world on the basis of perceiving the sorts of things human beings are capable of perceiving. On the prevailing conception, what was “directly” available to unaided perception alone was not part of that “external” world at all. The idea was that we could perceive everything we strictly speaking perceive without knowing anything about any such world, or even without the world’s being anything like what we believe it to be.

This represented a truly formidable challenge. It seemed to me that some proposed solutions that accepted the challenge on its own terms (e.g. phenomenalism) simply did not work, and that others did not strictly fulfill the conditions that any successful solution would have to meet. At one point or another they all surreptitiously ascribed to the human subjects they were studying some knowledge of just the kind they were supposed to be accounting for. Paying close attention to the distinctive character of the completely general question of knowledge, and to the demanding requirements for a fully successful answer to it, made it look more and more as if no satisfactory account could be found. A negative or sceptical answer to the question of knowledge seemed to me the only acceptable outcome. Not only necessary truths but also no truths about the world

at all seemed knowable on that conception of perception and knowledge.

One response to this predicament, if it was recognized at all, was to reject the general epistemological problem as somehow illegitimate. Since a negative, sceptical answer to the question of human knowledge in general is absurd, paradoxical, or even in a sense unthinkable, it was tempting to conclude that there really must be no serious or perhaps even fully intelligible epistemological problem of that form. One strong version of this strategy was the roughly Kantian idea that the very formulation of the question carries within it a guarantee of an unthreatening positive solution to it. Our having any conception of ourselves as knowers of a world at all was thought to be enough to imply the possibility of human knowledge of such a world. This encouraging conclusion could perhaps look reachable on Kant’s own view of the world as “empirically real” but “transcendentally ideal”. But on any more reasonable conception of a world independent of us and our responses it was difficult to explain how a necessary connection could be found, or forged, between our thinking of the world in certain ways and the world’s actually being, and being knowable as being, any of those ways. This was the problem I drew attention to in “Transcendental Arguments” of 1968. Without transcendental or any other form of idealism to appeal to, it looked as if only something like an implausible verifiability account of meaning could connect intelligible thoughts and beliefs about the world with the guaranteed possibility of knowledge of the world we thereby think about.

What seemed to me to be needed for any real progress was greater understanding of the sources of the general epistemological problem. In the early chapters of *The Significance of Philosophical Scepticism* (1984) I tried to present the problem clearly from the ground up, hoping to identify the sources of those demanding requirements on the understanding of knowledge that seemed so difficult to fulfill. Later chapters investigated different attempts to dismiss or expose the philosophical question, or to sidestep or overcome it in some way. Something is to be

learned from each, but I think none of them gets to the heart of the matter.

Explicit presentation of the sceptical reasoning has been followed by a variety of fruitful responses to the problem since about the middle 1980s. One reaction has been not to challenge directly the reasoning by which the sceptical philosophical conclusion is reached, but to question the very significance or import of that conclusion itself. By drawing attention to the relativity to context or to current interests and inquiries that is characteristic of claims involving “know” and other epistemic terms in everyday life, this brings into question the implications of conclusions reached in “philosophical” reflection for anything we ordinarily think or claim to know when not philosophizing.

It remains unclear to me whether a “contextualist” strategy along such lines even countenances the possibility of a positive answer, or at least acknowledges the intelligibility of an all-encompassing question, about human knowledge. If philosophical scepticism is the best answer to a question that arises only in that special context called “philosophy”, is there none the less a question about the possibility and explanation of human knowledge in general? And is there a “contextualist” form of answer to that question about knowledge? Or can we at best simply describe the many different contexts in which we say “Human beings know many things” and ask “How do they know them”?

Another kind of reaction has taken the form of careful responses to one or another step of the sceptical reasoning. This is not to treat the question as unreal or unintelligible, but rather to attempt to disarm the threat of a sceptical answer to it by exposing and rejecting one or another of the assumptions on which it rests. It can leave room for a positive answer to the general epistemological problem once the alleged fallacies of scepticism have been set aside. In at least some quarters these matters are still in dispute.

One suggestion was to deny the necessity of an assumption apparently central to the sceptical case – that if one knows that *p* and knows that “It is true that *p*” implies “It is true that *q*”, then one knows that *q*. But blanket rejection

of that principle would seem to leave a great deal of our otherwise unproblematic knowledge unaccounted for. A more plausible strategy was to deny that one’s knowing that *p* implies one’s knowing that one knows that *p*. This has encouraged a variety of “externalist” or “reliabilist” conceptions of knowledge according to which conditions sufficient for one’s knowing that *p* can be fulfilled even if one does not know or have reason to believe that they are fulfilled. The success of any such theory depends first on finding a set of conditions that correctly captures all possible positive and negative instances of knowledge – a continuation of the Gettier “definitional” project that I have not entered into. But even a “definition” that is correct in that sense will give us what we want only if a non-“externalist” conception of knowledge actually plays an essential role in the sceptical reasoning and nothing else within that way of thinking still represents an obstacle to the kind of understanding of knowledge that we seek.

I think the major obstacle in the traditional epistemological project is its commitment to the restrictive view of the proper objects of perception that has been with us in one form or another since at least the seventeenth century. That is what gives sense to the very idea of the world as “external” to or “beyond” everything we could ever directly perceive to be so. In “Understanding Human Knowledge in General” (1989), “Scepticism, ‘Externalism’, and Goal of Epistemology” (1994), and “Perceptual Knowledge and Epistemological Satisfaction” (2004) I have tried to illustrate the inevitable dissatisfaction I think even acceptance of an enlightened “externalist” conception of knowledge would leave us in as long as we continue to hold any such restrictive view of perception. The idea is that even if, according to an “externalist” conception of knowledge that we accept, we do know some particular thing about the world around us we still will not satisfactorily understand our possession of that knowledge if we grant that we can never actually *perceive* that things are a certain way in the world we claim to know about.

What I think is needed to undermine the appeal of any such view of perception



is more than an “externalist” conception of knowledge. We need a better understanding of the conditions of our even thinking of ourselves as having the thoughts, beliefs, and perceptions involved in making sense of the philosophical question in a way that could seem to present a completely general challenge. The full Kantian aspiration of deriving substantive conclusions about the world and our knowledge of it from no more than the conditions of our thinking of the world in certain ways still seems to me too demanding to promise results we could find palatable. But deeper understanding of the conditions of our even thinking and believing what we do about the world might serve to undermine the sceptical threat without going so far as to imply that the world is a certain way or that we do in fact know the truth of those things philosophical scepticism says we do not know. Establishing the denial of scepticism is not the only possible way to overcome what looks like a sceptical threat. It would be enough to find that we cannot consistently get into the position of understanding and recognizing the force of any such general threat to whatever knowledge we have.

I have outlined in very programmatic form one version of a strategy along these lines and begun to explore some of its epistemological implications in “Kantian Argument, Conceptual Capacities, and Invulnerability” (1994), “Radical Interpretation and Philosophical Scepticism” (1994), and “The Goal of Transcendental Arguments” (1999). The idea is that in order to investigate human knowledge in general we must at least attribute to the human beings we would study, including ourselves, certain thoughts and beliefs and experiences concerning the world in which we take those people to live. There is a question of the conditions of anyone’s having the particular thoughts, beliefs, and experiences that we attribute, and there is a question of the conditions of our being able to ascribe such thoughts, beliefs, and experiences to anyone.

The thought behind this strategy is that to understand people as having certain thoughts and experiences, and to be in a position to attribute such attitudes to them (which is

required for thinking of them as faced by the general epistemological problem), we must understand those psychological states as connected in some ways with the surrounding non-psychological world that we take the thoughts and experiences in question to be about. In “The Epistemological Promise of Externalism” (2005) I tried to sketch how we must move beyond “externalism” about knowledge to a broader “externalist” or “anti-individualist” understanding of all mental contents. That would mean that we can make the attributions needed to formulate the general question of knowledge only because we ourselves are engaged in and have experiences and beliefs about the very non-psychological world that those attitudes are directed towards. This obviously does not imply that those beliefs of ours are true or that the world is just as we experience it to be. But it does mean that anything we must believe about the world in order to attribute to ourselves or others certain beliefs and experiences cannot consistently be regarded as false while we at the same time hold that we or others do have the beliefs and experiences we attribute to them. Whatever is indispensable to making any such attributions cannot itself be considered false consistently with our making the attributions in question.

The inconsistency involved here would be structurally similar to that of the paradoxical sentence “I believe that it is raining, and it is not raining”. Although it is possible for both conjuncts of that sentence to be true together, no one could consistently hold that what the whole sentence says is true. Applied to the general problem of our belief in a world independent of us and our responses, this would mean that if there are certain things that must be believed by anyone who thinks of himself and others as believing in such a world, then the (admitted) possibility that we believe in such a world while those certain things are false is not a possibility we could consistently believe to be actual. Realizing that that is so, we would recognize no pressing need to explain how we know that that conjunctive possibility is not actual.

Whatever is indispensable in this way to pursuit of a traditional epistemological inquiry

into the credentials of our beliefs could be said to enjoy a certain kind of epistemological invulnerability. It is not something that could consistently be brought into question by any epistemological assessment that requires acceptance of that very belief as true. This does not imply that the indispensable belief in question is true, or known to be true, or that it could not be abandoned in the face of weighty considerations against it. It would not have been shown to be invulnerable to all possible attack. The invulnerability in question, given the indispensability, would be only invulnerability against a certain kind of negative epistemological assessment. But that is just the kind of assessment to which the traditional epistemological project would subject our beliefs about the world in general.

So far this is no more than a sketch of the skeleton of what I see as a promising line of investigation into the sources of the traditional epistemological enterprise and what has looked like its inevitably sceptical outcome. It remains to be seen how far it can take us. But even a fully successful investigation along these lines would not yield a reassuringly positive anti-sceptical answer to a general question about human knowledge. It would at best expose the traditional epistemological challenge as something we cannot get into a position to be faced with. And that can leave us philosophically dissatisfied – as if there is something we still lack. We shall have put the completely general epistemological problem forever behind us only when we have managed to overcome that apparently undying desire to explain human knowledge as a single, comprehensible whole, and from a position somehow recognizably outside it. Such a view from nowhere can appear to be reveal the truth of a completely general philosophical scepticism. I think it is better to try to explain why there can be no such thing as a view from nowhere than to argue that with greater care we can see from there that scepticism is not really true.

#### BIBLIOGRAPHY

Stroud, Barry: *The Significance of Philosophical Scepticism* (Oxford: Oxford University Press, 1984).

Stroud, Barry: "Perceptual Knowledge and Epistemological Satisfaction," in *Sosa and His Critics*, ed. J. Greco (Oxford: Blackwell, 2004).

Stroud, Barry: "The Epistemological Promise of Externalism," in *The Externalist Challenge: New Studies in Cognition and Intentionality*, ed. R. Schantz (Berlin and New York: De Gruyter, 2005).

Stroud, Barry: *Understanding Human Knowledge: Philosophical Essays* (Oxford: Oxford University Press, 2000). (Essays not mentioned here are published in this collection.)

## Michael Williams

### 1. THREE THEMES

My work in epistemology explores three themes: opposition to epistemological foundationalism, a diagnostic approach to epistemological scepticism, and a certain scepticism about the very idea of epistemology, that is, about the idea that knowledge and justification are appropriate objects of theory. I take my three themes to be closely related. Foundationalism is typically seen as a response to the threat of scepticism. But I think that scepticism arises out of the quest for foundations, a quest that is itself bound up with the very idea of knowledge as an object of theory. Thus, the cure for epistemological scepticism is scepticism about epistemology. Weaving these strands together leads to a picture of knowledge and justification that I call "contextualist", though my contextualism differs from some familiar contextualist views. The main influences on my thinking have been Sellars, Quine, Wittgenstein, and my teachers, A. J. Ayer and Richard Rorty.

In speaking of scepticism about epistemology, I do not mean to deny that we can make our unreflective understanding of epistemological concepts more explicit. But whether knowledge itself is theorizable is another question. A useful comparison is with the deflationary approach to truth (which I endorse). Deflationary "theories" of truth have two components: an account of the linguistic rules governing the use of "true", and an account of why it is useful for us to have such a word. Although deflationists about

truth differ with respect to the first component, all agree that the function of truth-talk is expressive: it enables us to endorse (or reject) sentences (or propositions) that we cannot state, either because there are too many of them ("Every sentence of the form 'P or not-P' is true"), or because we do not know exactly what they are ("Everything John told you is true"). Such views of truth are deflationary because they take these modest claims about "true" to exhaust the topic of truth, there being no further questions about the nature of "truth itself". I approach epistemic notions in a similar spirit.

## 2. KNOWLEDGE AND JUSTIFICATION

I have never been sanguine about solving Gettier's problem. But, in so far as we can say in general terms what it takes for someone to have knowledge, my sympathies are with the "justified, true belief" account. Initially, however, this preference amounts to little more than a terminological decision. A justified belief is one that it is "epistemically appropriate" to hold. Since by definition beliefs are never more epistemically appropriate than when they amount to knowledge, anyone who knows that P has justification to believe that P. The interesting questions concern the factors that make for justification. Here, I think, we find ourselves pulled in two directions. Traditional accounts of justification tend to emphasize the importance of a subject's having and making proper use of good reasons for believing as he does. Such accounts are therefore broadly "responsibilist" or deontological: knowledge (or justified belief) depends on epistemically appropriate conduct. Since regulating and being accountable for one's epistemic conduct requires knowing what one is up to, the responsibilist approach to justification is broadly "internalist". Today, however, many epistemologists approach knowledge from a "reliabilist" perspective, holding that a true belief amounts to knowledge when acquired and retained through the operation of some reliable cognitive process. Reliabilist epistemology is externalist, the epistemic status of a belief depending

solely on the de facto reliability of the processes that produce and sustain it. Knowing that-p does not require even being in a position to know how or why. Though reliabilist ideas can be found in antiquity, their re-emergence in the latter half of the last century constitutes a revolution in epistemology.

Among its many strengths, reliabilism explains why we are interested in knowledge and justification, including justification that takes the special form of basing beliefs on reasons. To cope with the ever-changing circumstances in the world around us, we need to acquire true beliefs while avoiding errors. Justificatory, in the sense of truth-reliable, procedures are the means to this end. However, much knowledge arises from the unself-conscious operation of basic cognitive capacities, or through social mechanisms like testimony. Weighing evidence, marshalling reasons, is a rather special case of a truth-conducive procedure. Responsibilist epistemologies seem over-intellectualized. But pure reliabilism pushes this point too far, losing sight of the normative character of terms of epistemic appraisal. Unlike animals, human beings are held accountable for their beliefs in much the way that they are held accountable for their actions. Responsibilism recognizes this essential point.

I think, then, that both approaches to epistemology have genuine intuitive appeal. The problem is to combine them. This brings me to scepticism.

## 3. SCEPTICISM

In common parlance, the name "scepticism" is often given to a (generally desirable) critical stance. The sceptic is not easily taken in: he requires good reasons for what he believes. This ordinary scepticism presupposes a workable distinction between beliefs that are appropriately grounded and beliefs that are not. As a result, it is selective.

Philosophical scepticism is a different animal. Philosophical sceptics argue that knowledge – either knowledge *tout court* or knowledge of facts belonging to certain very broad domains, such as facts pertaining to

the external world – is impossible. More radically, they argue that nothing we believe is even so much as justified. Thus, knowledge is beyond us, not because we fail to meet exacting standards but because we fail to meet even the most modest, and on a very broad front. But, while such radical and general conclusions are difficult to swallow, arguments for them can strike us as very persuasive. They seem “natural” or “intuitive”, invoking or presupposing only the most banal, lowest-common-denominator ideas about knowledge and justification. If they are not intuitive, it takes some showing. In any case, we only understand traditional epistemology’s goal of explaining how knowledge is possible because there are seemingly persuasive arguments to the effect that knowledge is impossible. The most important such arguments belong to either of two families, which I call “Agrippan” and “Cartesian”.

Agrippan arguments turn on an apparently disastrous trilemma. The trilemma, which I call Agrippa’s Trilemma after the ancient sceptic said to be the first to have articulated it, seems to be implicit in the very idea of epistemic justification. If I have a (supposedly) justified belief, there is presumably something that makes it justified. So what is this something? If I explain, the sceptic asks me whether what I have just said is itself justified. If I offer a further explanation, the question of justification can be renewed. And so on. If I play along, there seem to be three possible outcomes:

- (i) The justificatory chain goes on for ever.
- (ii) The justificatory chain stops with a proposition for which I have no justification.
- (iii) In tracing my justification for justification for justification . . . to believe that P, we are led back to reasons that include my belief that P.

According to the sceptic, all three outcomes are bad. If the justificatory chain goes on for ever, I am trapped in a vicious infinite regress. If the chain stops with a belief for which I have no justification, I am making a brute assumption. Finally, if the justificatory

chain loops back on itself, my “justification” involves circular reasoning. In no case do we find genuine justification. But there is no fourth option. The result is scepticism.

Cartesian arguments work differently. Their hallmark is an appeal to sceptical hypotheses: for example, the hypothesis that I am systematically deceived by an Evil Deceiver or that I am a brain in a vat. In such a sceptical situation, things would seem just as they do in a “normal” world. How, therefore, do I know that I am not in a sceptical situation? Descartes’ original problem about the external world provides a template for several other sceptical problems: other minds and the past, for example.

The proper way to formulate Cartesian arguments is controversial. In my view, they are best-understood as posing underdetermination problems. In the case of beliefs about the external world, all I ultimately have to go on is the evidence of my senses: how things, in experience, appear to me. Unfortunately, this evidence seems to be neutral with respect to our ordinary view and various sceptical alternatives. If this is so, I can have no reason to think that a sceptical alternative does not obtain.

The Agrippan problem has exerted a powerful influence on epistemological theorizing. Accepting the Agrippan sceptic’s three options, traditional epistemologists argue that we can put a better face on one of them. Because the regress strikes most theorists as obviously vicious, we have two choices:

- Foundationalism: The justificatory chain terminates, but not with a brute assumption.
- Coherentism: There are closed inferential loops, but their presence does not imply justificatory circularity.

Because these approaches accept the problem in the sceptic’s terms, they amount to direct responses to scepticism. For the foundationalist, justification stops with basic beliefs that are “intrinsically credible”. For the coherentist, justification is grounded on certain general principles, which define coherence.

Coherentists also tend to presuppose that we have more – or less unproblematic access to what we believe. In my view, coherentism is a variant form of foundationalism.

Cartesian arguments, too, have shaped the space of epistemological theories. Reductionist theories (e.g. phenomenism) treat claims in the problematic class as logically elaborated from claims in the privileged (evidence-conferring) class; criterial theories postulate a priori evidential connections between privileged and problematic beliefs; abductive approaches treat, say, belief in the external world as justifiable by an inference to the best explanation (from certain features of the course of perceptual experience). Again, these responses to the sceptic are direct. Under the influence of Agrippan scepticism, they accept the sceptic's conception of our ultimate evidential impoverishment: if they are to be intrinsically credible, basic beliefs are not likely to be ambitious in their content. But once we accept a narrow basis for our beliefs we need to explain how we transcend it.

#### 4. THEORETICAL DIAGNOSIS

I hold that it is a mistake to try to answer the sceptic in his own terms. In particular, it is a mistake to let the sceptic bully us into accepting a foundational view of knowledge and justification. The key foundationalist idea – that of intrinsic credibility – cannot be defended. And, even if it could, it would be a case of out of the Agrippan frying pan into the Cartesian fire.

Scepticism is best-approached diagnostically. Crucially, this means destroying the air of naturalness of intuitiveness that makes sceptical arguments so troubling. This means isolating, with a view to contesting, the sceptic's unacknowledged theoretical presuppositions. I call this task "theoretical diagnosis", in contrast to "therapeutic diagnosis", which tries to convict the sceptic of "misusing language" or otherwise failing to speak in a fully intelligible way. I think that we do understand the sceptic, at least well enough to see how his arguments determine the space of epistemological theories.

Given the externalist revolution, one obvious diagnostic strategy is to charge the Agrippan sceptic with an unmotivated epistemic internalism. The sceptic assumes that someone is not justified in believing that-P unless he can state his justification, exactly what externalism denies. But, while there is something to this suggestion, as a diagnosis of scepticism it falls short. It says how scepticism goes wrong, but doesn't explain scepticism's appeal. Further, it threatens to trade scepticism for an almost equally implausible meta-scepticism. If we are favourably placed in the world, we shall know lots of things. If not, we won't. But, given the Cartesian problem, how do we know where we stand?

In any case, I reject pure externalism. The sceptic goes wrong not in connecting being justified with the ability to justify but in how he makes the connection. In threatening us with a vicious regress of justification, the sceptic assumes that epistemic challenges (such as "How do you know that?") are indefinitely iterable because invited by anyone who claims knowledge. This assumption of claimant-challenger asymmetry reflects a particular model of epistemic justification, in which justification always depends on positive authorization. I call it the Prior Grounding conception. Whenever a belief or claim is justified, there is some statable positive property in virtue of which it is justified. All the sceptic wants is to be told what it is: reasonably enough, given the model. However, the principle of claimant-challenger asymmetry will be false if challenges are themselves subject to justificatory constraints. But, if challenges are not always legitimate, a belief may be justified because, in the circumstances, there is nothing to be said against it. Globally, justification will exhibit what Robert Brandom calls a "default and challenge structure". On this view, statable positive authorization will be an essentially local phenomenon. Justification thus connects with the ability to justify, or give reasons, in the following way: one must be able to give reasons in the context of a legitimate challenge. So, while justified belief does not always require positive authorization, it does always carry a qualified defence commitment.



The factors influencing default justification are contextually sensitive. An important positive task for my kind of contextualism is to identify these factors. Some are dialectical (are there standing objections to a claim?). Others have to do with our real-world situation. Still others have to do with the costs of error (or rewards of getting things right). The view might look like a kind of foundationalism. I don't argue about the name. What matters is to realize that default-credible commitments cannot be classified by their contents. They do not constitute an epistemic kind. This point shows how my response to Agrippan scepticism works against Cartesian scepticism, too. Without epistemic kinds, Cartesian scepticism falls to the ground.

Cartesian scepticism rests on a metaphysics of knowledge that I call "epistemological realism". This is not realism as a position within epistemology or metaphysics (the view that there is a real world out there) but rather an extreme realism about the objects of epistemological inquiry. The Cartesian sceptic imagines that it is just a fact, written into the human condition, that our beliefs about the external world depend for their justifiability on experience: awareness of how things appear to us. On this view, our beliefs, by virtue of certain broad features of their content, fall into natural epistemological kinds, standing in natural relations of epistemological priority. But there are no such relations and no such kinds. What can be invoked to justify what varies with the context of inquiry.

If this is right, Cartesian scepticism is rooted in the very idea of trying to understand our knowledge of the external world in the fully general way typical of traditional epistemology. The problem is with the category itself. The category has no obvious theoretical integrity. It looks like a rag-bag: comparable to "things that happen on Wednesdays", which no one would make the object of systematic investigation. However, the category starts to look more sensible when we realize that its real basis is epistemological. What unites beliefs about the external world is not any common subject-matter but their uniform dependence on experiential evidence. If

this idea of general dependence is rejected, "beliefs about the external world" goes back to being a rag-bag. The sceptic challenges us to explain how we know anything at all about the external world. My response is that there is nothing to understand.

All epistemologists agree that we do not face sceptical problems – Agrippan or Cartesian – in everyday justificational contexts. The problems come to light only in the context of trying to reflect in an unusually general and systematic way on knowledge and justification. But, in my view, this kind of systematic reflection does not bring to light the hidden structure of everyday justification. Rather, the structure supposedly discovered is a self-generated requirement on this kind of epistemological theorizing.

This is why I say that scepticism about epistemology is the key to epistemological scepticism. It is also my point of contact with other forms of contextualism. All contextualists agree that scepticism's plausibility is tied to the peculiar context of "doing epistemology". However, many contextualists think that the sceptic seems right because, in the context of doing epistemology, he is right. In the context of epistemology, everyday knowledge goes missing. My response is far less concessive. I think that the sceptic's investigative context is fatally flawed, with the result that he is never more than conditionally correct.

Why, then, do sceptical arguments even seem intuitive? For several reasons. One is that we do not go around stating the obvious. Accordingly, things worth talking about are often not default credible, so the need to be able to justify seems more prevalent than it is. Another reason is that patterns of sceptical questions often mimic patterns of ordinary questioning: thus, introducing sceptical hypotheses looks similar to raising ordinary counter-possibilities to a particular claim, albeit in the context of an unusually systematic examination of our knowledge of the world. And even that systematic examination can be made to look more innocent than it is. Exploration of how this is brought about is a central task of theoretical diagnosis.

## 5. RESPONSIBILITY AND RELIABILITY

A good picture of knowledge and justification should reconcile responsibility with reliability. Traditional epistemologists have had a hard time with this, and their accounts of knowledge and justification often seem over-intellectualized. The source of these difficulties is that the project of reconciling responsibility with reliability has been assimilated to that of showing, in some very general way, that epistemically responsible belief-management conduces to holding true beliefs: in effect, the project has been assimilated to that of answering the sceptic. Once we see that the sceptic does not need to be answered directly, the project looks much more tractable. Since we have all sorts of default entitlements, the task of responsible belief-management is less onerous than has often been supposed. In many everyday contexts, the reliability of our more – or less automatic cognitive processes can be legitimately presupposed, no positive authorization necessary. However, such reliability-presuppositions are vitally important as a source of motivated challenges. If, in a particular context, we are given reasons to think that these presuppositions fail to hold, default entitlement lapses.

Sellars put his finger on the deepest flaw in the picture of knowledge as resting on foundations: its peculiarly static character. I think that Sellars could have said the same about traditional coherentism (though he didn't). Both foundationalism and coherentism are time-slice theories: they seek to identify some structural feature of our belief-system in virtue of which the elements comprising it are justified. This picture already starts to make the question "How are we justified in believing anything at all?" look like a good one. By contrast, the contextualist vision is dynamic: we are always already *in medias res*, with a complex set of beliefs about the world and about our reliability as knowers. Questions of justification always arise in such rich, historically evolving contexts. As problems come up, we make revisions in our beliefs about the world and in our epistemic beliefs, too. But improvement is a retail affair, and generally speaking the epistemologist is not our guide.

## WORKS BY MICHAEL WILLIAMS

- Groundless Belief* (Oxford: Basil Blackwell, 1977).  
*Groundless Belief*, 2nd edn (Princeton, NJ: Princeton University Press, 1999).  
*Unnatural Doubts* (Princeton, NJ: Princeton University Press, 1992).  
*Problems of Knowledge* (Oxford: Oxford University Press, 2001).  
 "Knowledge, Reflection and Sceptical Hypotheses," *Erkenntnis* 61 (2005).  
 "Why (Wittgensteinian) Contextualism Isn't Relativism," *Episteme* 4 (2007).  
 "Responsibility and Reliability," *Philosophical Papers* (forthcoming).

**Timothy Williamson**

Epistemology is the theory of knowledge. Unlike many epistemologists, I treat knowledge as the basic epistemological phenomenon: not for etymological reasons but because it is also the key to understanding the other phenomena that rightly concern epistemology, such as perception, memory, testimony, inference, evidence, justification and probability.

Knowledge manifestly involves a relation between the subject and the world, the knower and the known. It has both a subjective and an objective side. Knowing involves a mental state: it requires the subject to believe the truth. It is also factive, in the sense that, necessarily, only truths are known. One knows only when the world is as one believes it to be. Arguably, knowing is the most general factive mental state. Many philosophers regard the two-sided nature of knowledge as a reason for treating it not as a basic phenomenon but as a hybrid, whose complexity must be decomposed into more basic components, some purely subjective, some purely objective. After Gettier's refutation of the analysis of knowledge as justified true belief, an industry arose for analysing knowledge into truth, belief and further conditions. Notoriously, the series of analyses, counter-examples and added epicycles has not converged. If I am right, the reason is that knowledge does not decompose in the way assumed. Just as it is futile to ask what one must add to being coloured to get being red, it is equally futile

to ask what one must add to believing truly to get knowing; a necessary condition need not be a component of a non-circular necessary and sufficient condition.

The two-sided nature of knowledge is typical of central mental phenomena: other examples are *awareness*, *reference* and *action*. They, too, involve a relation between the subject and the world. That is the nature of the mental. It is not something inside the subject, but a way in which the subject relates to the world. In theorizing about it, our primary need is for success terms such as “know” and “do”, rather than success-neutral terms such as “believe” and “try”. Of course, believing and trying are genuine mental phenomena, too, but the proposed strategy for understanding them is to treat them as derivative from their success conditions. Failure makes sense only in relation to a standard of success, malfunction only in relation to function.

One test for this reorientation is to show that factive mental states such as knowing play a robust role in the causal explanation of action. It is often claimed that knowing, unlike believing, is not a causal-explanatory mental state because what explains one's actions is one's beliefs, not one's knowledge, for one's actions are the same whether one knows or merely believes. Philosophers have tended to focus on basic actions such as moving your arm, to which the contribution of the world is minimal – although it makes a difference whether you know or falsely believe that you have an arm. The neural firings that were the immediate cause of the movement occur at a sub-personal level of which the subject is typically ignorant; they are not intentional actions, not what was to be explained. Suppose that attempts to perform basic actions are proposed as the primary explananda. Most actions are not basic actions: consider eating an apple, walking home. . . . The strategy of explaining such non-basic actions by providing the immediate explanations of their constituent basic actions faces severe problems. One is that the reasons for the non-basic action need not figure in the immediate explanation of any of its constituent basic actions. My reasons for walking home explain my decision to walk home, but

then the intention to walk home takes over, and the original reasons are no part of the immediate explanation of any individual step. Thus, the rationality of walking home disappears from view. Another problem is a loss of significant generality. There are many ways to walk home; exactly which steps I take depends on a mass of accidental circumstances (such as traffic). A combined explanation of each of my individual steps is highly specific; it does not cover other possibilities in which I walk home for the same reasons by a slightly different sequence of steps – unlike the explanation of my walking home in terms of my original reasons. But significant generality is an explanatory virtue. Often, what we need is an explanation of an extended non-basic action in terms of the subject's original reasons, and attributing knowledge plays an explanatory role that attributing belief cannot. For example, in explaining why I walked home – not just why I initially attempted to – it is more relevant that I knew that home was within walking distance than that I believed that it was. Such arguments can be extended to show that knowing plays just the sort of robust causal-explanatory role that one would expect of a genuine mental state.

The concept of knowledge is first-personal, in the sense that the subject can self-attribute knowledge or ignorance, just like the presence or absence of other mental states such as belief, without having to observe her own behaviour as someone else might. Sometimes we know that we know; sometimes we know that we do not know. However, the concept of knowledge is not fully operational. We are not always in a position to know whether we know. For example, when one has been hoaxed, one has good evidence that one knows something that is in fact false; one lacks knowledge without being in a position to know that one lacks knowledge. One can also know without being in a position to know that one knows. If concepts of some other mental states, such as belief, were fully operational, that would be an important difference. For then part of epistemology might be expected to be fully operationalized: it would judge (advise, praise, blame) the

subject by criteria whose application was in principle entirely within the subject's ken. It would not employ the concept of knowledge. The theory of justification, rationality and evidence has often been conceived along such lines. Subjective Bayesian epistemology is a mathematically if not philosophically sophisticated attempt to achieve operationalization.

I have argued that full operationalization is an impossible ideal. Call a condition "luminous" if, whenever it obtains, the subject is in a position to know that it obtains. For instance, the condition that one seems to see a red patch might be thought to be luminous. The anti-luminosity argument is an argument to the conclusion that only trivial conditions are luminous. Since full operationalization requires non-trivial luminous conditions, full operationalization is impossible. The anti-luminosity argument works by considering a process of gradual change from a state in which the given condition clearly obtains to a state in which it clearly fails to obtain. A human subject with limited powers of discrimination is not in a position to know during the process exactly how far it has gone. When the condition still obtains at moments just before moments at which it has ceased to obtain, the subject relies on its obtaining almost as much at the latter moments (when it fails) as at the former ones. Thus, the reliance at the former moments is too indiscriminate for knowledge. Just before the condition ceases to obtain, the subject is not in a position to know that it obtains; therefore it is non-luminous.

One can describe the anti-luminosity argument by saying that it uses a "margin for error" or "safety" constraint on knowledge. However, this should not be envisaged as a potential component of a non-circular analysis of knowledge. One cannot specify how wide a margin or how much safety knowledge requires except in terms of knowledge itself: enough for knowing. This point is overlooked by many criticisms of the anti-luminosity argument and of the safety condition. Other challenges exploit borderline cases that are easily interpretable as safe and knowledgeable but also easily interpretable as unsafe and unknowledgeable; by confusing the two

interpretations one can convince oneself that one has a case of unsafe knowledge. We are not guaranteed to be in a position to know which interpretation is correct. I have defended the robustness of the anti-luminosity argument by showing that versions of it go through under a range of variations in the initial assumptions. Analogous arguments apply even to probabilistic notions.

Many issues in epistemology have been distorted by attempted operationalization. For example, evidence as ordinarily understood in science is obviously non-operational. A scientist may not be in a position to know that an experimental result is not part of the evidence, because he has been deceived by a disgruntled lab technician. On a fully operationalized notion of evidence, one would always be in a position to know whether something was part of one's evidence. Thus, epistemologists have been driven to conceiving one's evidence as one's present phenomenal states, despite their apparent irrelevance to most of what one needs evidence for. The anti-luminosity argument implies that not even phenomenal states meet such epistemologists' demands for full operationalization, since one is not always in a position to know whether one is in a given phenomenal state. Abandoning those impossible demands frees us to return to a more realistic conception of evidence. In fact, I have argued that one's total evidence is simply the totality of what one knows, and developed a related theory of evidential probability. Since justification depends on the subject's evidence, it depends on the subject's knowledge. Instead of justification explaining knowledge, as in analyses with justified true belief and further conditions, knowledge explains justification.

This "externalist" conception of evidence and justification impacts on the problem of scepticism. Most epistemologists assume without argument that the victim of a perfect sceptical scenario (a dreamer or brain in a vat) has exactly the same evidence and justification as the corresponding undeceived subject. On the externalist conception, it is begging the question in favour of the sceptic to assume that the two subjects have the same evidence, for unless scepticism holds

the undeceived subject knows more than the victim of the sceptical scenario. The two scenarios are indiscriminable at least for the sceptic's victim, since, for all he knows, he is the undeceived subject – although the undeceived subject knows many things that entail that he is not the sceptic's victim. But evidence and justification can differ across indiscriminable situations. After all, many indiscriminable differences can add up to a discriminable difference, as in a sorites series: if evidence and justification were identical across indiscriminable situations, they would have to be identical across discriminable situations, too, which is absurd.

Justification varies between the sceptic's victim and the undeceived subject in another way: the content of their justified beliefs. For instance, when I am the undeceived subject, I am justified in believing that this hand is warm. The sceptic's victim is not justified in believing that *this* hand is warm, for this hand is my hand, and none of his evidence bears on my hand. In general, externalism about content can be shown to force differences in justification between intrinsic duplicates. The conception of knowing as a mental state extends externalism about the mental from the content of propositional attitudes to the attitudes themselves.

If knowledge is what justifies belief, to the degree to which it is justified, the natural conclusion is that a belief is *fully* justified only if it constitutes knowledge. I have defended a similar view of assertion, the closest analogue in speech to belief. Specifically, the constitutive rule for assertion is that one should assert only what one knows. Alternative accounts of assertion often reflect the futile search for a fully operationalized criterion.

One application of the theory just sketched is to the problem of vagueness. Safety considerations predict that, even if vague terms have sharp boundaries, speakers will not always be in a position to know which side of the boundary a given case is on. Moreover, the obstacle is specific to vagueness, because cases in the relevant margin for error involve variations in the reference of the term at issue. Thus, our ignorance in border-

line cases does not show that vague terms lack sharp boundaries. I have developed an epistemic theory of vagueness, on which classical logic and semantics apply to vague languages but speakers are in no position to trace exact boundaries for the reference of their terms. Critics of epistemicism tend to rely on over-optimistic assumptions about our capacity to know how the use of an expression determines its reference.

My work on indiscriminability relates to the preceding themes. I understand discriminating between objects or properties in terms of knowing that they are different (non-identical). The logic of indiscriminability, including its non-transitivity, can then be derived from the interaction of epistemic logic with the logic of identity.

Another application of the anti-luminosity argument is to claims that the truth of ascriptions of "knowledge" and other terms of epistemic appraisal is sensitive to features of the context – that of the ascriber or of the subject, depending on the theory – traditionally considered irrelevant, such as practical interests and conversational salience. I have argued that the examples used to support such claims can be understood, consistently with the denial of such contextual effects, in terms of a hierarchy of knowledge, knowledge that one knows, and so on, implicit in the anti-luminosity argument.

Externalism about evidence also has implications for philosophical method. Analytic philosophers tend to speak as though our evidence consists of our "intuitions": not their content (which may be false) but the fact that we have them. This psychologization of evidence results from an implicit attempt to operationalize the epistemology of philosophy. It is sometimes defended on the grounds that we have to start from where we are, with our current beliefs. But that does not justify the idea that our evidence consists of our own psychological states. For where we are includes our current knowledge, not simply our current beliefs. We start with an awareness of the world, not merely of ourselves. The data to be explained include data about the external world as well as data about our "intuitions". Someone might object



that the data should not be described in terms that some philosophers will refuse to accept. Unfortunately, given *any* description of the data, some philosopher will refuse to accept it. For example, strict eliminativists about mental states will reject descriptions of the data in terms of “intuitions”; they require descriptions in physical terms. In recent work, I have reinterpreted the epistemology of philosophy in more externalist terms.

One kind of limit to knowledge comes from the anti-luminosity argument. Another kind comes from the so-called “paradox of knowability”. This is a formal proof, first published by Fitch, that, on weak assumptions, if all truths are knowable, then all truths are eventually known. Since various forms of anti-realism describe truth as outrunning what is eventually known but not what could be known, the proof has obvious philosophical significance. Why realists should call it a “paradox” rather than a “result” is unclear. I have explored the scope of Fitch’s proof, its use of classical logic and the possibility of weakening the premises still further and generalizing the conclusion to other attitudes. Since anti-realists were persuaded to take the result seriously, they have tended to respond with qualifications to the claim that all truths are knowable that are both *ad hoc* and hard to reconcile with their original arguments for the claim.

I have used the resources of modal and epistemic logic to conduct other formal investigations in epistemology. For example: Do verificationist theories of meaning collapse truth into verifiability? What constraints are imposed on epistemic logic by the assumption that the subject’s computational powers do not exceed those of a Turing machine? Philosophers have often been suspicious of epistemic logic, not least because its practitioners have often treated subjects as logically omniscient. In consequence, few of them have taken advantage of the opportunities that it affords for the rigorous exploration of the consequences of epistemological ideas. One way is to tailor the interpretation of the operators to validate the epistemic logic at issue, for instance by considering what follows from what the subject knows rather than

simply what the subject knows. Another way is to idealize explicitly in order to explore the effect of obstacles to knowledge other than failure of logical omniscience. There are still further ways.

In general, epistemological arguments in ordinary language often obscure far more logically relevant structure than one might think. For example, they can conceal tacit elisions of knowing and knowing that one knows. Although formal methods by themselves cannot solve epistemological problems, they can produce clarifications that would be hard to achieve without them. Sometimes the required degree of formalization is slight, as when one tests an argument by laying it out in labelled premises and conclusion in regimented English. Sometimes it is more serious, as when one proves the consistency of some epistemological claims by constructing a formal model in which they are all true.

One technique uses formal semantic theories of natural languages. Jason Stanley and I argued by these means, *contra* Ryle, that all knowing how is knowing that: roughly, knowing how to do something is knowing, of an appropriately presented way, that it is the way to do it. The conclusion reinforces the centrality of propositional knowledge, and undermines attempts to classify knowing what it is like to see red or knowing what a word means as knowing how to the exclusion of knowing that. According to some critics, these appeals to the semantics of English yield conclusions only about the English word “know”, whereas epistemologists want to know about knowledge itself. But the point of such applications of semantic theories is to determine more exactly which arguments in the natural language are valid. Having ascertained the validity of an argument in English, where the word “know” is used but not mentioned, we can use the argument to reach a conclusion about knowledge, not merely about “know”.

In brief: Knowledge is central; the gap between truth and knowability is pervasive. Formal methods have a place in epistemology; properly applied, they make it a bit more objective, and a bit more fun.

## BIBLIOGRAPHY

- Williamson, T.: *Identity and Discrimination* (Oxford: Blackwell, 1990).
- Williamson, T.: *Vagueness* (London: Routledge, 1994).
- Williamson, T.: *Knowledge and Its Limits* (Oxford: Oxford University Press, 2000).
- Williamson, T.: "Contextualism, Subject-sensitive Invariantism and Knowledge of Knowledge," *Philosophical Quarterly* 55 (2005), 213–35.
- Williamson, T.: *An Essay on Philosophical Method* (Oxford: Blackwell (2008).
- Williamson, T. and Stanley, J.: "Knowing How," *Journal of Philosophy* 98 (2001), 411–44.

## Crispin Wright

One of the most ancient philosophical challenges is to respond to the Sceptical thought that no objective difference separates our knowledge, or more generally those of our opinions we are inclined to regard as best, from convictions we dismiss as superstitious, prejudicial or dogmatic. My own first attempts to do so were conceived in reaction against a climate, during the late 1970s, when epistemologists tended to divide into two broad camps: those who, in Wittgenstein's name or under the lingering influence of "ordinary language" philosophy, were prone to dismiss sceptical doubt as nonsensical; and those who tamely conceded that the strongest versions of the sceptical challenge are simply unanswerable, permitting nothing better by way of response than a Strawsonian, or Lewisian, shrug of the shoulders. My conviction then, reinforced by important writings in the early 1980s of Nozick and Stroud, was that it should be possible to do much better: that scepticism is neither nonsense nor unassailable but the dead-end conclusion of a range of (in detail, significantly differing) *paradoxes*. Paradoxes are things that, if we are rational and smart enough, we should be able to dispel or solve. My researches over more than a quarter of a century have been driven by the quest for a head-on, rationalistic response to these paradoxes, with a direction for so doing, broadly inspired by elements in Wittgenstein's notes *On Certainty*, developing

gradually in a fashion whose most significant stages are marked in "Facts and Certainty" (1985), "Scepticism and Dreaming: Imploding the Demon" (1991),<sup>1</sup> "(Anti)-Sceptics, Simple and Subtle" (2002)<sup>2</sup> and "Warrant for Nothing (and Foundations for Free):" (2004). Although there continues to be little consensus among epistemologists about the proper handling of scepticism, and some continue to regard it as philosophically unimportant, I believe my work has contributed to the predominant change of outlook that is manifest nowadays in the widespread interest in and proliferation of writings on the topic.

My ideas about how to respond to scepticism have been structured by a number of "rules of engagement". First, it really is important to drop the idea of "the Sceptic" as a philosophical opponent since to think otherwise makes things seem easier than they are. For example, a philosophical opponent may be trapped into self-refutation and thereby defeated; but a paradox is not solved just by pointing out that its conclusion could not be accepted by one who accepted its premises. Second, the fundamental issue isn't so much about knowledge as *warrant* – the property that distinguishes belief held in good intellectual conscience from its disreputable opposite. Originally, I directed this claim against those who, like Russell, were inclined to try to make peace with scepticism by suggesting that, although sceptical argument may show that we do not *know* that there is an external world, say, we may nevertheless be well enough warranted in believing that there is. The (best of the) paradoxes, however, need no assumption of the factivity of the epistemic operator concerned and therefore attack not just knowledge but a wide class of virtuous epistemic states whose achievement we would normally have no hesitation in claiming. Yet, in view of the increasing number of epistemologists who draw anti-sceptical comfort from *externalist* conceptions of knowledge and justification, a version of this point merits re-emphasis. Unless you think not merely that knowledge is "external" – that is, roughly, that knowing consists in truly believing for causes and in circumstances that may not be apparent to reflection (for instance, the

belief in question is generated by a reliable method, or holds good in all circumstances that might easily have obtained, or wouldn't have been held had it been false . . . ) – but that *all* forms of doxastic merit are external, the kinds of response to the paradoxes which externalism makes possible won't engage either their application to our pretensions to these other merits – in particular, to warranted belief – or address a certain kind of higher-order sceptical doubt. We may indeed know that there is an external world, if we are lucky enough to live in such a world and, say, our ways of forming beliefs about it are reliable. But to emphasize that congenial possibility takes us no closer towards being in position *rationally to claim* that it is realized. Moreover, when knowledge may be externally grounded, it need not suffice for being in that position that one knows that one knows; rather what is wanted are considerations that allow the claim that one knows to be made in good conscience – something which merely externally constituted second-order knowledge need not do. The tendency of the sceptical paradoxes, properly gauged, is not to assail the possibility of knowledge – an attack that externalism might well deflect – but to engineer a crisis of *intellectual conscience*: to bring about a dialectical situation where we seem to be in no position to lay claim to the knowledge which, if externalism is right, it is possible we have.

A third – and particularly demanding – constraint emerges with this perspective, which I have sometimes expressed by saying that Scepticism wins a tie. Suppose we have a sceptical argument: a derivation from certain premises of the conclusion that none of our beliefs of a certain kind that we routinely take to be fully warranted is in fact so. What status must its premises have in order for this to pose a problem? It's natural to think that they must be motivated – must have some *prima facie* plausibility at least. But if the warrant operator is closed under (known) entailment, so is the operator, "It is not warranted that not". So the paradox will convert to a second-order version whose conclusion – still unacceptable – will be that, just provided we have no warrant actually to *deny* any

of the premises of the argument, we have no warrant to deny that none of our beliefs of the kind in question is warranted, and hence no warrant to affirm that they *are* warranted. It follows that a satisfactory response to the paradox must involve actively challenging one of its premises – it isn't good enough merely to point out, even if it is true, that they have not all been well motivated. This constraint completely scotches any attempt to address scepticism by arguing merely for the "non-compulsory" character of certain of the premises of the sceptical reasoning.

I have found it illuminating to distinguish two broad forms of sceptical paradox. On the one hand, there is *Cartesian scepticism*: the scepticism which makes play with subjectively undetectable scenarios of cognitive impairment – envathood, sustained deception by a powerful *malin génie*, sustained coherent dreaming, and so on. Some of these fantasies are inconsistent with the truth of most of what we take ourselves to know, some merely inconsistent with our ability to know it. It is sometimes supposed that scepticism depends on accepting an insinuated confinement of the scope of our direct awareness to our own occurrent mental lives, leaving us then in a predicament of chancy inference "outwards" to conclusions about the external material world, others' mental states, the past and future, and so on. The punch of Cartesian scepticism is that, properly formulated, it needs no such presupposition of retrenchment but goes to work on any cognitive faculty which is presumed to afford direct engagement with a certain subject matter. Just two premises are needed, plus a proof theory for the targeted epistemic operator that includes a suitable form of closure. The premises are, first, that the acquisition of warrant for a large class of beliefs depends on the proper functioning of the targeted cognitive faculty (and that this is something which we are warranted in believing); and, second, that there is no (acquiring a) warrant for thinking that one is in a state of such proper functioning, rather than in a radical impairment scenario providing a perfect subjective counterfeit of that state. This simple set-up suffices for the conclusion that we are

in no position to lay claim to warrant for the large class of beliefs in question. The reader will probably be able to fill in the details for himself/herself; they are fully explicit in “Facts and Certainty” and “Imploding”.

*Humean scepticism*, by contrast, engages large classes of our beliefs for which we conceive of even the strongest warrant as essentially based on indirect, defeasible evidence. The form of paradox concerned makes no play with dreams, demons, or crazed scientists but – in a fashion for which Hume’s doubts about enumerative induction provide a prototype – argues that our best procedures involve a vicious evidential circularity. The template for this form of scepticism is schematized by what I have called the *I-II-III argument*, laid out in “Facts and Certainty” and revisited in detail in “Simple and Subtle”. It is nicely illustrated, ironically, by an adaptation of G. E. Moore’s celebrated “Proof of an External World” (1939) – in essence, this train of thought: “Here is a hand. If there is a hand here, then there is an external world. Therefore, there is an external world.” The kernel of the Humean sceptical doubt is that Moore’s premise, “Here is a hand”, rests on the evidence of the appearances but that the evidential support so provided needs the setting of a *prior* acceptance of the conclusion of the “Proof”, that there is an external world (and one, moreover, broadly manifest to us in routine sense-experience). Only with that thesis already in place as background “theory” does routine sense-experience have any tendency to support claims about material reality. But a background acceptance of the thesis, in turn, the paradoxical reasoning contends, is rational just in case one has warrant for it. So, in effect, the empirical warrant – which one may indeed have – for one’s belief that there is a hand where one seems to see it actually depends on one’s *already having* warrant for the conclusion of Moore’s argument. The “Proof” is therefore circular, or, as I have usually put it, it *fails to transmit warrant* from its premises to its conclusion.

The application of this form of sceptical paradox to undermine the warrants generated by perception does indeed depend on the repudiation of direct realism implicit

in insisting on a separate evidential layer of appearances. But a corresponding direct realist move is only dubiously available in the case of knowledge of other minds, and unavailable for knowledge of the past beyond living memory, the future and general laws. Wherever you draw the line, *some* of our claims to knowledge are going to turn out to be essentially based on defeasible evidence; and for those cases the charge of circularity – of presupposition, in order for the evidence types in question to have their normally attributed warranting power, of certain very general “hinge” or framework beliefs for which warrant can be foreseen only by inferring them in turn from more local beliefs putatively justified by the kinds of evidence in question (just as Moore tried to do) – will arise.

The notion of failure of transmission of warrant (across a valid argument) is independently interesting and important. Its varieties, proper characterization and role in the sceptical landscape and (arguably) other paradoxes (for example, McKinsey’s argument that content externalism wars with the possibility of “armchair” psychological self-knowledge) take centre stage in several of my papers and have recently received increasing attention from others. Transmission-failure needs sharply distinguishing from the idea of failure of *closure* of warrant (across a valid argument) with which some of the critics of closure, like Nozick and Dretske, have tended to confuse it. For example, a famous alleged counter-example to closure of Dretske’s: the inference from (i) “The animals in that cage are zebras” to (ii) “The animals in that cage are not cleverly disguised mules”, when the warrant for the first is given by casual observation too crude to distinguish the animals in vision from cleverly disguised mules, is arguably a failure of transmission: it is not that one can have a casual observational warrant for (i) but *no* warrant for (ii), but rather that it is only when (ii) (and a range of kindred propositions) are presupposed that casual observation warrants (i). The difference between transmission-failure and the breakdown of closure can be put thus: when closure fails, one may be warranted in believing the premises of a valid argument yet have

no warrant for its conclusion. When transmission fails, one will have a warrant for the conclusion of the argument if one has warrant for the premises, but it will not be possible to *acquire* warrant for the conclusion by reasoning from the premises; rather one will need to warrant the conclusion *en route* to acquiring warrant for the premises. In essence, when transmission fails, the epistemic order – the order of warrant acquisition – inverts the deductive order.

I believe that both the Cartesian and Humean forms of sceptical paradox make a correct point about the epistemic architecture of the beliefs of ours that they target. We should agree that since the acquisition of, e.g., perceptual warrant does indeed require, as a matter of conceptual necessity, that one not be dreaming, or envatted, one is in position to claim such warrant in any particular case only in so far as one is in position to claim that no such uncongenial possibility obtains. And we should also agree that one is in position to claim warrant for, e.g., one's beliefs about the mental states of another on the basis of evidence concerning what they say and do only in so far as there is warrant to believe in the first place that there are other minds besides one own, whose states are broadly revealed in others' sayings and doings. So, to resist the sceptical conclusion, a case must be made that we *are* in position to discount the uncongenial scenarios of dreaming, envatment, and so on, and that we *do* have the requisite prior warrant for the general hinges, like "there are other minds" and "the world did not come into being five minutes ago, replete with apparent traces of a much more ancient history", on which the possibility of defeasible justification of specific beliefs about other minds and the past turns. But what is that case?

In "Imploding" (1991), I argued directly against the key premise in any form of Cartesian scepticism, viz. that one has no warrant to deny that one is in a situation of subjectively undetectable impairment of the relevant cognitive faculty. On the contrary, we must think of ourselves as warranted in believing that we are not now dreaming or envatted, that our memories are not

systematically impaired, and so on. The details of the argument and its ramifications are complex, but the basic idea was a simple one, turning on a version of the Cartesian paradox that, as it were inadvertently, provides a *reductio* of the conjunction of its own premises. These, recall, may be schematized like this:

*Premise (I):* There is no warrant for thinking one is not now cognitively impaired (in the relevant way).

*Premise (II):* One is warranted in accepting that [for any of a large class of beliefs, P, one has no warrant for accepting P unless one is not so cognitively impaired].

These entail (assuming closure) that one has no warrant for taking P as warranted, for any of the class of beliefs concerned. The central idea of "Imploding" was to turn this paradox against the faculties involved in engaging with the paradox itself, grouped together under the heading of "intellection". So we envisage a kind of state – *maundering* – in which all the subjectivity of lucid intellection is present but the real (internal causal) conditions for cognitive achievement by means of intellection are not met. The two resulting premises allow of exactly analogous support to that which they receive in all instances of the Cartesian paradox, and the conclusion duly follows that we have no warrant to regard as warranted any belief whose warrant would depend on intellection. One such belief, however, is given by the material in square brackets in Premise (II). So our conclusion implies that there is no warrant for Premise (II). Since this is false, and since Premise (II) is actually (one may argue) true, Premise (I) has to be false. The complexity of the paper is then occasioned by the attempt to generalize this finding to all versions of the paradox, including the classic case of the Dreaming argument itself.

So why did it seem otherwise? In particular, why is there any even *prima facie* plausibility in the suggestion that one has, in the end, no



warrant to discount the idea that one's present experience is but a lucid dream? There are various sceptical ploys to support this premise, but all rely on an equation of warrant with *evidence* and then make a case that the identification of evidence for thinking of one's present experience as broadly veridical, waking experience rather than a dream will rest on the presupposition that one really has the evidence and isn't just dreaming it up! So the point, in essentials, is that we cannot non-question-beggingly identify an *evidential warrant* to believe we are not dreaming.

This is the crucial point of contact with Humean scepticism. The Humean sceptical argument makes a case that we cannot attain an evidential warrant for a large "hinge" proposition – for instance, that there is an external material world – which, it contends, had better be warranted if we are to be justified in accepting the evidence for specific propositions within the "domain" of the hinge that we do. So what emerges is that both forms of paradox rely on the assumption that the only admissible kind of warrant either to dismiss the possibility that one might be dreaming right now, or to ground the belief that there is an external world, is evidential. This observation paves the way for the "unified strategy" of response to Cartesian and Humean scepticism broached in "Warrant for Nothing": both paradoxes can be resisted if it can be argued that rational warrants extend beyond – can consist in something other than – the provision of evidence. In particular, if it is possible to have *non-evidential* warrants for acceptance of "I am not dreaming now" and of "There is an external world", neither form of sceptical paradox can reach its depressing conclusion.

"Warrant for Nothing" pursues this strategy by making a case for several different kinds of non-evidential warrant, gathered under the catch-all term *entitlement*. The one most germane to the possibility of vindicating the rationality of believing that one is not now dreaming is *entitlement of cognitive project*. It is proposed that one has such an entitlement to accept, or presuppose, P if the following conditions are satisfied:

- (i) P is a *presupposition* of the cognitive project at hand – in the sense that to doubt P (in advance) would rationally commit one to doubting the significance or competence of the project;
- (ii) There is no extant sufficient reason to believe that P is false; and
- (iii) The attempt to justify P would involve further presuppositions in turn of no more secure a prior standing . . . and so on without limit; so that someone pursuing the relevant enquiry who accepted that there is nevertheless an onus to justify P would implicitly undertake a commitment to an infinite regress of justificatory projects, each concerned to vindicate the presuppositions of its predecessor.

When these conditions are met, to run the project in question, and to take its findings on board, rationally requires that one take it for granted that P, even absent any evidence in its favour. So if the project concerned is of a kind – for example, deciding what is so, or inferring – which is of the very essence of rational activity, then it is rational – so long as all three conditions are met – to have no doubt that P. I leave it to the reader to think through the suggestion that the proposition that one's cognitive faculties are functioning properly in conditions conducive to their doing so passes the test. If so, there is a rational entitlement to take it that one is not now dreaming.

It is another matter whether this notion of entitlement can extend to all the "hinge" propositions whose apparent evidential isolation is argued for and exploited in the various paradoxes. The suggestion of "Warrant for Nothing" was that this is not so: that other modes of entitlement will need to be explored if the unified strategy is to succeed.

Wittgenstein's response to scepticism of *On Certainty* is often read as naturalistic – as emphasizing the idea of the blamelessly "animal" in our habits of belief-formation, and there are certainly passages that support that reading. I like to see my own work in this field as responding to a different, rationalistic tendency in Wittgenstein's last notes. Those

notes were written in reaction to Moore's "Proof of an External World" and "Defence of Common Sense", and recently there has also been a refreshment of the discussion of Moore's contributions, with a very good debate developing between so-called "dogmatists" (Pryor, Peacocke), who view his notorious "Proof" as epistemically unexceptionable, though wanting – perhaps – when produced in certain specific dialectical settings, and so-called "sceptics", such as myself. This renewed interest in Moore's ideas is proving extremely rich in its philosophical implications, fostering consideration both of fundamental issues about the epistemic architecture of perceptual warrant, and about the conditions that valid reasoning, in general, has to satisfy in order to be cogent – to produce warrant for a conclusion. The existence and vivacity of these discussions, which my work has, I believe, been partly responsible for stimulating, is gratifying.<sup>3</sup>

## NOTES

- 1 Henceforward "Imploding".
- 2 Henceforward "Simple and Subtle".
- 3 It is a difficult exercise to provide an effective conspectus, at this length, of one's development and publications over more than a quarter of a century on a major, ramified set of philosophical issues. I am greatly indebted to Annalisa Coliva for permission to draw on her own detailed and skilful map of the progression I have outlined here.

## CRISPIN WRIGHT: PUBLICATIONS IN EPISTEMOLOGY 1983–2008

- "Keeping Track of Nozick," *Analysis* 43 (1983), 134–40.
- "Second Thoughts about Criteria," *Synthese* 58 (1984), 383–405.
- "Comment on Lowe," *Analysis* 44 (1984), 183–5.
- "Facts and Certainty," the Henriette Hertz Philosophical Lecture for the British Academy, December 1985, in *Proceedings of the British Academy* 71 (1985), 429–72.
- "Scepticism and Dreaming: Imploding the Demon," *Mind* 100 (1991), 87–116.

"On Putnam's Proof That We Are Not Brains-in-a-vat," *Proceedings of the Aristotelian Society* 92 (1991), 67–94.

Extended version of "On Putnam's Proof That We Are Not Brains-in-a-vat," in *Reading Putnam*, ed. P. Clark and R. Hale (Oxford: Blackwell, 1994).

"Human Nature?," critical study of John McDowell's *Mind and World*, in the *European Journal of Philosophy* 4 (1996), 235–54.

"McDowell's Oscillation," contribution to a book symposium on John McDowell's *Mind and World* in *Philosophy and Phenomenological Research* 58: 2 (1998), 395–402.

"Cogency and Question-begging: Some Reflections on McKinsey's Paradox and Putnam's Proof," *Philosophical Issues* 10 (2000), 140–63.

"Replies," in *Philosophical Issues* 10 (2000), 201–19.

Reprint of "Human Nature" with a new postscript in *Reading McDowell: On Mind and World*, ed. Nicholas Smith (London: Routledge, 2002), 140–73.

"(Anti-)sceptics, Simple and Subtle: G. E. Moore and John McDowell," in *Philosophy and Phenomenological Research* 65 (2002), 330–48.

"Some Reflections on the Acquisition of Warrant by Inference," in *New Essays On Semantic Externalism and Self-knowledge*, ed. S. Nuccetelli (Cambridge, MA: MIT Press/Bradford, 2003), 57–77.

"Wittgensteinian Certainties," in *Wittgenstein and Scepticism*, ed. D. McManus (London: Routledge, 2004), 22–55.

"Scepticism, Certainty, Moore and Wittgenstein," in *Wittgenstein's Lasting Significance*, ed. M. Kolbel and B. Weiss (London: Routledge, 2004), 226–46.

"Hinge Propositions and the Serenity Prayer," in *Knowledge and Belief*, Proceedings of the 26th International Wittgenstein Symposium, ed. W. Löffler and P. Weingartner (Vienna: Holder-Pickler-Tempsky, 2004), 287–306.

"Warrant for Nothing (and Foundations for Free)?," *Aristotelian Society Supplementary Volume* 78 (2004), 167–212.

"Intuition, Entitlement and the Epistemology of Basic Logical Laws," *Dialectica* 58 (2004), 155–75.

"Contextualism and Scepticism: Even-handedness, Factivity and Surreptitiously Raising Standards," *The Philosophical Quarterly* 55 (2005), 236–62.

"The Perils of Dogmatism," in *Themes from G. E. Moore: New Essays in Epistemology and Ethics*, ed. Susana Nuccetelli and Gary Seay (Oxford: Oxford University Press, 2007).

"New Age Relativism and Epistemic Possibility," in *Philosophical Issues* 17, *The Metaphysics of Epistemology* (2007).

Comment on McDowell "The Disjunctive Conception of Experience as Material for a Transcendental Argument," in *Disjunctivism: Perception, Action, Knowledge*, ed. A. Haddock and F. MacPherson (Oxford: Oxford University Press, 2008).

"Internal–External: Doxastic Norms and the Defusing of Skeptical Paradox," *The Journal of Philosophy* 9 (2008), 501–17.

## Linda Zagzebski

When Aristotle said that it is human nature to desire to know, he set down one of the most important constraints on the field of epistemology. I propose that if we can figure out why knowledge is desirable we shall not only go some distance towards understanding what it is, but we shall also discover some surprising things. One thing we shall discover is that a number of well-known theories of knowledge cannot explain what makes knowledge better than true belief. I call this "the value problem". Second, we shall discover that the common view that epistemic good is independent of moral good is largely an illusion. Not only are the epistemic goods intimately connected with other things we value, but also intellectual virtues are strongly tied to moral virtues. A mechanistic model of human belief acquisition is problematic because there is no truth-directed channel in the human mind that operates independently of those features of the human psyche that are evaluated morally – human emotions, choices, acts, and traits. A third thing we shall discover is that knowing is only one of the valuable epistemic relations we desire to have with the world. Valuable states such as understanding and wisdom have been neglected in the modern period, but were given much more attention in most of ancient and medieval philosophy. These are among the considerations that have led to what Wayne Riggs calls "the value turn" in epistemology. One of the things I have tried to do in my work is to influence epistemologists to make the value turn.

The natural desire for knowledge is due both to the delight we take in its object and to our fear of ignorance, but modern philosophers have been much more strongly driven by the latter than by the former. The modern quest for beliefs that can survive the skeptical attack led to a focus on individual belief states and an obsession with justification since justification is what is needed to defend a claim to know. It also led epistemologists to give most of their attention to the simplest and least ambitious candidates for knowledge: simple perceptual and memory beliefs and bits of information one picks up from others. The *a priori* has never been ignored for long, but even there the candidates that get the attention are the least controversial: beliefs in logic and mathematics. Few would dare suggest that there can be knowledge in metaphysics, religion, or aesthetics, and even moral knowledge is not often discussed in the context of a more general epistemological theory.

Suppose, however, that Aristotle is right when he says that even meager knowledge of celestial things is more pleasurable than all our knowledge of the world in which we live, just as a half glimpse of a person we love is more delightful than a clear and complete view of other things (*Parts of Animals* 644b32–5). That would suggest that an account of knowledge according to which we can know that there is a red wall in front of us but which ignores the epistemic status of answers to our deepest and most heartfelt questions is not likely to give us much comfort. Why would philosophers have devoted so much attention to knowledge in the history of philosophy if the most we can aim for is to avoid global ignorance?

In my work in epistemology I have aimed to develop a theory that includes the following features: (1) It explains what it is about knowledge that makes it better than true belief, worthy of the sustained attention it has received throughout the history of philosophy; (2) The account of knowledge avoids Gettier problems and related problems arising from situations in which the truth is reached accidentally; (3) It gives us guidance in our

epistemic behavior; (4) It respects the interaction between and among the cognitive, affective and conative aspects of the human being; (5) It generates an account of knowledge that applies to the more desirable objects of knowledge; that is, it has room for high-grade knowledge; (6) It recognizes the epistemic values of understanding, insight, and wisdom.

Of course, there are many other desiderata in an epistemological theory, but I mention the features on the above list because they were often missing from theories in the second half of the twentieth century. They are components of my version of virtue theory, a theory intended to be comprehensive enough to include a virtue epistemology and a virtue ethics within a single theory.

#### A. THE VALUE PROBLEM

One point of entry into the importance of value in epistemology is a puzzle I call “the value problem”. This is the problem of explaining what it is about knowledge that makes it more valuable than mere true belief.<sup>1</sup> I have argued that most forms of reliabilism have a particularly hard time handling this problem. According to standard reliabilist models, knowledge is true belief that is the output of reliable belief-forming processes or faculties. But the reliability of the source of a belief cannot explain the difference in value between knowledge and true belief because reliability *per se* has no value or disvalue. The way I have put it before, a reliable espresso-maker is good because espresso is good. A reliable water-dripping faucet is not good because dripping water is not good. The good of the product makes the reliability of the source that produces it good, but the reliability of the source does not then give the product an additional boost of value. The liquid in this cup is not improved by the fact that it comes from a reliable espresso-maker. If the espresso tastes good, it makes no difference if it comes from an unreliable machine. If the flower garden is beautiful, it makes no difference if it was planted by an unreliable

gardener. If the book is fascinating, it makes no difference if it was written by an unreliable author. If the belief is true, it makes no difference if it comes from an unreliable belief-producing source.

Now, suppose we succeed in identifying a value in the source of a true belief other than reliability or truth-conduciveness. That is still not sufficient to solve the value problem as long as we think of knowledge as the external product of the source. A cup of espresso is not made better by the fact that the machine that produces it is valuable, even when that value is independent of the value of good-tasting espresso. So the fact that the machine is properly functioning or is valuable in some other way is not sufficient to explain why the product of the machine is better than it would be if it were a product of a less valuable machine. The espresso analogy shows not only that a reliable cause does not confer value on its effect, but that there is a general problem in attributing value to an effect because of its causes even if the value of the cause is independent of the value of the effect.

It follows that, in the course of explaining what knowledge is, it is not enough to find another value besides true belief somewhere in the neighborhood. Rather, we must find another value in the right place. Knowing itself must include something good in addition to true belief. I think this means that we cannot solve the value problem if we persist in using the machine-product model of belief that is so common in epistemological discourse. Knowledge cannot be identified with true belief that is the output of a valuable cause, whether or not the cause has a value independent of the value of true belief.

It is not just externalist theories that have the value problem. Some theories that define knowledge as true justified belief analyze justification in such a way that its value is explained by its truth-conduciveness.<sup>2</sup> If truth-conduciveness is not sufficient to add value to true belief, then these theories also fail to solve to the value problem.

My conclusion at this point is that knowing is not an output of the agent, but is a feature of the agent. Knowing is an intellectual act

that gets epistemic value from the agent in the same way that overt acts get moral value from the agent. I do not mean to suggest that this is the only alternative to the machine-product model,<sup>3</sup> but the parallel with the moral evaluation of acts is illuminating. I suggest that in an act of knowing the agent gets to the truth because of the virtuous features of her belief-forming activity, in particular the motives out of which the act arises.<sup>4</sup> Just as we think that an overt act is morally better when it arises from morally virtuous motives, similarly an act of truly believing a proposition is epistemically better when it arises from epistemically virtuous motives.

A similar response to the value problem has been proposed by Wayne Riggs and John Greco, who propose that the extra value of knowing in addition to true belief is the state of affairs of the epistemic agent's getting credit for the truth that is acquired.<sup>5</sup> Likewise, Ernest Sosa says that, in a state of knowing, the truth is attributable to the agent as his or her own doing.<sup>6</sup> These three approaches are similar to mine, but notice that they solve the value problem only if we reject the machine-product model of knowledge.<sup>7</sup> For the same reason that the espresso in the cup is not made better by the fact that it is produced by a reliable espresso-maker or a properly functioning espresso-maker, it does not get any better if the machine gets credit for producing the espresso, or if the espresso is attributed to the machine as its own doing.

The agent-act model of knowing has other advantages besides handling the value problem. A number of philosophers have recently discussed the connection between knowledge, justification, or conscientious believing, and other things we care about. As applied to knowledge, this idea leads to one form of contextualism. I have argued that it is counter-intuitive to connect what we care about to the account of knowledge, but highly intuitive to connect it to conscientious or virtuous believing.<sup>8</sup> An even stronger motive for the agent-act model is the relationship between the moral permissibility/impermissibility of acting and the epistemic permissibility/

impermissibility of believing. I suggest that when something of moral importance is at stake when someone performs an act *S*, then if *S* is a case of acting on a belief *B* it is morally important that *B* be true. It is, therefore, morally impermissible for the agent to believe in a way that fails to respect the importance of the truth of *B*. That implies that the agent must believe out of certain motives. In particular, I suggest that the agent's motives must be such that they include a valuing of truth. In my account of virtue, a motive is the primary component of a virtue, and a good epistemic motive is the primary component of the intellectual virtues pertinent to the acquisition of belief. If the moral evaluation of acts depends upon the epistemic evaluation of beliefs, moral evaluation includes or presupposes an evaluation of the agent's epistemic motives, which in turn is connected with an evaluation of the agent's intellectual character – her intellectual virtues and vices. So one route into virtue epistemology is via the connection between the evaluation of acts and the evaluation of beliefs.<sup>9</sup> Another is the search for the source of the value of knowledge in addition to true belief. There are many others, and I shall look at some of them next.

## B. INTELLECTUAL VIRTUE AND EPISTEMIC VALUES

When I was working on *Virtues of the Mind* in the mid-1990s, I was struck by the fact that epistemologists were obsessed with disputes over justification, but the disputes seemed to be intractable, at least in part because it was not clear what "justified belief" means. "Justified" seems to mean something like "right", but to point out the parallel is merely to highlight the ambiguities since right in the sense of commendable differs from right in the sense of permissible (not wrong). There is also the distinction between right in the subjective sense, or what is right, given certain background beliefs (and perhaps motives) of the agent, and there is right in the objective sense, or what is right from the perspective of a being who is omniscient, and perhaps has



other features of an ideal observer. Once we get past these distinctions, there is then the problem that in ethics an account of a right act is connected with an account of the good, and the way that is done differs in different kinds of theory. So a right act can be defined in terms of a good outcome, a good motive, or a good natural end, or it can be more basic than good in any of these senses. It is very unlikely that disputes about rightness can be separated from disputes over the plausibility of the general theories in which the right and the good are embedded.

The same point applies to disputes over justification. These disputes cannot be separated from disputes over the plausibility of an epistemological theory that connects justified belief with epistemic good. Some theories of justification are modeled on deontological ethics, some on consequentialism. The model I propose in *Virtues of the Mind* is a virtue theory. There are stronger and weaker forms of virtue theory, but all of them make the traits of a good person central, and right acts derivative or less central. Similarly, virtue epistemology makes intellectual traits central and justified beliefs derivative or less central. I propose that the notion of justified belief that survives the shift to virtue theory is one that is not very interesting or important, nor is the notion of doing one's epistemic duty. Much more important is the notion of virtuously believing. Given that a virtue is a complex trait, a belief could be virtuous in one way but not another. So a belief could be virtuously motivated, it could lead to an epistemic good such as truth or understanding, it could imitate the behavior of virtuous persons in relevantly similar circumstances, and so on. A belief could have one of these features and not others, just as an act can have one of the parallel moral features and not others. I propose that a belief that is *good in every respect*, like an act that is good in every respect, has the following features:

- (1) It is virtuously motivated.
- (2) It imitates the behavior of virtuous persons in relevantly similar circumstances.
- (3) It reaches the truth because of features (1) and (2).

My position is that knowledge is belief that is good in every respect. (Notice that on this definition it is not necessary for knowledge that the knower have intellectual virtue since her virtuous motives and success in reaching epistemic goods may not be entrenched traits of character.) I also surmise that there are epistemic states other than beliefs that are good in every respect, but they lead to some epistemic good other than truth, such as understanding. I have argued that this account of knowledge both avoids Gettier problems and the value problem.

But doesn't this account of knowledge seem too stringent? As I mentioned above, some philosophers think of the paradigm cases of knowledge as states like truly believing that one is looking at a red wall, a state/belief that is usually automatic, mindless, lacking a motive, and hardly good in every respect. A motive for making states of this kind the best candidates for knowledge is empirical foundationalism, which persists in epistemology in spite of repeated attacks. And, even if knowing that I am looking at a red wall is not the *best* candidate for knowledge, it does seem to be a *good* candidate. Ernest Sosa distinguishes between animal knowledge and reflective knowledge in order to solve this kind of problem.<sup>10</sup> I have no argument against Sosa's distinction, but I doubt that it is a very important question whether the sensory beliefs that may (or may not) be foundational to high-grade knowledge count as knowledge themselves. Even on the foundationalist picture, there are states upon which knowing states are based that are not knowing states, but are sub-rational sensory states of the knowing subject. If sensory states support knowing states without being knowing states themselves, it seems to me that believing that what I sense has certain sensory properties can support knowing states without being a knowing state itself. So one motive for claiming that I know I see a red wall when I am looking at a red wall in normal circumstances is foundationalism, but this foundationalist picture (a) is not jeopardized if my state of believing I see a red wall is not a knowing state, and (b) may be false anyway.

But there is another reason why many philosophers think that believing I see a red wall typically is a knowing state, and that is that the belief normally can survive critical reflection and is sensitive to evidence (e.g. reasons for thinking there may be a red light shining on the wall). But a belief that satisfies these two conditions would also normally satisfy my criteria for knowledge, and conversely. In the typical situation, when I believe I see a red wall, I am doing what intellectually virtuous persons characteristically do in those conditions. Further, if our motives are revealed in what we *would* do in counterfactual circumstances, then as long as I would revise my belief if I had evidence that there is something suspicious about the circumstances, then even when the circumstances are not suspicious, my belief is properly motivated by love of truth, and the belief satisfies my conditions for knowledge. Perhaps beliefs of young children do not satisfy either my conditions or the conditions of sensitivity to evidence and survival of critical reflection, but I do not think that anything of importance hangs on the claim that the beliefs of young children constitute knowledge.

### C. EXEMPLARISM

In my most recent book, *Divine Motivation Theory*, I propose a virtue theory of ethics in which the first step in the construction of the theory is the same as the first step in moral learning: the identification of admirable persons, persons worth imitating. Exemplars of goodness are picked out by direct reference to persons identified through the emotion of admiration rather than through descriptions.<sup>11</sup> More recently, I have suggested that an application of exemplarism to epistemology has the potential to solve a host of problems, including the dispute over foundationalism, coherentism, and infinitism; the dispute between adherents of these theories and reliabilism; the dispute over contextualism and invariantism; the problem of the alignment of rationality with truth, and the problem of the clash of first-person and third-person

perspectives. All of these disputes are about something normative: the right way to structure a belief system, the right way to form beliefs in different contexts, why rationality is a good thing, and the meta-theoretical problem of the perspective that gives us the right answer to our set of questions. I suggest that the epistemological model of an Ideal Observer theory, or my own Ideal Agent theory, can be usefully adopted to answer these normative questions.

### NOTES

- 1 I have addressed the value problem in "From Reliabilism to Virtue Epistemology", "Intellectual Motivation and the Good of Truth", and "The Search for the Source of Epistemic Good".
- 2 Bonjour, 1985, pp. 7–8. DePaul insightfully discusses the problem of Bonjour and others in explaining the value of knowledge in chapter 2 of his 1993.
- 3 Another alternative is that knowledge is identified with the entire process culminating in the belief, and it gets value from the value in the process as well as the truth of the end product of the process. I have proposed that it would serve the purposes of Sosa's account of epistemic value to think of knowledge as an organic unity in the sense used by Brentano and Moore. That would permit the value of the whole to exceed the value of the sum of the parts. See "Epistemic Value Monism". Michael DePaul also discusses the possibility that knowledge is an organic unity in (2001), sect. VI.
- 4 *Virtues of the Mind*, pt III.
- 5 See Riggs, 1998; and Greco, 2003.
- 6 Sosa, 2003.
- 7 As far as I can tell, Greco and Riggs reject the machine-product model, but Sosa uses it repeatedly, for example in "The Place of Truth in Epistemology", the paper in which he proposes his way out of the value problem.
- 8 "The Ethics of Belief and the Primacy of What We Care About".
- 9 For another defense of this thesis, see Montmarquet, 1993.
- 10 See Sosa, "Knowledge, Animal and Reflective: A Reply to Michael Williams".
- 11 In "Ideal Observers and Ideal Agents in Epistemology", I apply direct reference to exemplars in order to solve some entrenched problems in epistemology.

## BIBLIOGRAPHY

- BonJour, L.: *The Structure of Empirical Knowledge* (Cambridge: Cambridge University Press, 1985).
- DePaul, M.: *Balance and Refinement* (New York: Routledge, 1993).
- Greco, J.: "Knowledge as Credit for True Belief," in *Intellectual Virtue: Perspectives from Ethics and Epistemology*, ed. L. De Paul and L. Zagzebski (Oxford: Clarendon Press, 2003).
- Montmarquet, J.: *Epistemic Virtue and Doxastic Responsibility* (Totowa, NJ: Rowman and Littlefield, 1993).
- Riggs, W.: "What Are the 'Chances' of Being Justified?," *The Monist* 81 (1988).
- Sosa, E.: "Knowledge, Animal and Reflective: A Reply to Michael Williams," *Proceedings of the Aristotelian Society*, supplementary volume 77 (2000), 113–30.
- Sosa, E.: "The Place of Truth in Epistemology," in *Intellectual Virtue: Perspectives from Ethics and Epistemology*, ed. L. De Paul and L. Zagzebski (Oxford: Clarendon Press, 2003).
- Zagzebski, L.: "From Reliabilism to Virtue Epistemology," in *Knowledge, Belief, and Character*, ed. G. Axtell (Oxford: Clarendon Press, 2003).
- Zagzebski, L.: "Intellectual Motivation and the Good of Truth," in *Intellectual Virtue: Perspectives from Ethics and Epistemology*, ed. L. De Paul and L. Zagzebski (Oxford: Clarendon Press, 2003).
- Zagzebski, L.: "The Search for the Source of Epistemic Good," *Metaphilosophy* 34 (2003), 12–28. Reprinted in *Moral and Epistemic Virtues*, ed. D. Pritchard and M. Brady (Oxford: Oxford University Press, 2003).
- Zagzebski, L.: "The Ethics of Belief and the Primacy of What We Care About," in *Immoral Believing*, special issue of *Philosophical Papers*, ed. W. Jones (2005).
- Zagzebski, L.: "Ideal Observers and Ideal Agents in Epistemology," in *Epistemology Futures*, ed. S. Hetherington (Oxford: Oxford University Press, 2005).
- Zagzebski, L.: "Epistemic Value Monism," in *Sosa and His Critics*, ed. J. Greco (Oxford: Blackwell, 2006).



Part III

Epistemology from A to Z





**a priori/a posteriori** The a priori/a posteriori distinction has been applied to a wide range of objects including concepts, propositions, truths, and knowledge. Our primary concern will be with the epistemic distinction between a priori and a posteriori knowledge. The most common way of articulating the concept is by reference to Immanuel Kant's claim that a priori knowledge is absolutely independent of all experience (*see* A PRIORI KNOWLEDGE in Part I):

- (APK) S knows a priori that p if and only if S's belief that p is justified a priori and the other conditions on knowledge are satisfied; and
- (APJ) S's belief that p is justified a priori if and only if S's justification for the belief that p does not depend on experience.

Kant's characterization is incomplete since it does not articulate the relevant senses of the key terms "depend" and "experience".

Contemporary analyses of a priori justification fall into two primary camps. Some regard (APJ) as insufficiently informative and favor a positive characterization of a priori justification:

- (APJ\*) S's belief that p is justified a priori if and only if S's belief that p is justified by source  $\phi$ .

Theorists in this camp typically invoke the concept of necessity in their descriptions of the source. Laurence Bonjour (1985) maintains that S's belief that p is justified a priori when S is able "to intuitively 'see' or apprehend that its truth is an invariant feature of all possible worlds". Alvin Plantinga (1993)

describes it as "finding yourself convinced that p is true and could not have been false". Since "seeing that p" and "finding yourself convinced that p" entail "believing that p", both accounts entail that if S's belief that p is justified a priori, then S believes that p is necessarily true. This condition has implausible consequences.

Consider a mathematician, S, who believes a theorem T on the basis of a generally accepted proof. S's belief that T is justified. Suppose that S lacks the concept of necessity and, as a consequence, does not believe that necessarily T. It is implausible to maintain that S's belief that T is not justified a priori merely because S lacks a concept that is not a constituent of T. But, even if S does possess the concept of necessity, there is a further problem. Suppose that S has some philosophical sophistication, but is in the grips of a false metaphysical theory, which holds that mathematical truths are not necessary. As a consequence, S refrains from believing that T is necessarily true. It is implausible to maintain that S's mathematical belief that T is not justified a priori merely because S has a false belief regarding a metaphysical issue.

Theorists in the second camp endorse (APJ), but there is controversy over the respect in which justification must be independent of experience. Some, such as Albert Casullo (2003), maintain that

- (I1) S's justification for the belief that p does not depend on experience just in case S's belief that p is justified by some non-experiential source.

Others, such as Philip Kitcher (1983), favor

- (I2) S's justification for the belief that p does not depend on experience just

in case S's belief that *p* is justified by some non-experiential source and cannot be defeated by experience.

(I2) faces a serious problem: it rules out the possibility that mathematical beliefs are justified both a priori and by experience.

Suppose that S's belief that *p*, where *p* is some mathematical proposition, is justified by some experiential source, such as counting collections of objects, reading textbooks, or consulting mathematicians. Such experiential justification for the belief that *p* is defeasible by an experientially justified belief that not-*p*. For example, if S's belief that *p* is justified by counting a collection of objects and arriving at a particular result, then it is possible that S recounts the collection and arrives at a different result. If S were to do so, S's original justification would be defeated by an experientially justified belief that not-*p*.

Alternatively, if S's belief that *p* is justified by a textbook (or mathematician) that states that *p*, then it is possible that S encounters a different textbook (or mathematician) that states that not-*p*. If S were to do so, S's original justification would be defeated by an experientially justified belief that not-*p*. Hence, if S's mathematical belief that *p* is justified by experience, then S's belief that not-*p* is also justifiable by experience. On the other hand, if S's belief that *p* is justified a priori, then, according to (I2), S's justification for the belief that *p* is not defeasible by experience. Since S's justified belief that not-*p* is a defeater for S's justified belief that *p*, it follows that if S's belief that *p* is justified a priori, then S's belief that not-*p* is not justifiable by experience. Therefore, S's belief that *p* cannot be justified both a priori and by experience. Since it is a substantive epistemological question whether a mathematical belief can be justified both a priori and by experience, any conception of a priori justification that settles it by fiat should be rejected.

Those who endorse (APJ) are faced with the task of articulating the relevant sense of "experience". Proponents of the a priori frequently maintain that there is a distinctive type of experience, often called "intuition" or

"intuitive apprehension", which is the source of a priori justification. Hence, there is a *broad* sense of "experience" on which a priori justification can depend. If the claim that intuition is a non-experiential source of justification is coherent, then there is another sense of "experience", the *narrow* sense, which requires articulation. There are two approaches to articulating this sense: providing a list of the relevant types of experiences, or providing a general characterization.

The enumerative approach faces three problems. The first is completing the list. Although there is general agreement that it should include the experiences associated with the five senses, there is disagreement regarding introspection, memory, and testimony. Second, even if the list is completed, it has little explanatory value; it does not explain why, for example, intuition is not on the list. Third, the resulting analysis rules out the possibility that there are (yet undiscovered) experiential sources of justification other than those on the list.

There are four general approaches. The first attempts to offer a general characterization in terms of phenomenological features. The problem here is that it is difficult to find phenomenological features common to the experiences associated with the five senses, let alone the more controversial sources. The second looks at the content of belief: experiential sources are those that justify beliefs about only the actual world. This approach rules out the possibility that some necessary truths are justified by experience. The third looks at the objects of belief: experiential sources are those that justify beliefs about only concrete objects. This approach rules out the possibility that some beliefs about abstract objects, such as mathematical beliefs, are justified by experience. The final approach looks at the relation between cognizer and objects of experience: experiential sources are those that stand in some causal relation to their objects. This approach rules out the possibility that the same belief can be justified both a priori and by experience.

Casullo (2003) proposes an alternative approach modeled on the semantics of

natural kind terms, such as “water”. Such terms are introduced into the language using local paradigms, prominent examples of the kind in question, and their extension is fixed by the underlying nature of those paradigms, which is discovered by scientific investigation. The alternative approach views “experience”, in the narrow sense, as a putative natural kind term whose reference is fixed by local paradigms: the cognitive processes associated with the five senses. The extension of “experience” is fixed by the underlying nature (if there be any) of the paradigms, which is discovered by the scientific study of human cognition rather than by a priori reflection.

## BIBLIOGRAPHY

- BonJour, Laurence: *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1985).  
 Casullo, Albert: *A Priori Justification* (New York: Oxford University Press, 2003).  
 Kitcher, Philip: *The Nature of Mathematical Knowledge* (New York: Oxford University Press, 1983).  
 Plantinga, Alvin: *Warrant and Proper Function* (New York: Oxford University Press, 1993).

ALBERT CASULLO

**abduction** Inductive reasoning tests hypotheses against experience: typically, we derive predictions from hypotheses and establish whether they are satisfied. An account of induction leaves unanswered two prior questions: How do we arrive at the hypotheses in the first place? And on what basis do we decide which hypotheses are worth testing? These questions concern the logic of discovery or, in Charles S. Peirce’s terminology, abduction. Many empiricist philosophers have denied that there is a logic (as opposed to a psychology) of discovery. Peirce, and followers such as N. R. Hanson, insisted that there is a logic of abduction.

The logic of abduction thus investigates the norms employed in deciding whether a hypothesis is worth testing at a given stage of inquiry, and the norms influencing how we

should retain the key insights of rejected theories in formulating their successors.

*See also* INDUCTION; PEIRCE.

## BIBLIOGRAPHY

- Hanson, N. R.: *Patterns of Discovery* (Cambridge: Cambridge University Press, 1958).  
 Peirce, C. S.: *Collected Papers*, vol. 7, ed. A. Burks (Cambridge, MA: Harvard University Press, 1958), pp. 89–164.

CHRISTOPHER HOOKWAY

**absurdity** An absurdity is any obviously, patently, or otherwise undeniably false proposition, such as  $0 = 1$  or, for some proposition  $p$ , the proposition  $p$  & *not*- $p$ . Absurdities play the most important role in *reductio ad absurdum* arguments conducted in classical logic. – One wants to demonstrate  $p$ . Assume that *not*- $p$ . Show that *not*- $p$  implies a false proposition  $A$ . Since any proposition that implies a false proposition is false, *not*- $p$  is false, so that *not*-*not*- $p$  is true, and *not*-*not*- $p$  is logically equivalent to  $p$ . That is, from the fact that *not*- $p$  implies  $A$ , one infers that  $p$ . Such a demonstration of  $p$  would be controversial if the falsehood of  $A$  were in doubt. Hence it would be best if  $A$  were patently or uncontroversially false, i.e. *absurd*.

## BIBLIOGRAPHY

- Kneale, W. and Kneale, M.: *The Development of Logic* (Oxford: Clarendon Press, 1962).

ROBERT S. TRAGESSER

**Academy (Plato)** Plato (c.429–347 BC) (*see* PLATO) founded his school, named after a district of Athens, between 387 and 367 BC. Studies centred on philosophy, mathematics and science. The conjecture that Plato’s *Republic* accurately describes its curriculum, however, seems wrong. It focused on Platonic thinking through the headships of Speusippus and Xenocrates (d. 314). It

was later dominated by the scepticisms of Arcesilaus (d. 241) and Carneades (d. 129). In the first century BC it was dominated by a Platonism that was assimilated to the views of Aristotle and Stoicism (see ARISTOTLE; STOIC EPISTEMOLOGY). Its subsequent history is unclear, and its activity apparently ceased with the closing of the pagan schools by Justinian in AD 529.

## BIBLIOGRAPHY

Zeller, E.: *Die Philosophie der Griechen in ihrer geschichtlichen Entwicklung*, 6th edn (Hildesheim, 1963), II.1.2 and III.1.1–2.

NICHOLAS P. WHITE

**act/object analysis** According to the act/object analysis of experience, every experience with content involves an object of experience to which the subject is related by an act of awareness (the event of experiencing that object). This is meant to apply not only to perceptions, which have material objects (whatever is perceived), but also to experiences like hallucinations and dream experiences, which do not. Such experiences none the less appear to represent *something*, and their objects are supposed to be whatever it is that they represent. Act/object theorists may differ on the nature of objects of experience, which have been treated as properties, Meinongian objects (which may not exist or have any form of being), and, more commonly, private mental entities with sensory qualities. (The term “sense-data” is now usually applied to the latter, but has also been used as a general term for objects of sense experiences, as in the work of G. E. Moore (see MOORE).) Act/object theorists may also differ on the relationship between objects of experience and objects of perception. In terms of representative realism (see REPRESENTATIVE REALISM), objects of perception (of which we are “indirectly aware”) are always distinct from objects of experience (of which we are “directly aware”). Meinongians, however, may simply treat objects of perception as *existing* objects of experience.

See also ADVERBIAL THEORY; DIRECT REALISM; EXPERIENCE; REPRESENTATIVE REALISM; SENSE-DATA.

## BIBLIOGRAPHY

Ayer, A. J.: *The Foundations of Empirical Knowledge* (London: Macmillan, 1940).  
 Jackson, F.: *Perception: A Representative Theory* (Cambridge: Cambridge University Press, 1977).  
 Moore, G. E.: *Philosophical Studies* (London: Kegan Paul, 1922).  
 Perkins, M.: *Sensing the World* (Indianapolis, IN: Hackett, 1983).

MICHAEL PENDLEBURY

**adverbial theory** In its best-known form the adverbial theory of experience proposes that the grammatical object of a statement attributing an experience to someone be analysed as an adverb. For example.

(1) Rod is experiencing a pink square

is rewritten as

Rod is experiencing (pink square)-ly.

This is presented as an alternative to the act/object analysis (see ACT/OBJECT ANALYSIS), according to which the truth of a statement like (1) requires the existence of an object of experience corresponding to its grammatical object. A commitment to the explicit adverbialization of statements of experience is not, however, essential to adverbialism. The core of the theory consists, rather, in the denial of objects of experience (as opposed to objects of perception) coupled with the view that the role of the grammatical object in a statement of experience is to characterize more fully the sort of experience which is being attributed to the subject. The claim, then, is that the grammatical object is functioning as a modifier, and, in particular, as a modifier of a verb. If this is so, it is perhaps appropriate to regard it as a special kind of adverb at the semantic level.

See also EXPERIENCE; REPRESENTATIVE REALISM.



## BIBLIOGRAPHY

- Chisholm, R. M.: *Perceiving: A Philosophical Study* (Ithaca, NY: Cornell University Press, 1957).
- Clark, R.: "Sensing, perceiving, thinking," in *Essays on the Philosophy of Roderick M. Chisholm*, ed. E. Sosa, *Grazer Philosophische Studien* (1981), 273–95.
- Cornman, J.: *Perception, Common Sense, and Science* (New Haven, CT: Yale University Press, 1975).
- Ducasse, C. J.: *Nature, Mind, and Death* (La Salle, IL: Open Court, 1951).

MICHAEL PENDLEBURY

**agnosticism** There are two forms of agnosticism: weak and strong. Consider *theism*: the proposition there is such a person as God – an almighty and all-knowing and wholly good creator of the world. A weak agnostic is someone who believes neither that there is such a person nor that there is not. In this respect an agnostic is to be contrasted with an atheist, who holds that there is no God, and a theist, who holds that there is. So the theist affirms theism; the atheist denies it; and the agnostic withholds it, having no view as to whether or not this proposition is true. A strong agnostic adds that it isn't *possible* to know or have a justified belief about the truth of theism, so that no one else should have a view on it either.

ALVIN PLANTINGA

**Alston, William P. (1921–2009)** Alston has contributed to epistemology on many topics: the analysis of justification and knowledge, the foundationalism–coherentism and internalism–externalism controversies, epistemic principles, religious epistemology, perception, and numerous others. He is known both for his own positions and for incisively developing distinctions now important in the literature.

His early papers on foundationalism (see FOUNDATIONALISM) distinguished *levels* of justification and thereby showed that, even if one does not hold, or is not directly (non-inferentially) justified in holding, the second-order belief that one is justified in believing *p*, one may be directly justified in believing *p*. Since

foundationalists as such need not require second-order justification regarding basic beliefs, this distinction undercuts much criticism previously considered decisive against foundationalism. In distinguishing many grades of privileged access, Alston also showed that neither foundationalists nor other epistemologists must regard infallibility (see INFALLIBILITY) or some version of Cartesian certainty as the only alternatives to coherentism (see COHERENTISM) in accounting for the varieties of justification. Regarding justification in general, Alston draws a contrast between deontological and "strong position" notions. Roughly, the former treat justification as fulfilment of epistemic duty, the latter as a matter of being in a good position with respect to the truth of *p*, e.g. being able simply to see that *p* is true. He argues that much of the literature on justification fails to take account of this distinction, and he shows how the distinction can explain major disagreements. For instance, the former conception goes well with an internalist view, since one has introspective access to grounds of one's obligations, such as a memory of having promised to do something or a conviction that lying is wrong; the latter suggests an at least partly externalist view on justification, since one does not in general have such access to the reliability of one's position with respect to discerning the truth of *p* (see VIRTUE EPISTEMOLOGY in Part I; RELIGIOUS BELIEF; EXTERNALISM/INTERNALISM).

Alston's own account of justification is a distinctive blend of internalism and externalism: if I justifiedly believe *p*, I must both have an appropriate access to my grounds and, by virtue of them, be in a good position vis-à-vis the truth of *p*. This condition normally also holds for knowledge, but Alston has rigorously argued that under special conditions knowledge is possible without justification and thereby on grounds to which one lacks access. For knowledge, as for justification, perception is a paradigmatic source. In both cases, moreover, first-order success is possible without second-order success; e.g. one can know that *p* without knowing that one does, or even that one's source, say perception, is reliable. But can we know or justifiedly believe

perception is reliable? In discussing epistemic circularity, Alston argues that, although a kind of circularity is implicit in plausible attempts to show the reliability of perception, it does not prevent one's justifiedly believing, or even knowing, that perception is reliable. To this extent, at least, scepticism is answerable.

In recent work, Alston has pursued at least three major epistemological projects. He has defended the theory of appearing as an account of perception. He has developed a doxastic practice approach in meta-epistemology, arguing, along lines suggested by the work of Reid (*see* REID), that justification is rooted in a certain kind of social practice. And, using these resources and many others inside and outside epistemology, he has built an account of the possibility of perception of, and thereby justified beliefs about, God.

His account of perception is extended in more recent work and is supported by his realist theory of truth. His latest epistemology book (2005) sets out a series of epistemic desiderata – including truth-conduciveness, permissibility, and accessibility – and argues that the traditional concern with explicating justification should be replaced by interconnected treatments of these diverse elements. The book is a theoretically sophisticated, constructive challenge to the standard view that there is a concept of justification of central importance in epistemology.

#### WRITINGS

"Varieties of privileged access," *American Philosophical Quarterly* 8 (1971), 223–41.

"Two types of foundationalism," *Journal of Philosophy* 73 (1976), 165–85.

"Concepts of epistemic justification," *The Monist* 68 (1985), 57–89.

"An internalist externalism," *Synthese* 74 (1988), 265–83.

*Epistemic Justification* (Ithaca, NY: Cornell University Press, 1989). (This contains all of the above.)

*Perceiving God* (Ithaca, NY: Cornell University Press, 1991).

*The Reliability of Sense-perception* (Ithaca, NY: Cornell University Press, 1993).

*A Realist Conception of Truth* (Ithaca, NY: Cornell University Press, 1996).

*Illocutionary Acts and Sentence Meaning* (Ithaca, NY: Cornell University Press, 2000).

*Beyond "Justification": Dimensions of Epistemic Evaluation* (Ithaca, NY: Cornell University Press, 2005).

ROBERT AUDI

**analyticity** The true story of analyticity is surprising in many ways. Contrary to received opinion, it was the empiricist Locke rather than the rationalist Kant who had the better informal account of this type of a priori proposition. Frege and Carnap, represented as analyticity's best friends in this century, did as much to undermine it as its worst enemies. Quine and Putnam, represented as having refuted the analytic/synthetic distinction, not only did no such thing, but, in fact, contributed significantly to undoing the damage done by Frege and Carnap. Finally, the epistemological significance of the distinction is nothing like what it is commonly taken to be.

Locke's account of analytic propositions (*see* LOCKE) was, for its time, everything that a succinct account of analyticity should be (Locke, 1924, pp. 306–8). He distinguishes two kinds of analytic propositions, identity propositions in which "we affirm the said term of itself", e.g. "Roses are roses", and predicative propositions in which "a part of the complex idea is predicated of the name of the whole", e.g. "Roses are flowers" (pp. 306–7). Locke calls such sentences "trifling" because a speaker who use them "trifles with words". A synthetic sentence, in contrast, such as a mathematical theorem, states "a real truth and conveys with it instructive real knowledge" (pp. 307–8). Correspondingly, Locke distinguishes two kinds of "necessary consequences", analytic entailments where validity depends on the literal containment of the conclusion in the premiss and synthetic entailments where it does not. (Locke did not originate this concept-containment notion of analyticity. It is discussed by Arnauld and Nicole, and it is safe to say it has been around for a very long time (cf. Arnauld, 1964, pp. 59–65).)

Kant's account of analyticity (*see* KANT), which received opinion tells us is the consummate formulation of this notion in modern

philosophy, is actually a step backwards. What is valid in his account is not novel, and what is novel is not valid. Kant repeats Locke's account of concept-containment analyticity, but introduces certain alien features, the most important being his characterization of analytic propositions as propositions whose denials are logical contradictions (Kant, 1783, p. 14). This characterization suggests that analytic propositions based on Locke's part-whole relation or Kant's explicative copula are a species of logical truth. But the containment of the predicate concept in the subject concept in sentences like "Bachelors are unmarried" is a different relation from the containment of the consequent in the antecedent in sentences like "If John is a bachelor, then John is a bachelor or Mary read Kant's *Critique*". The former is *literal* containment whereas the latter is, in general, not. Talk of the "containment" of the consequent of a logical truth in the antecedent in cases like our example is only metaphorical, a way of saying "logically derivable".

Kant's conflation of concept containment with logical containment caused him to overlook the issue of whether logical truths are synthetic a priori and the problem of how he can say mathematical truths are synthetic a priori when they cannot be denied without contradiction. Historically, the conflation set the stage for the disappearance of the Lockean notion. Frege, whom received opinion portrays as second only to Kant among the champions of analyticity, and Carnap, whom it portrays as just behind Frege, were jointly responsible for the disappearance of concept-containment analyticity.

Frege (see FREGE) was clear about the difference between concept containment and logical containment, expressing it as like the difference between the containment of "beams in a house" and the containment of a "plant in the seed" (Frege, 1953, p. 101). But he found the former, as Kant formulated it, defective in three ways: it explains analyticity in psychological terms; it does not cover all cases of analytic propositions; and, perhaps most important for Frege's logicism, its notion of containment is "unfruitful" as a definitional mechanism in logic and mathematics

(Frege, 1953, pp. 100–1). In an invidious comparison between the two notions of containment, Frege observes that with logical containment "we are not simply taking out of the box again what we have just put into it" (ibid., p. 101). To overcome these shortcomings, Frege defines analytic propositions as consequences of laws of logic plus definitions, consequences that "cannot be inspected in advance" (ibid., pp. 4, 100–1). This definition makes logical containment the basic notion. Analyticity becomes a special case of logical truth, and, even in this special case, the definitions employ the power of definition in logic and mathematics rather than mere concept combination.

Carnap (see CARNAP), attempting to overcome what he saw as a shortcoming in Frege's account of analyticity, took the remaining step necessary to do away explicitly with Lockean–Kantian analyticity. As Carnap saw things, it was a shortcoming of Frege's explication that it seems to suggest that definitional relations underlying analytic propositions can be extra-logical in some sense, say, in resting on linguistic synonymy. To Carnap, this represented a failure to achieve a uniform formal treatment of analytic propositions and left us with a dubious distinction between logical and extra-logical vocabulary. Hence, he eliminated the reference to definitions in Frege's explication of analyticity by introducing "meaning postulates", e.g. statements such as " $(\forall x) (x \text{ is a bachelor} \rightarrow x \text{ is unmarried})$ " (see Carnap, 1965, pp. 222–9). Like standard logical postulates on which they were modelled, meaning postulates express nothing more than constraints on the admissible models with respect to which sentences and deductions are evaluated for truth and validity. Thus, despite their name, meaning postulates have no more to do with meaning than any other statement expressing a necessary truth. In defining analytic propositions as consequences of (an expanded set of) logical laws, Carnap explicitly removed the one place in Frege's explication where there might be room for concept containment, and with it, the last trace of Locke's distinction between semantic and other "necessary consequences".

Quine, the staunchest critic of analyticity of our time (*see* QUINE), performed an invaluable service on its behalf – albeit one that has gone almost completely unappreciated. Quine made two devastating criticisms of Carnap's meaning postulate approach which expose it as both irrelevant and vacuous. It is irrelevant because, in using particular words of a language, meaning postulates fail to explicate analyticity for sentences and languages generally, that is, they do not define it for variable "S" and "L" (Quine, 1953, pp. 33–4). It is vacuous because, although meaning postulates tell us what sentences are to count as analytic, they do not tell us what it is for them to be analytic (*ibid.*, p. 33).

Received opinion has it that Quine did much more than refute the analytic/synthetic distinction as Carnap tried to draw it. Received opinion has it that Quine demonstrated there is no distinction, however anyone might try to draw it. But this, too, is incorrect. To argue for this stronger conclusion, Quine had to show that there is no way to draw the distinction outside logic, in particular, in linguistics. In the absence of a particular theory in linguistics corresponding to Carnap's, Quine's argument had to take an entirely different form. Some inherent feature of linguistics had to be exploited in showing that no theory in this science can deliver the distinction. But the feature Quine chose was a principle of operationalist methodology characteristic of the school of Bloomfieldian linguistics. Quine succeeds in showing that meaning cannot be made objective sense of in linguistics if making sense of a linguistic concept requires, as that school claims, operationally defining it in terms of substitution procedures which employ only concepts unrelated to that linguistic concept. But Chomsky's revolution in linguistics replaced the Bloomfieldian taxonomic model of grammars with the hypothetico-deductive model of generative linguistics, and, as a consequence, such operational definition was removed as the standard for concepts in linguistics. The standard of theoretical definition which replaced it was far more liberal, allowing the members of a family of linguistic concepts to be defined with respect to one another within a set of

axioms which state their systematic interconnections – the entire system being judged by whether its consequences are confirmed by the linguistic facts. Quine's argument does not even address theories of meaning based on this hypothetico-deductive model (Katz, 1988b, pp. 227–52; Katz, 1990, pp. 199–202).

Putnam, the other staunch critic of analyticity (*see* PUTNAM), performed a service on behalf of analyticity fully on a par with, and complementary to, Quine's. Whereas Quine refuted Carnap's formalization of Frege's conception of analyticity, Putnam refuted this very conception itself. Putnam put an end to the entire attempt, initiated by Frege and completed by Carnap, to construe analyticity as a logical concept (Putnam, 1962, pp. 647–58; 1970, pp. 189–201; 1975a, pp. 131–93).

However, as with Quine, received opinion has it that Putnam did much more. Putnam is credited with having devised science fiction cases, from the robot cat case to the twin earth case, that are counter examples to the traditional theory of meaning. Again, received opinion is incorrect. These cases are only counter examples to Frege's version of the traditional theory of meaning. Frege's version claims *both* (1) that sense determines reference, and (2) that there are instances of analyticity, say, typified by "cats are animals", and of synonymy, say typified by "water" in English and "water" in twin earth English. Given (1) and (2), what we call "cats" could not be non-animals and what we call "water" could not differ from what the twin earthers call "water". But, as Putnam's cases show, what we call "cats" could be Martian robots and what they call "water" could be something other than H<sub>2</sub>O. Hence, the cases are counter examples to Frege's version of the theory.

Putnam himself takes these examples to refute the traditional theory of meaning *per se* because he thinks other versions must also subscribe to both (1) and (2). He was mistaken in the case of (1). Frege's theory entails (1) because it defines the sense of an expression as the mode of determination of its referent (Frege, 1952, pp. 56–78). But sense does not have to be defined this way, or in any way that entails (1). It can be defined as (D).

- (D) Sense is that aspect of the grammatical structure of expressions and sentences responsible for their having sense properties and relations like meaningfulness, ambiguity, antonymy, synonymy, redundancy, analyticity, and analytic entailment (Katz, 1972; 1990, pp. 216–24).

(Note that this use of sense properties and relations is no more circular than the use of logical properties and relations to define logical form, for example, as that aspect of grammatical structure of sentences on which their logical implications depend.)

(D) makes senses internal to the grammar of a language and reference an external matter of language use – typically involving extra-linguistic beliefs. Therefore, (D) cuts the strong connection between sense and reference expressed in (1), so that there is no inference from the modal fact that “cat” refers to robots to the conclusion that “Cats are animals” is not analytic. Likewise, there is no inference from “water” referring to different substances on earth and twin earth to the conclusion that our word and theirs are not synonymous. Putnam’s science fiction cases do not apply to a version of the traditional theory of meaning based on (D).

The success of Putnam’s and Quine’s criticisms in application to Frege’s and Carnap’s theory of meaning together with their failure in application to a theory in linguistics based on (D) creates the option of overcoming the shortcomings of the Lockean–Kantian notion of analyticity without switching to a logical notion. This option was explored in the 1960s and 1970s in the course of developing a theory of meaning modelled on the hypothetico-deductive paradigm for grammars introduced in the Chomskyan revolution (Katz, 1972).

This theory automatically avoids Frege’s criticism of the psychological formulation of Kant’s definition because, as an explication of a grammatical notion within linguistics, it is stated as a formal account of the structure of expressions and sentences. The theory also avoids Frege’s criticism that concept-containment analyticity is not “fruitful” enough to encompass truths of logic and

mathematics. The criticism rests on the dubious assumption, part of Frege’s logicism, that analyticity *should* encompass them. (Benacerraf, 1981, p. 25). But in linguistics where the only concern is the scientific truth about natural languages, there is no basis for insisting that concept-containment analyticity encompass truths of logic and mathematics. Moreover, since we are seeking the scientific truth about trifling propositions in natural language, we will eschew relations from logic and mathematics that are too fruitful for the description of such propositions. This is not to deny that we want a notion of necessary truth that goes beyond the trifling, but only to deny that *that* notion is the notion of analyticity in natural language.

The remaining Fregean criticism points to a genuine incompleteness of the traditional account of analyticity. There are analytic relational sentences, for example, “Jane walks with those with whom she herself strolls”, “Jack kills those he himself has murdered”, etc., and analytic entailments with existential conclusions, for example, “I think”, therefore, “I exist”. The containment in these sentences is just as literal as that in analytic subject–predicate sentences like “Bachelors are unmarried”. I will now show how a theory of meaning constructed as a hypothetico-deductive systematization of sense as defined in (D) overcomes the incompleteness of the traditional account in the case of such relational sentences. (For a treatment of the existential sentences, see Katz, 1988a.)

Such a theory of meaning makes the principal concern of semantics the explanation of sense properties and relations like synonymy, antonymy, redundancy, analyticity, ambiguity, etc. Furthermore, it makes grammatical structure, specifically, sense structure, the basis for explaining them. This leads directly to the discovery of a new level of grammatical structure, and this, in turn, makes possible a proper definition of analyticity. To see this, consider two simple examples. It is a semantic fact that “male bachelor” is redundant and that “spinster” is synonymous with “woman who never married”. In the case of the redundancy, we have to explain the fact that the sense of the modifier “male” is



already contained in the sense of its head "bachelor". In the case of the synonymy, we have to explain the fact that the sense of "spinster" is identical to the sense of "woman who never married" (compositionally formed from the senses of "woman", "never" and "married"). But in so far as such facts concern relations involving the components of the senses of "bachelor" and "spinster" and in so far as these words are syntactic simples, there must be a level of grammatical structure at which syntactic simples are semantically complex. This, in brief, is the route by which we arrive a level of *decompositional semantic structure* that is the locus of sense structures masked by syntactically simple words.

Discovery of this new level of grammatical structure was followed by attempts to represent the structure of the senses found there. Without going into the details of sense representation, it is clear that, once we have the notion of decompositional representation, we can see how to generalize Locke's and Kant's informal, subject-predicate account of analyticity to cover relational analytic sentences. Let a simple sentence *S* consist of a *n*-place predicate *P* with terms  $T_1, \dots, T_n$  occupying its argument places. Then:

- (A) *S* is analytic in case, first, *S* has a term  $T_i$  which consists of an *m*-placed predicate *Q* ( $m > n$  or  $m = n$ ) with terms occupying its argument places, and second, *P* is contained in *Q* and, for each term  $T_j$  of  $T_1, \dots, T_{i-1}, T_{i+1}, \dots, T_n$ ,  $T_j$  is contained in the term of *Q* which occupies the argument place in *Q* corresponding to the argument place occupied by  $T_j$  in *P* (Katz, 1972).

To see how (A) works, suppose that "stroll" in "Jane walks with those with whom she herself strolls" is decompositionally represented as having the same sense as "walk idly and in a leisurely way". The sentence is analytic by (A) because the predicate "stroll" (the sense of "stroll") contains the predicate "walk" (the sense of "walk") and the term "Jane" (the sense of "Jane" associated with the predicate "walk") is contained in the term "Jane" (the sense of "she herself" associated with

the predicate "stroll"). The containment in the case of the other terms is automatic.

The fact that (A) itself makes no reference to logical operators or logical laws indicates that analyticity for subject-predicate sentences can be extended to simple relational sentences without treating analytic sentences as instances of logical truths. Further, the source of the incompleteness is no longer explained, as Frege explained it, as the absence of "fruitful" logical apparatus, but is now explained as mistakenly treating what is only a special case of analyticity as if it were the general case. The inclusion of the predicate in the subject is the special case (where  $n = 1$ ) of the general case of the inclusion of an *n*-place predicate (and its terms) in one of its terms. Note that the defects Quine complained of in connection with Carnap's meaning postulate explication are absent in (A). (A) contains no words from a natural language. It explicitly uses variable "*S*" and variable "*L*" because it is a definition in linguistic theory. Moreover, (A) tells us what the property is in virtue of which a sentence is analytic, namely, redundant predication, that is, the predication structure of an analytic sentence is already found in the content of its term structure.

Received opinion has been anti-Lockean in holding that necessary consequences in logic and language belong to one and the same species. This seems wrong because the property of redundant predication provides a non-logical explanation of why true statements made in the literal use of analytic sentences are necessarily true. Since the property ensures that the objects of the predication in the use of an analytic sentence are chosen on the basis of the features to be predicated of them, the truth conditions of the statement are automatically satisfied once its terms take on reference. The difference between such a linguistic source of necessity and the logical and mathematical sources vindicates Locke's distinction between two kinds of "necessary consequences".

Received opinion concerning analyticity contains another mistake. This is the idea that analyticity is inimical to science. In part, the idea developed as a reaction to certain

dubious uses of analyticity such as Frege's attempt to establish logicism (see LOGICISM) and Schlick's, Ayer's and other logical positivists' attempts (see LOGICAL POSITIVISM) to deflate claims to metaphysical knowledge by showing that alleged synthetic a priori truths are merely empty analytic truths (Schlick, 1949, p. 285; Ayer, 1946, pp. 71–87). In part, it developed as also a response to a number of cases where alleged analytic, and hence, necessary, truths, e.g. the law of excluded middle, had subsequently been taken as open to revision. Such cases convinced philosophers like Quine and Putnam that the analytic/synthetic distinction is an obstacle to scientific progress.

The problem if there is one is not analyticity in the concept-containment sense, but the conflation of it with analyticity in the logical sense. This made it seem as if there is a single concept of analyticity which can serve as the grounds for a wide range of a priori truths. But, just as there are two analytic/synthetic distinctions, so there are two concepts of concept (see CONCEPTS). The narrow Lockean/Kantian distinction is based on a narrow notion of concept on which concepts are senses of expressions in the language. The broad Fregean/Carnapian distinction is based on a broad notion of concept on which concepts are conceptions – often scientific ones – about the nature of the referent(s) of expressions (Katz, 1972, pp. 450–2; and, curiously, Putnam, 1981, p. 207). Conflation of these two notions of concept produced the illusion of a single concept with the content of philosophical, logical and mathematical conceptions but with the status of linguistic concepts. This encouraged philosophers to think that they were in possession of concepts with the content to express substantive philosophical claims, e.g. such as Frege's, Schlick's, Ayer's, etc., and with a status that trivializes the task of justifying them by requiring only linguistic grounds for the a priori propositions in question.

Thus, there is no need to reject the analytic/synthetic distinction *in toto* to prevent analyticity from being put to dubious uses. All that is necessary is to keep the original, narrow distinction from being broadened.

This insures that propositions expressing the content of broad concepts cannot receive the easier justification appropriate to narrow ones. Accordingly, in so far as the wholesale rejection of the analytic/synthetic distinction was based on a concern about dubious philosophy, particularly the possibility of blocking scientific progress, it threw out the baby with the bath water.

Finally, there is an important epistemological implication of separating the broad and narrow notions of analyticity. Frege and Carnap took the broad notion of analyticity to provide foundations for necessity and a priority, and, hence, for some form of rationalism, and nearly all rationalistically inclined analytic philosophers followed them in this. Thus, when Quine dispatched the Frege–Carnap position on analyticity, it was widely believed that necessity, a priority, and rationalism had also been dispatched, and, as a consequence, that Quine had ushered in an “empiricism without dogmas” and naturalized epistemology (see NATURALIZED EPISTEMOLOGY). But given there is still a notion of analyticity which enables us to pose the problem of how necessary, synthetic a priori knowledge is possible (moreover, one whose narrowness makes logical and mathematical knowledge part of the problem), Quine did not undercut the foundations of rationalism. Hence, a serious reappraisal of the new empiricism and naturalized epistemology is, to say the least, very much in order (Katz, 1990).

*See also* A PRIORI/A POSTERIORI; INTUITION AND DEDUCTION; KANT; LOCKE; MATHEMATICAL KNOWLEDGE; PHILOSOPHICAL KNOWLEDGE.

#### BIBLIOGRAPHY

- Arnauld, A.: *The Art of Thinking* (Indianapolis, IN: Bobbs-Merrill, 1964).
- Ayer, A. J.: *Language, Truth, and Logic* (London: Gollancz, 1946).
- Benacerraf, P.: “Frege: The Last Logician,” in *Midwest Studies in Philosophy* 6 (Minneapolis, MN: University of Minnesota Press, 1981).
- Carnap, R.: “Meaning Postulates,” in his *Meaning and Necessity*, 2nd edn (Chicago, IL: University of Chicago Press, 1965).

- Frege, G.: "On Sense and Reference," in *Translations from the Philosophical Writings of Gottlob Frege*, trans. P. T. Geach and M. Black (Oxford: Blackwell, 1952).
- Frege, G.: *Foundations of Arithmetic* (Oxford: Blackwell, 1953).
- Kant, I.: *Critique of Pure Reason* (1781), trans. N. Kemp Smith (London: Macmillan, 1964).
- Kant, I.: *Prolegomena to Any Future Metaphysic* (1783), trans. L. W. Beck (Indianapolis, IN: Bobbs-Merrill, 1950).
- Katz, J. J.: *Semantic Theory* (New York: Harper & Row, 1972).
- Katz, J. J.: *Cogitations* (New York: Oxford University Press, 1988a).
- Katz, J. J.: "The Refutation of Indeterminacy," *Journal of Philosophy* 85 (1988b), 227–52.
- Katz, J. J.: *The Metaphysics of Meaning* (Cambridge, MA: MIT Press, 1990).
- Locke, J.: *An Essay Concerning Human Understanding* (Oxford: Clarendon Press, 1924).
- Putnam, H.: "It Ain't Necessarily So," *Journal of Philosophy* 59 (1962), 658–71.
- Putnam, H.: "The Meaning of 'Meaning,'" in *Language, Mind, and Knowledge*, Minnesota Studies in the Philosophy of Science (Minneapolis, MN: University of Minnesota Press, 1975a).
- Putnam, H.: "Is Semantics Possible?," (1970); reprinted in his *Mind, Language, Reality* (Cambridge: Cambridge University Press, 1975b), 139–52.
- Putnam, H.: *Reason, Truth, and History* (Cambridge: Cambridge University Press, 1981).
- Quine, W. V.: "Two Dogmas of Empiricism," in his *From a Logical Point of View* (Cambridge, MA: Harvard University Press, 1953).
- Schlick, M.: "Is There a Factual A Priori?," in *Readings in Philosophical Analysis*, ed. H. Feigl and W. Sellars (New York: Appleton-Century-Crofts, 1949).

JERROLD J. KATZ

**anamnesis** "Recollection", or anamnesis, has several roles in Plato's epistemology (see PLATO). In the *Meno* (80–6) it is invoked to explain the behaviour of an uneducated boy who answers a geometrical problem that he has never heard. At the same time it is used to solve a paradox about inquiry and learning. In the *Phaedo* it is said to explain our possession of concepts, construed as knowledge of Forms, which we supposedly could not have gained from experience. Recollection also

appears in the *Phaedrus*, but is notably absent from important presentations of Plato's epistemological views in the *Republic* and other works.

#### BIBLIOGRAPHY

- Gosling, J. C. B.: *Plato* (London: Routledge, 1973), ch. 16.
- Gulley, N.: *Plato's Theory of Knowledge* (London: Methuen, 1962), ch. 1.
- White, N. P.: *Plato on Knowledge and Reality* (Indianapolis, IN: Hackett, 1976), chs 2–3.

NICHOLAS P. WHITE

**antinomy** An antinomy occurs when we are able to argue for, or demonstrate, both a proposition and its contradictory (see PRINCIPLE OF CONTRADICTION), but where we cannot now fault either demonstration. We would eventually hope to be able "to solve the antinomy" by managing, through careful thinking and analysis, eventually to fault either or both demonstrations.

Many paradoxes are an easy source of antinomies. For example, Zeno gave some famous let us say logical-cum-mathematical arguments which might be interpreted as demonstrating that motion is impossible. But our eyes as it were demonstrate motion (exhibit moving things) all the time. Where did Zeno go wrong? Where do our eyes go wrong? If we cannot readily answer at least one of these questions, then we are in antinomy. In the *Critique of Pure Reason*, Kant (see KANT) gave demonstrations of the same kind – in the Zeno example they were obviously not the same kind – of both, e.g. that the world has a beginning in time and space, and that the world has no beginning in time or space. He argues that both demonstrations are at fault because they proceed on the basis of 'pure reason' unconditioned by sense experience.

#### BIBLIOGRAPHY

- Beck, L. W.: "Antinomy of Pure Reason," in *Dictionary of the History of Ideas* ed. P. P. Wiener (New York: Charles Scribner's Sons, 1973), vol. 1.

Kant, I.: *Critique of Pure Reason*, trans. N. Kemp Smith (London: Macmillan, 1964).

Kneale, W. and Kneale, M.: *The Development of Logic* (Oxford: Clarendon Press, 1962).

ROBERT S. TRAGESSER

**apodeictic** A proposition *p* is apodeictic when it is demonstrable in a sense entailing not only that *p* is true, but also that it is not possibly false. For Aristotle, an apodeictically true proposition was one inferred by formal-logical syllogism from incontrovertibly true premisses (*Posterior Analytics*, 1.71b–72c). Sometimes “apodeictic” is used loosely to mean that a proposition is recognized to be beyond dispute; and sometimes it is taken to mean that it *must* be true (without any reference to demonstration).

ROBERT S. TRAGESSER

**aporia** Any difficult problem that arises when we are trying to extend our knowledge of a matter, and that threatens seriously to impede our further progress, is called an aporia, especially when there seem to be equally strong arguments for and against any solution. An antinomy is an especially nasty aporia (see ANTINOMY).

#### BIBLIOGRAPHY

Peters, F. E.: *Greek Philosophical Terms* (London: University of London, 1967).

ROBERT S. TRAGESSER

**apperception** This is Leibnitz’s term (see LEIBNITZ) for inner awareness or self-consciousness, in contrast with “perception” or outer awareness. He held, in opposition to Descartes (see DESCARTES), that adult humans can have experiences of which they are unaware; experiences which may affect what they do, but which are not brought to self-consciousness. Indeed there are creatures, such as animals and babies, which completely lack the ability to reflect on their experiences, and to become aware of them as

experiences of theirs. The unity of a subject’s experience, which stems from his capacity to recognize all his experiences as his, was dubbed by Kant the transcendental unity of apperception (see KANT). This apprehension of unity is transcendental, rather than empirical, because it is presupposed in experience and cannot be derived from it. Kant used the need for this unity as the basis of his attempted refutation of scepticism about the external world. He argued that my experiences could only be united in one self-consciousness if at least some of them were experiences of a law-governed world of objects in space. Outer experience is thus a necessary condition of inner awareness.

See also INTROSPECTION.

#### BIBLIOGRAPHY

Kant, I.: *Critique of Pure Reason*, trans. N. Kemp Smith (London: Macmillan, 1964).

Leibniz, G. W.: *New Essays on Human Understanding* (1704), trans. P. Remnant and J. Bennett (Cambridge: Cambridge University Press, 1981), esp. pp. 53–7.

DAVID MCNAUGHTON

**Aquinas, Thomas (1225–74)** Theologian and philosopher, born near Naples. Aquinas’s fundamental epistemic category is that of cognition (*cognitio*). He endorses the Aristotelian view that the soul is potentially all things and holds that cognition is its actually becoming a given thing or, as he sometimes puts it, its being assimilated to that thing in a certain way (*Summa theologiae* Ia.12.4, 17.3, 76.2.AD4, 84.2.AD2). His account of this basic notion of cognition involves both a metaphysical account of the two relata in the relation of cognitive assimilation – the human soul and the objects of human cognition – and a psychological theory identifying the sorts of powers the soul must possess and the processes it must engage in if cognitive assimilation of this sort is to be possible.

According to Aquinas’s metaphysics, the reality to which human beings are assimilated

in cognition is made up of basic elements (particular substances and accidents) that are joined together in various ways to form complexes (accidents inhering in particular substances). He holds that the soul must possess cognitive powers capable of rendering it isomorphic with an external reality of this sort. The intellective soul's activity of understanding (*intellectus*) allows human beings both to grasp, via sense perception, the natures of substances and accidents and to link them together into complexes (subject–predicate propositions). But cognition is not restricted to the sort of intake of information made possible by sense perception and understanding; human beings are also able, by virtue of a distinct activity of discursive thought (*ratio, ratiocinatio*), to acquire new cognition of things by drawing inferences from things already cognized. Aquinas's strictly epistemological views are to be found within this broad metaphysically and psychologically oriented account of cognition (*Commentary on Aristotle's Posterior Analytics* Prologue; *Summa theologiae* Ia.75–86; *Disputed Questions on Truth* I, X).

Aquinas's most detailed epistemological reflections occur in the context of his discussion of the propositional attitude *scientia*, which he conceives of as the paradigm of knowledge. To have *scientia* with respect to a given thing is to have complete and certain cognition of its truth; that is, to hold a given proposition on grounds that guarantee its truth in a certain way. Following Aristotle, Aquinas holds that grounds of this sort are provided only by demonstrative syllogisms, and so he maintains that the objects of *scientia* are propositions one holds on the basis of demonstrative syllogisms. To have *scientia* with respect to some proposition *p*, then, is to have a particular sort of inferential justification for *p* (*Commentary on Aristotle's Posterior Analytics* I.4).

Now Aquinas holds that because the sort of justification essential to *scientia* is inferential, it is also derivative: *scientia* acquires its positive epistemic status from the premisses of the demonstrative syllogism and the nature of the syllogistic inference. Hence, he holds a principle of inferential justification according

to which one is justified in holding the conclusion of some demonstration only if one is justified in holding the demonstration's premisses. The premisses that ground *scientia* are not only logically but also epistemically prior to the conclusion (*Commentary on Aristotle's Posterior Analytics* I.6).

Aquinas argues that our justification for holding the premisses of demonstrative syllogisms cannot in every case be inferential: some propositions must have their positive epistemic status not by virtue of an inference (*per demonstrationem*) but non-inferentially, by virtue of themselves (*per se*) (*Commentary on Aristotle's Posterior Analytics* I.4, 7). Propositions that are known by virtue of themselves (*per se nota*) are Aquinas's epistemic first principles, the foundations of his foundationalist account of *scientia*. He offers two sorts of argument for his epistemological foundationalism (see FOUNDATIONALISM). The first proceeds by attacking rival accounts of justification, concluding that inferential justification is possible only if there is non-inferential justification. If one holds that all justification is inferential and if a person is inferentially justified in holding some proposition only if he is justified in holding the premisses of the relevant inference, then one is committed to an infinite regress of justification (see INFINITE REGRESS ARGUMENT). If the regress is linear, then there can be no justification since one cannot possess an infinite number of distinct inferences. But if one tries to avoid this sceptical result by maintaining that the regress of justification circles back on itself in such a way that an inferentially justified conclusion appears as part of the (proximate or remote) justification for itself, then one is committed to absurdities such as that one and the same proposition can be at once epistemically both prior and posterior to some other proposition and that some proposition can be epistemically prior to itself. Aquinas concludes that if inferential justification is to be possible, there must be non-inferential justification (*Commentary on Aristotle's Posterior Analytics* I.7–8).

His second argument for foundationalism rests on his positive characterization of the nature of non-inferential justification. He



holds that certain propositions (immediate propositions) are knowable *per se* by virtue of the fact that their predicates belong to the definition (*ratio*) of their subjects (*Commentary on Aristotle's Posterior Analytics* I.5). We are non-inferentially justified in holding propositions of this sort because when we are aware that a proposition's predicate belongs to the definition of its subject, we are directly aware of the proposition's necessary truth and cannot be mistaken about it (*Commentary on Aristotle's Posterior Analytics* I.7, 19, 20, 44; *Summa theologiae* Ia.17.3.AD2). To be aware that a given proposition is of this sort we must conceive its subject and predicate, which requires us to have attained explicit grasp of the real natures referred to by the subject and predicate terms (*Commentary on Aristotle's Posterior Analytics* I.2, 4, II.8). For example, when we have grasped the real nature *human being* (the real definition of which is *rational animal*), we cannot help but be aware of the necessary truth of the proposition "A human being is an animal". Aquinas holds, however, that attaining explicit grasp of the real natures of things can be difficult, and so he holds that not all propositions in which the predicate belongs to the definition of the subject are actually known to be of that sort. The subjects and predicates of certain purely formal, a priori propositions (e.g. those of logic and mathematics) are more easily cognized by human beings (because of their relative independence from matter), and so most people will recognize first principles of this sort as such and be non-inferentially justified in holding them. The subjects and predicates of a posteriori propositions (e.g. those of natural science), however, are accessible only with difficulty, and hence objective first principles of this sort may not be recognized as such. Aquinas holds that each of us has experience of being directly acquainted with the necessary truth of at least some first principles and that this provides us with sufficient reason for thinking that there is non-inferential justification (*Commentary on Aristotle's Posterior Analytics* I.4, 25, 41; *Commentary on Boethius's De Trinitate* 5).

This account of non-inferential justification requires him to give an account of our

cognitive relations to universal real natures (such as *human being*), the elements out of which complex (propositional) knowledge is built. He thinks of his account as resolving an ancient epistemological puzzle. The puzzle is how human beings, whose senses provide access to a world of irreducibly particular corporeal objects, can have cognition of universals. Aquinas rejects Platonist and Neoplatonist solutions that postulate the possibility of some sort of extrasensory contact with independent, immaterial universals. He not only rejects the existence of universals of this sort, but also holds that because human beings are by nature unified corporeal substances whose natural form of access to the world is through the bodily senses, all human cognition arises from sense perception. Aquinas's solution to the epistemological puzzle is a theory of intellective abstraction: cognition of universals, like all human cognition, originates from sense perception, and so from the external world of material particulars; but human beings possess a cognitive capacity (in particular what he calls an active intellect), which acts on sensory data to produce intelligible universals. We cognize the universal real natures that constitute the subjects and predicates of epistemic first principles when we possess actually intelligible species or forms abstracted by this mechanism from the material conditions that render them merely potentially intelligible (*Commentary on Aristotle's Posterior Analytics* II.20; *Summa theologiae* Ia.79, 84–6; *Disputed Questions on Truth* X.6).

According to Aquinas, then, we have paradigmatic knowledge when we hold a first principle by virtue of seeing that its predicate belongs to the definition of its subject (i.e. when we possess understanding – *intellectus* – of it) or hold a proposition on the basis of a demonstrative syllogism the premisses of which we hold in that way (i.e. have *scientia* with respect to it). He recognizes, however, that these conditions restrict paradigmatic knowledge to a very narrow range, and he allows that there is knowledge other than paradigmatic knowledge. First, following the ancient Greek distinction between demonstrative and dialectical reasoning, he allows that dialectical

(*probabile, persuasoria*) reasoning can provide epistemic justification. Reasoning of this sort is distinguished by virtue of its producing conclusions that are not certain but merely probable. So-called probable arguments rely on premisses that are not necessary and certain but possess some positive epistemic status (propositions held by most people, on good authority, inductive grounds, etc.) and make use of broadly inductive argument forms (enumerative induction, analogy, probabilistic argument forms, etc.). Justificatory grounds of this sort give rise not to *scientia* but to opinion (*opinio*) or belief (*fides*), and Aquinas holds that one can be justified in holding propositions one holds in this way (*Commentary on Aristotle's Posterior Analytics* Prologue, I.44; *Commentary on Boethius's De Trinitate* 2.1.AD5; *Summa theologiae* IIaIIae.2.1, 2.9.AD3; *Summa contra gentiles* I.9).

Second, Aquinas extends the strict account of demonstrative reasoning to take account of the particular condition of human epistemic subjects. Because human beings' epistemic access to the world is through the bodily senses, propositions about sensible objects are in many cases epistemically prior for us both in the sense that we acquire them first and in the sense that we find them easiest to accept as true (*Commentary on Aristotle's Posterior Analytics* I.4, 42). Consequently, he allows that propositions about sensible objects can function as epistemic first principles grounding what is for us (though not unqualifiedly) *scientia*. Demonstrations the ultimate premisses of which are not paradigmatic but only qualified first principles of this sort are not paradigmatic, explanatory demonstrations (*demonstrations propter quid*) but non-paradigmatic, merely factual demonstrations (*demonstrationes quia*). They establish the truth of the conclusion but fail to provide for it the certain and complete justification that would be provided by a demonstration deriving from necessary truths about the universal real natures of things (*Commentary on Aristotle's Posterior Analytics* I.23; *Summa theologiae* Ia.1–2). Since Aquinas holds that it is difficult for us to grasp the real natures of corporeal objects, he claims that the sort of knowledge we can have of the natural world

via natural science is for the most part non-paradigmatic *scientia* that relies on merely factual demonstrations the ultimate premisses of which are propositions about the accidental properties of sensible corporeal objects (*Commentary on Aristotle's Posterior Analytics* I.4). Similarly, since we cannot grasp God's essence, our knowledge of divine matters via natural theology is non-paradigmatic *scientia* deriving from propositions about sensible objects (*Commentary on Boethius's De Trinitate* 2; *Summa theologiae* Ia.1; *Summa contra gentiles* I.6–9).

#### WRITINGS

- Commentary on Aristotle's Posterior Analytics*, trans. F. R. Larcher (Albany, NY: Magi Books, 1970).
- Disputed Questions on Truth*, trans. R. W. Mulligan, J. V. McGlynn and R. W. Schmidt, 3 vols (Chicago, IL: Henry Regnery, 1952–4).
- Division and Methods of the Sciences* (Questions 5–6 of *Commentary on Boethius's De Trinitate*), trans. A. Maurer (Toronto: Pontifical Institute of Mediaeval Studies, 1986).
- Faith, Reason, and Theology* (Questions 1–4 of *Commentary on Boethius's De Trinitate*), trans. A. Maurer (Toronto: Pontifical Institute of Mediaeval Studies, 1987).
- Summa contra gentiles*, trans. A. C. Pegis, J. F. Anderson, V. J. Bourke and C. J. O'Neil, 5 vols (Notre Dame, IN: University of Notre Dame Press, 1975).
- Summa theologiae*, trans. Fathers of the English Dominican Province, 5 vols (Westminster: Christian Classics, 1981).

#### BIBLIOGRAPHY

- Aertsen, J.: *Nature and Creature: Thomas Aquinas's Way of Thought* (Leiden: Brill, 1988).
- Grabmann, M.: *Die theologische Erkenntnis- und Einleitungslehre des heiligen Thomas von Aquin* (Freiburg: Paulusverlag, 1948).
- Kretzmann, N.: "Infallibility, Error, and Ignorance," in *Aristotle and His Medieval Interpreters*, ed. M. Tweedale (supplementary volume to *Canadian Journal of Philosophy*, 1992).
- Lonergan, B.: *Verbum: Word and Idea in Aquinas* (Notre Dame, IN: University of Notre Dame Press, 1967).

MacDonald, S.: "Aquinas's Epistemology," in *Aquinas*, ed. N. Kretzmann and E. Stump (Cambridge: Cambridge University Press, 1992).

SCOTT MACDONALD

**argument** A group of statements, some of which purportedly provide support for another. The statements which purportedly provide the support are the *premises* while the statement purportedly supported is the *conclusion*. Arguments are typically divided into two categories depending on the *degree* of support they purportedly provide. *Deductive* arguments purportedly provide *conclusive* support for their conclusions while *inductive* arguments purportedly provide only *probable* support. Some, but not all, arguments succeed in providing support for their conclusions. Successful deductive arguments are valid while successful inductive arguments are strong. An argument is *valid* just in case if all its premisses are true then its conclusion *must* be true. An argument is *strong* just in case if all its premisses are true its conclusion is *only probably* true. Deductive logic provides methods for ascertaining whether or not an argument is valid whereas inductive logic provides methods for ascertaining the degree of support the premisses of an argument confer on its conclusion.

See also INFORMAL FALLACIES; INDUCTION; PROOF.

#### BIBLIOGRAPHY

- Haack, S.: *Philosophy of Logics* (Cambridge: Cambridge University Press, 1978).  
 Jeffrey, R.: *Formal Logic: Its Scope and Limits*, 2nd edn (New York: McGraw-Hill, 1981).  
 Kahane, H.: *Logic and Philosophy*, 6th edn (Belmont, CA: Wadsworth, 1990).  
 Salmon, W.: *Logic*, 2nd edn (Englewood Cliffs, NJ: Prentice-Hall, 1973).  
 Skyrms, B.: *Choice and Chance*, 3rd edn (Belmont, CA: Wadsworth, 1986).

ALBERT CASULLO

**argument from analogy** This argument is intended to establish our right to believe in

the existence and nature of other minds (*see* OTHER MINDS). It admits that it is *possible* that the objects we call persons are, other than ourselves, mindless automata, but claims that we none the less have sufficient reason for supposing this not to be the case. There is more evidence that they are not mindless automata than that they are.

The classic statement of the argument comes from J. S. Mill. He writes:

I am conscious in myself of a series of facts connected by a uniform sequence, of which the beginning is modifications of my body, the middle is feelings, the end is outward demeanour. In the case of other human beings, I have the evidence of my senses for the first and last links of the series, but not for the intermediate link. I find, however, that the sequence between the first and last is as regular and constant in those other cases as it is in mine. In my own case I know that the first link produces the last through the intermediate link, and could not produce it without. Experience, therefore, obliges me to conclude that there must be an intermediate link; which must either be the same in others as in myself, or a different one; . . . by supposing the link to be of the same nature . . . I conform to the legitimate rules of experimental enquiry.

(1867, pp. 237–8)

As an inductive argument this is very weak, because it is condemned to arguing from a single case. But to this we might reply that none the less we have more evidence that there are other minds than that there are not.

The real criticism of the argument is due to Wittgenstein (1953) (*see* WITTGENSTEIN). It is that the argument assumes that we at least *understand* the claim that there are subjects of experience other than ourselves, who enjoy experiences which are like ours but not ours; it only asks what reason we have to suppose that claim true. But if the argument does indeed express the ground of our right to believe in the existence of others, it is impossible to explain how we are able to achieve that understanding. So if there is a place for the argument from analogy, the problem of other minds – the real, hard problem, which is how we acquire a conception of another mind – is insoluble. The argument is either redundant or worse.

See also OTHER MINDS; PRIVATE LANGUAGE ARGUMENT; SOLIPSISM; WITTGENSTEIN.

# BIBLIOGRAPHY

- Malcolm, N.: "Knowledge of Other Minds," *Journal of Philosophy* 55 (1958), 969–78; reprinted in his *Knowledge and Certainty* (Englewood Cliffs, NJ: Prentice-Hall, 1963).
- Mill, J. S.: *An Examination of Sir William Hamilton's Philosophy*, 3rd edn (London: Longmans, 1867).
- Wittgenstein, L.: *Philosophical Investigations* (Oxford: Blackwell, 1953).

JONATHAN DANCY

**argument from illusion** This is usually intended to establish that certain familiar facts about illusion disprove the theory of perception called naïve or direct realism (see DIRECT REALISM). There are, however, many different versions of the argument which must be distinguished carefully. Some of these distinctions centre on the content of the premisses (the nature of the appeal to illusion); others centre on the interpretation of the conclusion (the kind of direct realism under attack). Let us begin by distinguishing importantly different versions of direct realism which one might take to be vulnerable to familiar facts about the possibility of perceptual illusion.

## DIRECT REALISM

A crude statement of direct realism might go as follows. In perception, we sometimes directly perceive physical objects and their properties; we do not always perceive physical objects by perceiving something *else*, e.g. a sense-datum (see SENSE-DATA). There are, however, difficulties with this formulation of the view. For one thing a great many philosophers who are *not* direct realists would admit that it is a mistake to describe people as actually *perceiving* something other than a physical object. In particular, such philosophers might admit, we should never say that we *perceive* sense-data. To talk that way would be to suppose that we should model our

understanding of our relationship to sense-data on our understanding of the ordinary use of perceptual verbs as they describe our relation to the physical world, and that is the last thing paradigm sense-datum theorists should want. At least many of the philosophers who objected to direct realism would prefer to express what they were objecting to in terms of a technical (and philosophically controversial) concept such as *acquaintance*. Using such a notion we could define direct realism this way: In *veridical* experience we are directly acquainted with parts (e.g. surfaces) or constituents of physical objects. A less cautious version of the view might drop the reference to veridical experience and claim simply that in all experience we are directly acquainted with parts or constituents of physical objects (see KNOWLEDGE BY ACQUAINTANCE/BY DESCRIPTION).

Because one can interpret the relation of acquaintance or awareness as one that is not *epistemic*, i.e. not a kind of propositional knowledge, it is important to distinguish the above views read as *ontological* theses from a view one might call *epistemological direct realism*: In perception we are, on at least some occasions, non-inferentially justified in believing a proposition asserting the existence of a physical object (see DIRECT REALISM).

What relevance does illusion have for these two forms of direct realism?

## ILLUSION, HALLUCINATION AND DIRECT REALISM

The fundamental premiss of the argument from illusion seems to be the thesis that things can appear to be other than they are. Thus, for example, a straight stick when immersed in water looks bent; a penny when viewed from a certain perspective looks elliptical; something that is yellow when placed under red fluorescent light looks red. In all of these cases, one version of the argument goes, it is implausible to maintain that what we are directly acquainted with is the real nature of the object in question. Indeed, it is hard to see how we can be said to be aware of the real physical object at all. In the above illusions the things we were aware of actually



were bent, elliptical and red, respectively. But, by hypothesis, the real physical objects lacked these properties. Thus we were not aware of the real physical objects.

So far, if the argument is relevant to any of the direct realisms distinguished above, it seems relevant only to the claim that in *all* sense experience we are directly acquainted with parts or constituents of physical objects. After all, even if in illusion we are not acquainted with physical objects, their surfaces, or their constituents, why should we conclude anything about the nature of our relation to the physical world in *veridical* experience?

We are supposed to discover the answer to this question by noticing the similarities between illusory experience and veridical experience and by reflecting on what makes illusion possible at all. Illusion can occur because the nature of the illusory experience is determined, not just by the nature of the object perceived, but also by other conditions, both external and internal. But all of our sensations are subject to these causal influences and it would be gratuitous and arbitrary to select from the indefinitely many and subtly different perceptual experiences some special ones as those which get us in touch with the “real” nature of the physical world. Red fluorescent lights affect the way things look, but so does sunlight. Water reflects light, but so does air. We have no unmediated access to the external world.

Still, why should we conclude that we are aware of something other than a physical object in experience? Why should we not conclude that to be aware of a physical object just is to be appeared to by that object in a certain way? (See *ADVERBIAL THEORY*.) There are indefinitely many ways in which objects appear to us but why not construe all of those ways of appearing as simply different kinds of direct appearing *relations* holding between us and the physical world? There is no need to infer that there is anything other than our yellow object under those red lights. We need only describe that yellow object as standing in the relation of appearing red to us.

At this point, it might be profitable to move from considering the possibility of illusion to considering the possibility of hallucination.

Instead of comparing paradigmatic veridical perception to illusion, let us compare it to complete hallucination. For any experience or sequence of experiences we take to be veridical, we can imagine qualitatively indistinguishable experiences occurring as part of an hallucination. For those who like their philosophical arguments spiced with a touch of science, we can imagine that our brains were surreptitiously removed in the night, and unbeknown to us are being stimulated by a neurophysiologist so as to produce the very sensations that we would normally associate with a trip to the Grand Canyon. Now let us ask what we are aware of in this complete hallucination. Obviously we are not aware of physical objects, their surfaces, or their constituents. Nor can we even construe the experience as one of an object’s appearing to us in a certain way. It is after all a complete hallucination and the objects we take to exist before us are simply not there. But if we compare hallucinatory experience with the qualitatively indistinguishable veridical experiences, should we not conclude that it would be *ad hoc* to suppose that in veridical experience we are aware of something radically different from what we are aware of in hallucinatory experience? Again, it might help to reflect on our belief that the *immediate* cause of hallucinatory experience and veridical experience might be the very same brain event, and it is surely implausible to suppose that the effects of this same cause are radically different – acquaintance with physical objects in the case of veridical experience; something else in the case of hallucinatory experience.

This version of the argument from hallucination would seem to address straightforwardly the ontological versions of direct realism. The argument is supposed to convince us that the ontological analysis of sensation in both veridical and hallucinatory experience should give us the same results, but in the hallucinatory case there is no plausible physical object, constituent of a physical object, or surface of a physical object with which we can identify experience. If we add one additional premiss we would also get an argument against epistemological direct realism. That premiss is that in a vivid hallucinatory



experience we might have precisely the same justification for believing (falsely) what we do about the physical world as we do in the analogous, phenomenologically indistinguishable, veridical experience. But our justification for believing that there is a table before us in the course of a vivid hallucination of a table is surely not non-inferential in character. It certainly isn't if non-inferential justification is supposed to consist in some unproblematic access to the fact that makes true our belief – by hypothesis the table doesn't exist. But if the justification that hallucinatory experience gives us is the same as the justification we get from the parallel veridical experience, then we should not describe a veridical experience as giving us non-inferential justification for believing in the existence of physical objects. In both cases we should say that we believe what we do about the physical world on the basis of what we know directly about the character of our experience.

#### CRITICISMS

In this brief space, I can only sketch some of the objections that might be raised against arguments from illusion and hallucination. Let us begin with a criticism that accepts most of the presuppositions of the arguments. Even if the possibility of hallucination establishes that in some experience we are not acquainted with constituents of physical objects, it is not clear that it establishes that we are never acquainted with a constituent of physical objects. Suppose, for example, that we decide that in both veridical and hallucinatory experience we are acquainted with sense-data. At least some philosophers have tried to identify physical objects with "bundles" of actual and possible sense-data (see PHENOMENALISM). Hallucinatory experience on this view is non-veridical precisely because the sense-data one is acquainted with in hallucination do not bear the appropriate relations to other actual and possible sense-data. But if such a view were plausible one could agree that one is acquainted with the same kind of thing in veridical and non-veridical experience but

insist that there is still a sense in which in veridical experience one is acquainted with constituents of a physical object.

A different sort of objection to the argument from illusion or hallucination concerns its use in drawing conclusions we have not stressed in the above discussion. I mention this objection to underscore an important feature of the argument. At least some philosophers (Hume, for example) have stressed the rejection of direct realism on the road to an argument for general scepticism with respect to the physical world (see SCEPTICISM). Once one abandons epistemological direct realism, one has an uphill battle indicating how one can legitimately make the inference from sensation to physical objects. But a philosopher who appeals to the *existence* of illusion and hallucination to develop an argument for scepticism can be accused of having an epistemically self-defeating argument. One could justifiably infer sceptical conclusions from the existence of illusion and hallucination only if one justifiably believed that such experiences exist, but if one is justified in believing that illusion exists one must be justified in believing at least some facts about the physical world (for example, that straight sticks look bent in water). The key point to stress in replying to such arguments is that strictly speaking the philosophers in question need only appeal to the *possibility* of vivid illusion and hallucination. Although it would have been psychologically more difficult to come up with arguments from illusion and hallucination if we did not believe that we actually had such experiences, I take it that most philosophers would argue that the possibility of such experiences is enough to establish difficulties with direct realism. Indeed, if one looks carefully at the argument from hallucination discussed earlier, one sees that it nowhere makes any claims about actual cases of hallucinatory experience.

Another reply to the attack on epistemological direct realism focuses on the implausibility of claiming that there is any process of *inference* involved in our beliefs about the world around us. Even if it is possible to give a phenomenological description of the subjective character of sensation, it requires

a special sort of skill that most people lack. Our perceptual beliefs about the physical world are surely direct at least in the sense that they are unmediated by any sort of conscious inference from premisses describing something other than a physical object. The appropriate reply to this objection, however, is simply to acknowledge the relevant phenomenological fact and point out that from the perspective of an epistemologist it is beside the point. In attacking epistemological direct realism, the philosopher is attacking a claim about the nature of our justification for believing propositions about the physical world. Such a philosopher need make no comment at all about the causal genesis of such beliefs.

I have mentioned that proponents of the argument from illusion and hallucination have often intended it to establish the existence of sense-data, and many philosophers have attacked the so-called sense-datum inference presupposed in some statements of the argument. When the stick looked bent, the penny looked elliptical and the yellow object looked red, the sense-datum theorist wanted to infer that there was something bent, elliptical and red, respectively. But such an inference is surely suspect. In general we do not infer that because something appears to have a certain property, there is something that has that property. When I say that Jones looks like a doctor, I surely wouldn't want anyone to infer that there must actually be someone there who is a doctor. In assessing this objection it will be important to distinguish different uses of words like "appears" and "looks". At least sometimes to say that something looks F is only to express a tentative belief that it is F and the sense-datum inference from an F 'appearance' in this sense to an actual F would be hopeless. However, it also seems that we use the "appears"/"looks" terminology to describe the phenomenological character of our experience and the inference might be more plausible when the terms are used this way. Still, it does seem to me that the arguments from illusion and hallucination will not by themselves constitute strong evidence for a sense-datum theory. Even if one concludes that there is something common to both the hallucination of a red thing and a veridical

visual experience of a red thing, one need not describe that common constituent as awareness of something red. The adverbial theorist would prefer to construe the common experiential state as "being appeared to redly", a technical description intended only to convey the idea that the state in question need not be analysed as relational in character. Those who opt for an adverbial theory of sensation (see ADVERBIAL THEORY) need to make good the claim that their artificial adverbs can be given a sense that is not parasitic upon an understanding of the adjectives transformed into adverbs. Still other philosophers might try to reduce the common element in veridical and non-veridical experience to some kind of intentional state, more like belief or judgement. The idea here is that the only thing common to the two experiences is the fact that in both I spontaneously take there to be present an object of a certain kind. The above objections can be stated within the general framework presupposed by proponents of the arguments from illusion and hallucination. A great many contemporary philosophers are, however, uncomfortable with the intelligibility of the concepts needed to even make sense of the theories attacked. Thus at least some who object to the argument from illusion do so not because they defend direct realism. Rather they think there is something confused about all this talk of direct awareness or acquaintance. Contemporary externalists, for example, usually insist that we understand epistemic concepts by appeal to nomological connections (see EXTERNALISM/INTERNALISM). On such a view the closest thing to *direct* knowledge would probably be something like a judgement which is not *caused* by other beliefs. If we understand direct knowledge this way, it is not clear how the phenomena of illusion and hallucination would be relevant to the claim that on at least some occasions our judgements about the physical world are reliably produced by processes that do not take as their input beliefs about something else.

*See also* DIRECT REALISM; PHENOMENALISM; PROBLEM OF THE EXTERNAL WORLD; REPRESENTATIVE REALISM.

## BIBLIOGRAPHY

- Austin, J. L.: *Sense and Sensibilia* (Oxford: Oxford University Press, 1968).
- Ayer, A. J.: *The Problem of Knowledge* (Edinburgh: Penguin, 1956).
- Barnes, W. H. F.: "The Myth of Sense-Data," *Proceedings of the Aristotelian Society* 45 (1944–5), 89–117.
- Price, H. H.: *Perception* (London: Methuen, 1932).
- Russell, B.: *The Problems of Philosophy* (Oxford: Oxford University Press, 1959).
- Sellars, W.: *Science. Perception and Reality* (London: Routledge and Kegan Paul, 1963); see especially "Empiricism and the Philosophy of Mind."

RICHARD FUMERTON

**Aristotle (384–322 BC)** Greek philosopher, born in Stagira in northern Greece. Aristotle is influenced by questions raised in Plato's dialogues (especially the *Meno* and *Theaetetus*) about the definition of knowledge and the conditions for justification (see PLATO). Though none of his works is devoted exclusively or primarily to epistemology, he discusses many epistemological problems. Different discussions seem to rely on different, sometimes conflicting, epistemological assumptions. (We might seek to resolve this conflict by tracing a development in Aristotle's thought.) He discusses several of the issues that become explicit topics of debate among Hellenistic philosophers; indeed both sides in some Hellenistic debates can claim support from Aristotle. It is particularly useful to trace his tendencies towards acceptance of a foundationalist view of justification, and some tendencies towards accepting some form of coherence as the appropriate condition for justification.

## 1. THE CONCEPT OF KNOWLEDGE

The Aristotelian terms usually translated by "know" are "*epistasthai*" (cognate abstract noun: "*epistēmē*"), "*gignōskein*" ("*gnōsis*") and "*eidenai*". "*Gignōskein*" and "*eidenai*" are used in many contexts, and applied to knowledge of many different sorts of propositions (about general laws, particular facts, observable and

non-observable states of affairs, and so on); "*epistasthai*", by contrast, sometimes refers specifically to scientific knowledge. Moreover, "*epistēmē*" may refer either (a) to a body of truths known, or (b) to the state of someone who knows them; hence in sense (a) mathematics or astronomy counts as an *epistēmē* (so that "science" is the proper translation), and in sense (b) someone who knows such a science counts as having *epistēmē* (so that "knowledge" is the proper translation). The primary example of an *epistēmē* (in sense (a)) is a demonstrative science (see 2 below), but it is not the only example. Aristotle does not confine his use of the term "*epistēmē*" to demonstrative science; crafts and disciplines that lack a rigorous demonstrative structure are also cases of *epistēmē*.

There is no reason to deny that Aristotle is discussing questions about knowledge; but the examples of knowledge that he has in mind are not always those that would seem most natural to us (though they come to seem more natural if we think of epistemology as continuous with the philosophy of science). His explicit comments are mostly about *epistēmē*; and sometimes we need to take this into account in order to see what he assumes about knowledge in general.

## 2. THE NATURE OF KNOWLEDGE

In the *Posterior Analytics* Aristotle states explicit conditions for knowledge. This is Aristotle's major work on the structure of scientific knowledge (*epistēmē*); it is not an account of how to acquire such knowledge (on which see 4 below), but an account of what we must have acquired if we are to count as possessing such knowledge.

A scientific theory that expresses genuine knowledge must be demonstrative. A demonstrative science has a rigorously deductive structure; it is arranged in demonstrative syllogisms which present the theorems of the science as deductions from first principles that are necessarily true, prior to and better known than the conclusions, and explanatory of the conclusions derived from them.

Aristotle assumes (cf. Plato, *Meno* 98a) that if I know that *p*, then (1) I can justify my belief that *p*, and (2) I know the justification *q*. He insists on (2) because it does not seem satisfactory if I can simply *state q*; it seems reasonable to demand that I should also know why *q* is true and why *q* justifies *p* (*Post. An.* i 2).

To satisfy this second condition, three options are available: (a) An infinite regress; (b) A circle: my sequence of justifications avoids an infinite regress because it eventually comes back to the original belief that I sought to justify; (c) Foundationalism: some propositions are self-justifying, so that we can know them (or have some other suitable cognitive relation to them) without justifying them by appeal to any other propositions, and these are the basis for our knowledge of other propositions.

Hellenistic Sceptics, especially Agrippa, reject all three options, and so conclude that justification, and hence knowledge, is impossible. Aristotle agrees with them (in anticipation) in rejecting the first two options; he disagrees with them, however (*Post. An.* i 3), in so far as he accepts the third option; in his view, we have a self-justifying intuitive grasp (*nous*: *Post. An.* ii 19) of the first principles of each science.

Aristotle's argument, then, is a statement of a foundationalist position (see FOUNDATIONALISM). His candidates for fundamental, self-justifying propositions are rather surprising; they are the basic principles of specific sciences. These are different from, say, the basic axioms recognized by Descartes as objects of intuition; Aristotelian *nous* extends to these, but not only to these. Aristotle's first principles differ even more sharply from the sensory states or propositions that count as basic for an empiricist foundationalist (though Aristotle may be a foundationalist about these too; see the next two sections). Given Aristotle's conception of a science, the first principles cannot be justified by anything more fundamental within the science of which they are principles. But it is natural to argue that the scientific theory as a whole has to be justified by reference to the empirical evidence and other considerations that lead us to construct one sort of

theory rather than another. (This might be taken as an objection to Aristotle, or as an objection to the interpretation just suggested.) Aristotle certainly agrees that scientific theory has to be founded on the applicable empirical observations and other relevant considerations. But he does not explain how he reconciles this relation of theory to experience with the epistemological status that he claims for the first principles of a scientific theory.

### 3. PERCEPTION AND KNOWLEDGE

Aristotle's discussion of the epistemology of sense-perception appears in *De Anima* ii–iii, as part of his account of the nature of perception as a capacity of the soul, to be compared and contrasted with imagination, thought and desire. He offers three formulae for the understanding of perception: (F1) The perceiver becomes like the object (417a18). (F2) The perceiver that was potentially F (e.g. white) becomes actually F when it perceives the actually F object (418a3). (F3) The perceiver acquires the form, but not the matter, of the object (424a18–24). F3 is intended to capture the truth in F1–2 without the misleading suggestion that the perceiver becomes physically similar to the object. The “form without the matter” expresses a systematic correspondence between features of the perceiver and features of the object without implying physical similarity. (The nose does not smell like an onion when we smell an onion.)

The three formulae of perception assume a realist view of perception (so that an object is white, square, etc. in its own right, whether or not we perceive it as such). Aristotle seems to intend to hold this realist view throughout his discussion of perception, but some of his remarks do not seem to be consistent with it. At one point he claims that when an object is perceived as red, it becomes actually red, and its becoming red consists in its being perceived as red. This claim seems to conflict with F1–3 for they assume the objective existence of real perceptible qualities independent of the perceiver, but the claim about when something is “actually red” seems to

imply that the existence of red etc. depends on their being perceived. (For Aristotle's attempt to resolve this apparent conflict in his view, see 426a20–6.) Aristotle seems to take this view partly because he believes that each sense is infallible about its special objects (427b8–16); and he intends this view as a defence against sceptical attacks on the senses (*Met.* 1010b14–26). But his claim about “actual” redness does not show that the senses are infallible about the external (perceiver-independent) qualities of objects; and so it does not seem to meet the most serious sceptical arguments.

Aristotle attributes to a “common sense” the perception of size, shape and number, which are all perceived by motion (425a14–20). These are related to, though not the same as, the properties that Plato calls “common”. Plato argues that certain properties are common, in contrast to the proper objects of particular senses, because they are grasped by the rational soul itself, not by the senses (*Tht.* 184–6); Aristotle, however, rejects Plato's argument, suggesting instead that these properties are grasped by a unified faculty of perception. He may be influenced by the assumption that perception by itself, without interpretation or further inference, is a reliable foundation for claims to knowledge; and this assumption may rest in turn on foundationalist assumptions about justification.

#### 4. METHOD

Aristotle discusses the proper method of inquiry in empirical science and in philosophical argument. The similarities and differences that he sees between the two forms of inquiry throw some light on his epistemological outlook.

His descriptions of empirical inquiry suggest that it proceeds from perceptual observations (“appearances”, *phainomena*) through induction (*epagôgê*) until we reach experience (*empeiria*), which is the best way to reach first principles: “Hence it is a task for experience to supply the principles about a given area. . . . For if our inquiry (*historia*) leaves out none of the facts that truly hold of things, we will be

able to find and produce a demonstration of whatever admits of demonstration, and if something does not admit of demonstration, to make this evident also” (*Prior An.* 46a17–27). Aristotle's own *Historia Animalium*, for instance, is a collection of appearances introduced as preliminary to the exposition of the theory (see *HA* 491a7–14, *PA* 640a12–16). Aristotle recognizes observation and experience not only as the basis for forming theories, but also as a means of testing them; and for this purpose he advocates the collection of new observations (see *GA* 760b28–33). His general description of empirical inquiry is intelligible in the light of the foundationalist attitudes we have traced elsewhere.

His attitude to philosophical inquiry seems different from his attitude to empirical inquiry, even though he describes both forms of inquiry in rather similar terms, as beginning from “appearances”. In describing philosophical argument, he says:

We must, as in the other cases, set out the appearances, and then first raise the puzzles [about them]. In this way we must prove, ideally, all the common beliefs, and if not all of them, at least most of them, and the most important. For if the difficulties are dissolved, and the common beliefs are left standing, that will be a sufficient proof.

(*Nic. Ethics* 1145a2–7)

This method is dialectical, and Aristotle thinks dialectical inquiry is the appropriate way to advance towards the first principles of the sciences (*Topics* 101a36–b4). In this case the “appearances” are “common beliefs” (*endoxa*), the views of “the many and the wise” (*Topics* 104a8); they are not confined (as they characteristically are in the case of empirical inquiry) to perceptual observations. Appearances do not appear to play the foundational role in dialectical inquiry that they play in empirical inquiry. In dialectical inquiry Aristotle recognizes that it may be appropriate to reject some common beliefs if they do not fit with what seems, on other grounds, to be the best theory. In this case his arguments seem to appeal to considerations of coherence that reject any appeal to an infallible foundation (see COHERENTISM). The



epistemology underlying Aristotle's own philosophical argument is difficult to reconcile with some of the epistemological assumptions that he sometimes seems to accept. We can illustrate this tension in Aristotle's thinking by examining some of his anti-sceptical arguments.

## 5. REPLIES TO SCEPTICISM

In *Metaphysics* iv Aristotle discusses several challenges to propositions that he regards as basic first principles. He includes a discussion (iv 5) of sceptical arguments about the senses. These arguments are derived from conflicting appearances, as follows: (1) Different people's perceptual appearances conflict (e.g. the water seems cold to me and hot to you, the tower seems square to me and round to you). (2) There is no reason to prefer one appearance over the other (they are "equipollent"). (3) Hence we should suspend judgement about which appearance is true (*Met.* 1009a38–b12; cf. Sextus, *Pyrr. Hyp.* i 8, 10).

Aristotle denies that conflicting perceptual appearances are always equipollent. He insists that we often have a satisfactory criterion for preferring one of a pair of conflicting appearances over the other. He recognizes that the criterion itself may be challenged; but he rejects the challenges (1010b3–11). Since discrimination between appearances is not as difficult as the sceptic suggests it is, the sceptic's problem does not arise.

The sceptic assumes that (1) we are justified in believing that *p* only if we can prove *p* by appeal to some further principle *q*; and (2) we can prove *p* in this way only if we are justified in believing *q* independently of believing *p*, and we can produce the proof of all the predecessors of *p* (i.e. of *q*, of the justification of *q*, . . . etc.). We can stop an infinite regress only if the sceptic agrees to a principle; but he will not agree to one, because at each stage he can demand a further proof (Sextus, *Pyrr. Hyp.* ii 53). Aristotle rejects the sceptic's assumptions. He argues that his opponents have the wrong conception of the proper sort of justification; they look for demonstrations when they should not

(1006a5–11, 1011a3–13). In asking why we believe the doctor, or how we can tell we are awake, the sceptics keep asking for a further principle, so that they will not have to take anything on trust without a demonstration. But this demand will never yield the right sort of justification; "it would go on to infinity, so that there would be no demonstration that way either" (1006a8–9). Once the sceptics see that the demand for demonstration is the source of their puzzles, and is itself an unreasonable demand, they will give up their challenge (1011a11–16).

Aristotle might mean that a good reply to the sceptic will point to infallible self-justifying foundations that we can rely on to provide the sorts of criteria that are beyond sceptical doubt or question. In fact, however his reply to the sceptic seems to undermine the assumptions that would lead us to look for that sort of foundation. For he seems to suggest that a criterion need not be, or rest on, some infallible foundation; it only needs to be one that we have good reason – in the light of the rest of our beliefs – for accepting. If this is what Aristotle means, then he appeals to coherence. But he does not pursue the issue far enough to formulate or defend the epistemological assumptions underlying his reply to scepticism.

See also SEXTUS EMPIRICUS.

## WRITINGS

A convenient translation of Aristotle's complete works is *The Complete Works of Aristotle: The Revised Oxford Translation* 2 vols (Princeton: Princeton University Press, 1984). The Clarendon Aristotle series contains useful commentaries on individual works: J. Barnes (trans. and ed.) *Posterior Analytics* (Oxford: Clarendon, 1975); D. W. Hamlyn (trans. and ed.) *De Anima* (1968); C. A. Kirwan (trans. and ed.) *Metaphysics* iv. v, vi (1971).

## BIBLIOGRAPHY

A helpful survey of Aristotle's epistemology (including discussion of his moral epistemology) is C. C. W. Taylor, "Aristotle's Epistemology,"

in *Epistemology*, ed. S. Everson (Cambridge: Cambridge University Press, 1990), ch. 6. For a longer treatment of the topics discussed in this entry, see T. H. Irwin, *Aristotle's First Principles* (Oxford: Clarendon Press, 1988), chs 2–3, 6–9, 14. For further bibliography, see Everson (above) and *Articles on Aristotle*, ed. J. Barnes, M. Schofield, and R. Sorabji (London: Duckworth, 1975–9), especially vols 1 and 4. On the connection of some issues in Aristotle with issues in more recent epistemology, see R. M. Chisholm, *The Foundations of Knowing* (Minneapolis, MN: University of Minnesota Press, 1982).

TERENCE IRWIN

**Armstrong, David M. (1926– )** Born in Melbourne, Australia, Armstrong taught at the University of London and at Melbourne University, and was Challis Professor of Philosophy at Sydney University, 1964–91. Besides his work in epistemology, his main contributions to philosophy have been in philosophy of mind and metaphysics.

There are two strands to Armstrong's work in epistemology, which make it distinctive. First, Armstrong revives an idea that originated in the work of F. P. Ramsey, which he calls reliabilism (see RELIABILISM). The central core of reliabilism (for non-inferential knowledge) is the identification of knowledge with nomically reliable true belief, i.e. belief that arises as a result of the law-like connection between ourselves and the world. If one is placed before a tree and has all of one's faculties, then the belief that is formed as a result of the perception counts as knowledge because the nomic relation between the tree and the perceptual process serves as an objective justification of the belief. This justification is objective rather than subjective because the individual need not, and indeed may not, be aware of the requisite nomic relation in order to be justified in having the belief (see OBJECTIVE/SUBJECTIVE).

The analogy that Armstrong favours for this model of knowledge is given by the thermometer. A thermometer is a reliable guide to the temperature because there is a nomic relation between the two. The thermometer correctly represents the amount of heat in a room in the same way that a belief about

the tree correctly represents the external environment. Thus, in a sense, belief measures the environment.

For inferential knowledge Armstrong appeals to the framework of classical deductive logic and scientific induction (see INDUCTION). The use of the latter he justifies by inference to the best explanation (see INFERENCE TO THE BEST EXPLANATION). A sighting of many black ravens and no non-black ones serves to justify the generalization "All ravens are black", in the sense that it is more probable given the evidence than any alternative hypothesis (see HEMPEL'S PARADOX OF THE RAVENS). Thus the best explanation for the sighting of only black ravens is that all ravens are, in fact, black. Armstrong is thus opposed to a Humean scepticism concerning induction (see PROBLEMS OF INDUCTION) without feeling the need to align himself to any formal inductive logic in the manner of Carnap (see CARNAP).

Armstrong argues that this response to inductive scepticism follows from a belief in strong laws of nature. Armstrong conceives of laws as contingent relations between universals, and calls these strong laws. Since laws are conceived of as more than simply universal generalisations the justification of induction is able to rest upon its shoulders. The property of Blackness is tied to the property of Ravenhood, and that is why it is reasonable to assert the generalisation that all ravens are black given a sample.

The second major strand to Armstrong's thought in epistemology is a belief in the Moorean certainties. Like Moore (see MOORE), and unlike Russell (see RUSSELL), Armstrong believes that some of our beliefs are so fundamental that philosophical doubt cannot be rationally entertained. He believes, for example, that one cannot seriously entertain a rational doubt that one has a body. Any philosophical speculation designed to produce such a doubt would require an argument with some contingent premiss that is more assertible than the doubted proposition and, in this case, such a one cannot be found.

The belief in Moorean certainties is intimately related to Armstrong's realism (see

REALISM). The existence of the external world is a Moorean certainty, its character the object of scientific discovery, and the only entities a metaphysics should postulate are those required by good scientific explanations.

See also CAUSAL THEORIES IN EPISTEMOLOGY; DIRECT REALISM; INTROSPECTION; KNOWLEDGE AND BELIEF; OTHER MINDS; SELF-KNOWLEDGE AND SELF-IDENTITY.

#### WRITINGS

*A Materialist Theory of Mind* (London: Routledge and Kegan Paul, 1968), esp. ch. 10.  
*Belief, Truth and Knowledge* (Cambridge: Cambridge University Press, 1973).  
 “Epistemological Foundations for a Materialist Theory of Mind,” in his *The Nature of Mind* (Brighton: Harvester Press, 1981).  
*What Is a Law of Nature?* (Cambridge: Cambridge University Press, 1983).  
 “What Makes Induction Rational?,” in *Dialogue* 83 (1991), 503–11.  
*A World of States of Affairs* (Cambridge: Cambridge University Press, 1997).  
*Truth and Truthmakers* (Cambridge University Press, 2004).

ADRIAN HEATHCOTE

**association** A relation between two or more ideas such that the appearance of one of them in the mind naturally leads one or more of the others to appear there. Locke appealed to “the association of ideas” to explain certain sorts of error that the mind seems naturally prone to, as when the “wrong” idea appears instead of the associated “right” one. For Hume (see HUME) the discovery of the “principles of association” was as central to the task of a positive science of the human mind as he thought the principle of universal gravitation had been to an understanding of physical nature. No particular associations were seen as intrinsic to a given idea; they are established in each individual mind only by repeated experience. Hume thought all comings and goings of mental items, or all transitions in our thinking, are instances of one or another of only three “principles of

association”: resemblance, contiguity, and cause and effect. These are “to us the cement of the universe”. This led to an “associationist” movement or school of psychology which envisaged systematic explanations of all human behaviour along similar lines. Early experimental psychology studied the ways in which associations are originally established.

#### BIBLIOGRAPHY

Hume, D.: *A Treatise of Human Nature* (1739), ed. L. A. Selby-Bigge, revised P. H. Nidditch (Oxford: Oxford University Press, 1978).  
 Hume, D.: *An Enquiry Concerning Human Understanding* (1748), in *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, ed. L. A. Selby-Bigge, revised P. H. Nidditch (Oxford: Oxford University Press, 1975).  
 Locke, J.: *An Essay Concerning Human Understanding* (1690), ed. P. H. Nidditch (Oxford: Clarendon Press, 1975).

BARRY STROUD

**ataraxia** A Greek term sometimes translated by “tranquillity”, it refers to a state of mind characterized by the absence of disturbance or stress, which was cultivated by Pyrrhonian sceptics in antiquity. Sceptics equated *ataraxia* with the end (*telos*) of scepticism, that for the sake of which sceptical argumentation was undertaken, and the most desirable state of human existence. The Pyrrhonian view was that *ataraxia* resulted from suspending belief about the nature of reality. To that end sceptics argued against all dogmatic philosophical theories. Claiming Pyrrho as a model of philosophical detachment and tranquility, sceptics regarded philosophy primarily as a practical enterprise and a way of life.

See also PYRRHONISM; SEXTUS EMPIRICUS.

CHARLOTTE STOUGH

**Augustine, St (354–430)** Theologian, Bishop of Hippo in North Africa. Augustine builds his epistemology around an account

of our certain knowledge of necessary truth. His paradigms of this sort of truth include basic mathematical and logical truths such as “ $7 + 3 = 10$ ” and “there is one world or it is not the case that there is one world” (*De libero arbitrio* II.8.83, *Contra academicos* III.10.23), but also propositions about value and morality (“what is incorruptible is better than what is corruptible”, “we should live justly”; *De libero arbitrio* II.10.110–14). Certain knowledge of truths of this sort rests on direct awareness of their necessity, immutability and eternity. He groups together with these logically necessary propositions a small group of contingent propositions that can be known with certainty (“I exist”, “I seem to see white”; *De civitate Dei* XI.26, *Contra academicos* III.11.26). He appeals to our knowledge of particular truths of these sorts to establish against the sceptic the general possibility of knowledge.

Augustine argues that it follows from the nature of the paradigm objects of knowledge that truth is perceptible only by the mind or reason (*mens, ratio*) and not by sense perception (see PLATO). Since all objects of the senses are contingent and mutable we cannot have knowledge through sense perception. We discern intelligible objects directly by turning within the immaterial soul and away from sense perception and the material world (*De libero arbitrio* II.8, *Confessiones* III–IV, VII.10). He develops his notion of direct acquaintance in terms of the metaphors of light and vision. Just as our seeing material objects depends on their being illuminated by the light of the sun, our intellectual vision of intelligible objects depends on their being illuminated by an intelligible light, truth itself (*De libero arbitrio* II.8.92, 9.108, 12.130, *Soliloquia* 18.15). Hence, knowledge of immutable, eternal truths requires direct acquaintance not only with certain kinds of objects but also with the fact that those objects have the property of being necessary, immutable and eternal. Augustine identifies truth itself with God, who is himself necessary, immutable, and eternal, and hence maintains that our knowledge of truth rests on divine illumination (*Confessiones* IV.15, VII.10, *Soliloquia* I.6.12).

Augustine distinguishes between beliefs grounded in this sort of intellectual vision and beliefs justified in other ways (*De utilitate credendi* II.25). The former sort constitutes paradigm or strict knowledge (*scientia, sapientia*), and when a belief is grounded in this way we can be said to have understanding (*intellectus*). The latter constitutes mere belief or (when the justification is sufficient) knowledge only in a broad sense (*Retractiones* I.14.3). The justification associated with understanding differs from that associated with mere belief not only in degree but also in kind. Understanding of a proposition requires evidence that is internally related to the proposition, so that one possesses the reason for the truth of the proposition. Other sorts of evidence – for example, testimony (see TESTIMONY) – can provide justification but are only externally related to the propositions they support. One can be said to know (in a broad sense) a theorem of geometry, for example, when one believes it on the testimony of a geometer, but one can be said to understand that theorem only when one grasps its proof. Augustine holds that a vast number of our beliefs – for example, all those about events and places we have not ourselves experienced and about other people’s beliefs and attitudes – rest on the testimony of others and that, despite the fact that beliefs based on testimony lack the paradigm sort of justification provided by intellectual vision, we are nevertheless epistemically justified in holding many beliefs of this sort (*De utilitate credendi* 9.22–10.23, 11.25).

Augustine’s famous recommendation that, in theological matters, one ought to believe in order that one may attain understanding is based in part on his distinction between kinds of justification. The testimony of Scripture and the Church (validated both by their historical reliability and by miracles) sufficiently grounds propositions of Christian doctrine so that we are justified in accepting them (*Confessiones* VI.5, *De utilitate credendi* 14.32–16.34). Nevertheless creatures possessing reason can (and perhaps are obligated to) acquire understanding of those propositions to some degree; that is, acquire for those propositions a basis of the sort provided by

intellectual vision (*Epistola CXX*). Augustine's work in philosophical theology is an attempt to provide for propositions of Christian doctrine, which are justifiably believed on testimony, the sort of internal justification necessary for understanding.

#### WRITINGS

- Confessions*, trans. R. S. Pine-Coffin (Harmondsworth: Penguin, 1961).  
*Contra academicos*, trans. M. P. Garvey (Milwaukee, WI: Marquette University Press, 1957).  
*De civitate Dei*, trans. H. Bettenson (Harmondsworth: Penguin, 1972).  
*De libero arbitrio*, trans. A. S. Benjamin and L. H. Hackstaff (New York: Macmillan, 1964).  
*De utilitate credendi and Soliloquia*, trans. J. H. S. Burleigh, in *Augustine: Earlier Writings* (Philadelphia, PA: Westminster Press, 1953), 291–323 and 23–63.  
*Epistola CXX*, trans. W. Parsons, in *Saint Augustine: Letters 83–130* (New York: Fathers of the Church, 1953), 300–17.  
*Retractiones*, trans. M. I. Bogan (Washington, DC: Catholic University of America Press, 1968).

#### BIBLIOGRAPHY

- Gilson, E.: *The Christian Philosophy of Saint Augustine*, trans. L. E. M. Lynch (New York: Random House, 1960), part 1.  
 Kretzmann, N.: "Faith Seeks, Understanding Finds: Augustine's Charter for Christian Philosophy," in *Christian Philosophy*, ed. T. P. Flint (Notre Dame, IN: University of Notre Dame Press, 1990), pp. 1–36.  
 Markus, R. A.: "Augustine: Reason and Illumination" and "Augustine: Sense and Imagination," in *Cambridge History of Later Greek and Early Medieval Philosophy*, ed. A. H. Armstrong (Cambridge: Cambridge University Press, 1967), chs 23 and 24.  
 Nash, R. H.: *The Light of the Mind: Saint Augustine's Theory of Knowledge* (Lexington, KY: The University Press of Kentucky, 1969).

SCOTT MACDONALD

**Austin, John L(angshaw) (1911–60)**  
 British philosopher, who spent his working life at Oxford. Austin, like Ryle (see RYLE),

adopted a revolutionary approach to philosophy. Like Ryle he saw philosophy as an attempt to map families of concepts, such as excuses in "A plea for excuses", and, like Ryle, his primary interest was in the rich and complex terms of everyday discourse. Austin's interest in epistemology cannot therefore be regarded as a straightforward contribution to its traditional debates. He sought instead to show that traditional issues and rivalries were often spurious or confused. Two publications in particular mark Austin's attention to this project, namely his influential paper "Other minds" and his book *Sense and Sensibilia*. In the former he examines a traditional scepticism about knowledge; in the latter he queries a traditional defence of phenomenalism largely due to A. J. Ayer (see PHENOMENALISM). In two papers, "Truth" and "Unfair to facts", Austin outlines his account of truth. Beyond that, his work on moral issues in "A plea for excuses" and "If's and can's", and on speech acts in "Performative utterances" and *How to Do Things with Words* falls strictly outside the scope of epistemology, though it follows the same revolutionary path.

In "Other minds" the novelty of his approach is made clear from the start, for he claims not to understand the very terms in which the traditional debate about knowledge is couched. He seeks to make headway instead by examining the ways in which we use the term "know". He considers a number of examples of "statements of particular current empirical fact" (1961, p. 45) which we may claim to know in order to explore the ways in which such a claim might properly be queried and defended. If my claim "There's a bittern" is queried by asking "Do you know?" or "How do you know?" I may reply by citing, *inter alia*, my past experience ("I was brought up in the Fens"), my current evidence ("I hear it") or some specific feature ("Because of its booming"). Such responses in their appeal to expertise, authority and to possible specific illusions or mistakes throw light on the commitments involved in saying appropriately "I know" rather than merely "I believe".

Austin held that there was some analogy between the commitments involved in saying "I know" and in saying "I promise". He



thought also that mistakes had been made in theory of knowledge through what he called the “descriptive fallacy”, that is, the mistake of assuming that uses of language such as “I know” could function only in a descriptive way. He thought that fallacy in part was responsible for the view that “belief” and “knowledge” described mental states in which the latter is simply a superior, and perhaps infallible, version of the former. His own view quite explicitly was that we are always liable to be mistaken and that it was futile to embark on a theory of knowledge which denied this liability.

One common criticism has been that Austin’s account deals only with the conditions for *claiming to know* rather than with the conditions for *knowing*. It has also been thought that a failure to recognize this distinction led Austin into the fallacy of supposing that if it is appropriate to make a claim, then that claim must be true. But Austin explicitly marks the required distinction when he says, “We are often right to say we know even in cases where we turn out subsequently to have been mistaken” (1961, p. 66). Austin’s interest was more in exploring and classifying our uses of “know” than in providing a traditional definition of knowledge in terms of necessary and sufficient conditions. His own programme moved away from what he regarded as a sterile tradition towards a more progressive task. It did not require him to suppose that his own enquiry was actually a contribution to that sterile debate.

In *Sense and Sensibilia* Austin applied the same approach to traditional debates about perception. His target in this context was principally, though not exclusively, the phenomenism outlined in Ayer’s *Foundations of Empirical Knowledge* (see AYER). Here, too, Austin believed that traditional assumptions had been unclear or erroneous, and that progress was to be made not by clinging to the traditional framework but by rejecting it. Austin recognized that Ayer’s theory had the advantage of a modern formulation, but he thought of Ayer’s project as irredeemably traditionalist.

His main objections to Ayer are that the phenomenists’ crucial introduction of the

term “sense-datum” was based on errors and misunderstandings, and that its subsequent use led philosophical thinking about perception astray. He thought that the naive opposition between “physical (or material) object” and “sense-datum” presented a misleadingly simple picture of the objects of perception which in turn encouraged an impoverished choice between naïve realism (see NAÏVE REALISM) and phenomenism. Austin’s procedure here parallels that of other philosophers, such as Kant, Wittgenstein and Ryle, who sought to show that opposed traditional theories were all mistaken because each of them, and the belief that one had to choose between them, rested on erroneous assumptions.

Austin claimed that the traditional argument from illusion (see ARGUMENT FROM ILLUSION) failed to distinguish “illusion” from “delusion” and “hallucination”, failed to give an adequate account of such terms as “looks”, “seems” and “appears”, and subverted the common use of the term “real”. Since Austin held that these complex discriminations marked our understanding of perceptual error he thought that there was no single coherent “argument from illusion” and that its generalized conclusion introducing the term “sense-datum” was therefore unjustified. He points out the error of thinking that in cases of illusion, and in uses of “appear”, what is presented to us is not part of a public, physical world. He believed that in arguments such as Ayer’s the opposition “physical object/sense-datum” had been erroneously conflated with the “real/apparent” distinction.

Ayer defended himself against Austin’s attack in “Has Austin refuted sense-data?”, but he did not succeed in reviving phenomenism. Ayer makes a number of concessions to Austin’s criticism, but where he rejects that criticism he tends to assume that terms like “sense-datum statement” or his preferred alternative “experiential statement” are perfectly clear and legitimate. He thus tends to beg the question against Austin’s central complaint. Ayer persists in the simple opposition between “experiential statement” and “physical object statement”, though he now concedes that this distinction no longer needs the traditional argument from illusion. But Ayer’s

new account of this distinction remains questionable. For the claim is that physical object statements allow the possibility of mistake just because they “go beyond” immediate presentation to the senses, while this is not true of “experiential statements”. Despite this Ayer now holds that experiential statements are not incorrigible. His position may not be simply inconsistent; but it remains unclear how we are to identify statements which are corrigible but not susceptible of mistake.

See also AYER; SCEPTICISM; SENSE-DATA.

#### WRITINGS

*Philosophical Papers* (Oxford: Oxford University Press, 1961).

*Sense and Sensibilia* (Oxford: Oxford University Press, 1962).

*How to Do Things with Words* (Oxford: Oxford University Press, 1962).

#### BIBLIOGRAPHY

- Ayer, A. J.: “Has Austin Refuted Sense-Data?,” *Synthese* 17 (1967); reprinted in Fann (1969).  
 Fann, K. T., ed.: *A Symposium on J. L. Austin* (London: Routledge and Kegan Paul, 1969).  
 Stroud, B.: *The Significance of Philosophical Scepticism* (Oxford: Oxford University Press, 1985).

GRAHAM BIRD

**avowals** The verb “to avow” has been adopted by many philosophers of mind as the translation of the German verb “*äussern*”. The usual alternative translations are “to express” or “to utter”.

In Wittgenstein’s later work (see WITTGENSTEIN) avowals are the keystone of a new philosophy of mind, founded on the rejection of the Cartesian idea (see CARTESIANISM) that a person discloses the contents of his mind by identifying inner objects and describing them. According to Wittgenstein, an avowal of an intention is not based on a self-examination which parallels the investigation of the world around us: it is only marginally liable to error; and in certain cases is an artificial

expression of the intention replacing a natural one (e.g. a raised fist). Each of these three points makes its contribution to the new philosophy of mind, which some of Wittgenstein’s followers have accepted in its entirety and which, perhaps, nobody can totally reject. But the third point may be the most important one, because it shows how language can develop directly out of behaviour which antedates it. This makes it possible to explain how we can learn, and communicate with, mentalistic language, which were things that remained mysterious when intentions, feelings, etc. were treated as private objects. So it prepares the way for a naturalistic (see NATURALISM), rather than an intellectualist answer to scepticism about other minds.

See also OTHER MINDS; SELF-KNOWLEDGE AND SELF-IDENTITY; WITTGENSTEIN.

#### BIBLIOGRAPHY

- Malcolm, N.: *Nothing Is Hidden* (Oxford: Blackwell, 1986), esp. ch. 8.  
 Wittgenstein, L.: *Philosophical Investigations trans.* G. E. M. Anscombe (Oxford: Blackwell, 1953).

DAVID PEARS

**axiomatization, axiomatics** A theory (see THEORY) usually emerges as a body of (supposed) truths that are not neatly organized, making the theory difficult to survey or study as a whole. The axiomatic method is an idea for organizing a theory (Hilbert, 1970): one tries to select from among the supposed truths a small number from which all the others can be seen to be deductively inferable. This makes the theory rather more tractable since, in a sense, all the truths are contained in those few. In a theory so organized, the few truths from which all others are deductively inferred are called *axioms*. David Hilbert had argued that, just as algebraic and differential equations, which were used to study mathematical and physical processes, could themselves be made mathematical objects, so axiomatic theories, like algebraic and differential equations, which are means of representing physical processes and

mathematical structures, could be made objects of mathematical investigation.

In the tradition (as in Leibniz, 1704), many philosophers had the conviction that all truths, or all truths about a particular domain, followed from a few principles. These principles were taken to be either metaphysically prior or epistemologically prior or both. In the first sense, they were taken to be entities of such a nature that what exists is “caused” by them. When the principles were taken as epistemically prior, that is, as *axioms*, either they were taken to be epistemically privileged (e.g. self-evident, not needing to be demonstrated) (see SELF-EVIDENCE) or (again, inclusive “or”) to be such that all truths do indeed follow from them (by deductive inferences). Gödel (1984) showed – in the spirit of Hilbert, treating axiomatic theories as themselves mathematical objects – that mathematics, and even a small part of mathematics, elementary number theory, could not be axiomatized, that, more precisely, any class of axioms which is such that we could effectively decide, of any proposition, whether or not it was in that class, would be too small to capture all of the truths.

See also GEOMETRY; SPINOZA; THEORY.

#### BIBLIOGRAPHY

- Gödel, K.: “On Formally Undecidable Propositions of Principia Mathematica and Related Systems I,” in *Kurt Gödel: Collected Works*, vol. 1, *Publications 1929–1936*, ed. S. Feferman et al. (Oxford: Clarendon Press, 1986), 145–95.
- Henkin, L., Suppes, P. and Tarski, A., eds: *The Axiomatic Method* (Amsterdam: North-Holland, 1959).
- Hilbert, D.: “Axiomatic thinking,” in *Hilbert: Toward a Philosophy of Modern Mathematics II*, ed. J. Fang (Hauppauge: Paideia, 1970), 187–98.
- Leibniz, G. W.: *New Essays on Human Understanding* (1704), trans. P. Remnant and J. Bennett (Cambridge: Cambridge University Press, 1981).

ROBERT S. TRAGESSE

**Ayer, A(lfred) J(ules) (1910–89)** British philosopher who worked in London and

Oxford. Though Ayer proved to be one of the most important modern epistemologists, there is relatively little epistemology in *Language, Truth and Logic*, his first and most famous book. Indeed, to the extent that epistemology is concerned with the a priori justification of our ordinary or scientific beliefs, the early Ayer explicitly abjures the subject, since the validity of such beliefs “is an empirical matter, which cannot be settled by a priori means” (1946, p. 33). But he does take positions which have bearings on epistemology. For example, he is a phenomenalist (see PHENOMENALISM), believing that material objects are logical constructions (see LOGICAL CONSTRUCTION) out of actual and possible sense-experiences, and an anti-foundationalist, at least in one sense, denying that there is a bedrock level of indubitable propositions on which empirical knowledge can be based. As regards the main specifically epistemological problem he addresses, the problem of our knowledge of other minds, he is essentially behaviouristic, since the verification principle pronounces that the hypothesis of the occurrence of intrinsically inaccessible experiences is unintelligible (1946, pp. 170–1; see LOGICAL POSITIVISM; OTHER MINDS; VERIFICATIONISM).

These themes are developed in Ayer’s second book, *The Foundations of Empirical Knowledge*. The argument from illusion (see ARGUMENT FROM ILLUSION) is deployed towards the conclusion that the objects of perception are sense-data (see SENSE-DATA). This was famously criticized by Austin in *Sense and Sensibilia* (see AUSTIN). Actually, Ayer’s official position is that there are two different “languages” for describing our experiences, sense-data language and material object language, and the question is whether there are “grounds for preferring one method of description to another” (Ayer, 1947, p. 28). But if the relationship between sense-data language and material object language is that of reducing language to reduced, ontological consequences would seem to follow, since it is natural to say that all that “really” exists is the ontology of the reducing language. Ayer goes on to sketch an essentially Humean account of the applicability of

material object language, except that whereas Hume is an error theorist about our commonsense notion of the external world, Ayer takes those aspects of our sense-data which Hume regarded as responsible for errors like belief in the continued existence of material objects, as being instead constitutive of what such phenomena consist in. Thus, according to Ayer, there is nothing more to the continued existence of material things than (a) resemblance between individual sense-data, (b) comparative stability of the contexts in which such resemblances are manifested, (c) systematic repeatability of sense-data, and (d) dependence of the repeatability upon the movements of the observer (1946, pp. 243–59; see also HUME; PROBLEM OF THE EXTERNAL WORLD).

Ayer's main contributions to epistemology are in his book *The Problem of Knowledge*, which he himself regarded as superior to *Language, Truth and Logic* (Ayer, 1985, p. 122). The book opens with general considerations about knowledge and how it is to be distinguished from true belief, and continues with discussions of scepticism about knowledge of (1) the external world, (2) the past and future and (3) the existence of other minds. About knowledge versus true belief, Ayer is a reliabilist (see RELIABILISM): a true belief is not knowledge if it is arrived at "by a process which is not generally reliable" (1956, p. 31). Ayer emphasizes the context-dependence of the reliability of a process, and is consequently in a position to accommodate "relevant alternative" cases (see RELEVANT ALTERNATIVES), in which a normally reliable process, such as looking at an array of objects, is unreliable, even though it might, by chance, lead to the formation of a true belief. So far as Gettier cases are concerned, he could say that the reliable process in the formation of inferential beliefs is application of deductive logic to true beliefs (see GETTIER PROBLEM). But this seems to blur a distinction between the process – in this case deductive reasoning – and that on which it operates. The "No False Lemmas" principle on which Gettier cases turn, that true beliefs arrived at by deductive reasoning are knowledge only if the premisses of the reasoning are also true, is something which

perhaps cannot be subsumed under an account of what makes for reliability in mechanisms of perceptual belief-formation. If so, there is then an interesting question of what makes the concept of knowledge univocal in its application to inferential and non-inferential cases.

Ayer then goes on to develop a fallibilist (see FALLIBILISM) type of foundationalism (see FOUNDATIONALISM), according to which processes of justification or verification terminate in someone's having an experience, but there is no class of infallible statements based on such experiences. Consequently, in making such statements based on experience, even simple reports of observation, we "make what appears to be a special sort of advance beyond our data" (1956, p. 78). And it is the resulting gap which the sceptic exploits. Ayer describes four possible responses to the sceptic: Naïve Realism, according to which material objects are directly given in perception, so that there is no advance beyond the data; Reductionism, according to which physical objects are logically constructed out of the contents of our sense-experiences, so that again there is no real advance beyond the data; a position according to which there is an advance, but it can be supported by the canons of valid inductive reasoning; and lastly a position called "Descriptive Analysis", according to which "we can give an account of the procedures that we actually follow . . . but there [cannot] be a proof that what we take to be good evidence really is so" (1956, p. 81).

Ayer considers each style of response in turn, most noticeably rejecting the second, phenomenalist, option, on the grounds of the impossibility of giving phenomenalist translations of material object discourse. Such translations would be in the form of counterfactuals about possible experiences, and Ayer stresses the main difficulty for phenomenism of analysing the spatio-temporal framework in which observations take place (1956, pp. 123–4). There is also, in the context of his own discussion, the problem that the more *recherché* the analyses become the more the gap between actual experience and what we claim on its basis widens, so that even a

successful phenomenalism would not help against the sceptic.

Ayer finally opts for Descriptive Analysis: “the reason why our sense-experiences afford us grounds for believing in the existence of physical objects is simply that sentences which are taken as referring to physical objects are used in such a way that our having the appropriate experiences counts in favour of their truth” (1956, p. 132). In other words, having such experiences is exactly what justification of our ordinary beliefs about the nature of the world *consists in*. The suggestion is, therefore, that the sceptic is making some kind of mistake or indulging in some sort of incoherence in supposing that our experience may not rationally justify our commonsense picture of what the world is like. Against this, however, is the familiar fact that the sceptic’s

undermining hypotheses seem perfectly intelligible and even epistemically possible. Ayer’s response seems weak relative to the power of the sceptical puzzles.

*See also* AUSTIN; the GIVEN.

#### WRITINGS

*Language, Truth and Logic* (London: Gollancz, 1946).

*The Foundations of Empirical Knowledge* (London: Macmillan, 1947).

*The Problem of Knowledge* (London: Macmillan, 1956).

*More of My Life* (Oxford: Oxford University Press, 1985).

GRAEME FORBES



# B

---

**Bacon, Francis, Lord Verulam (1561–1626)** English philosopher and statesman. While Bacon's philosophy in the wide sense embraces a whole cosmology, in the narrow sense it centres primarily on the identification of a fruitful method for the investigation of nature, with consequent utilitarian benefits. Indeed, it is roughly correct to say that Bacon's identification of knowledge with power was not contingent: it was the practical success of a theory which was for him the hallmark of its truth. Bacon began from a rejection of traditional Scholastic philosophy, which he held provided no method for advancing knowledge, and the oversimple empiricism of the alchemists. He also attacked what he called the Idols: it was false that the human senses are the measure of things; we each have our predispositions which are liable to mislead; our language is itself a further source of misapprehension; and the inherited unsubstantiated theories of the philosophers are mostly wrong.

To escape these causes of error we need to follow a new "method of induction", which could bring both knowledge and control over nature. Baconian induction has often been misinterpreted as a commitment to induction (see INDUCTION) by simple enumeration, and his whole programme dismissed as "inductive" in some pejorative sense. But Bacon explicitly rejected such a method. Rather, he should be seen as subscribing to a version of the hypothetico-deductive method: the positing of causal hypotheses (the forms), which are then brought to the test of experience. The putative identification of the forms, he held, could only follow a rigorous and systematic collection of data or "histories" of natural phenomena.

Aspects of Bacon's method found expression in the work of early members of the Royal

Society such as Boyle and Newton, but the *opera* or works to which Bacon aspired were not scholarly tomes but the practical achievements of the engineer and the chemist. And although Bacon was aware of rising sceptical challenges to knowledge claims, his work largely side-stepped the epistemological concerns that later feature in Descartes and Locke in favour of measuring theories by the pragmatic test of results. Successful theories, he held, might be regarded as certain.

*See also* NATURAL SCIENCE.

## WRITINGS

*Novum Organum* (1620), in *The New Organon and Related Writings*, ed. F. H. Anderson (New York: Liberal Arts Press, 1960).  
*The Advancement of Learning* (1605), Everyman's Library (London: Dent, 1915).

## BIBLIOGRAPHY

- Hesse, M.: "Francis Bacon's Philosophy of Science," in *A Critical History of Western Philosophy* ed. D.J. O'Connor (London: Macmillan, 1964), 141–52.  
Pérez-Ramos, A.: *Francis Bacon's Idea of Science and the Maker's Knowledge Tradition* (Oxford: Oxford University Press, 1988).  
Rossi, P.: *Francis Bacon. From Magic to Science* (London: Routledge and Kegan Paul, 1968).

G. A. J. ROGERS

**basic belief** A belief that is justified without receiving its justification from another belief, or, alternatively, a belief whose justification does not depend on having justification for any other beliefs. Foundationalists hold that there

is a sufficient number of basic beliefs to provide a foundation for non-basic beliefs. They face a threefold challenge: They must explain (i) by virtue of what a belief can enjoy the status of being basic; (ii) which beliefs are basic, and (iii) how non-basic beliefs receive justification from basic beliefs. Against foundationalism, coherentists argue either that there are no basic beliefs at all, or that there are too few of them to constitute a privileged class of beliefs playing the role of a foundation for other beliefs.

See COHERENTISM; EMPIRICISM; EXPERIENCE; FOUNDATIONALISM; the GIVEN; INFINITE REGRESS ARGUMENT; and FOUNDATIONS AND COHERENCE in Part I.

MATTHIAS STEUP

**behaviourism** Epistemologists have discussed a variety of doctrines, which have been called “behaviourism”. One of these, logical behaviourism, holds that sentences about the mental are equivalent in meaning to sentences about behaviour or behavioural dispositions. For example, “Smith feels depressed” might be translated by a behaviourist as “Smith is speaking in a flat monotone and is not exhibiting his usual animation”. A standard problem for such translations is that what is asserted by the first sentence may be true while the second assertion may be false. That might happen, for example, if the subject is depressed but is deliberately acting in a non-depressed manner. Another problem is that Smith might act so as to simulate depression, or at least be disposed to do that, even if he is not depressed. He might, for example, be disposed to act depressed so as to deceive someone into falsely believing that he really is depressed.

A second view, methodological behaviourism, holds that psychologists should, for methodological reasons, reject all mentalistic explanations. A basic problem for this view is that the question of what causes human behaviour appears to be an empirical question; if the evidence supports the postulation of mentalistic causes, it is difficult to see how all mentalistic explanations can be rejected

solely on methodological grounds. This problem is avoided by a third behaviouristic doctrine, associated with the late B. F. Skinner, which might be called “empirical behaviourism”. On this view, there are mental events but no mentalistic causes; behaviour is explainable in terms of such variables as: genetic make-up, states of deprivation, current environmental stimuli and histories of reinforcements and punishments. A particularly difficult problem for this view is to explain the acquisition and use of human language without appealing to mentalistic causes.

A fourth type of behaviourism metaphysical behaviourism, is the thesis that there are no mental events, states or processes. In philosophy, this view is more commonly called “eliminative materialism”. There are other types of behaviourism as well.

See also PSYCHOLOGY AND EPISTEMOLOGY.

#### BIBLIOGRAPHY

- Chomsky, N.: “Review of *Verbal Behavior*,” *Language* 35 (1959), 26–58.  
 Erwin, E.: *Behavior Therapy: Scientific, Philosophical and Moral Foundations* (New York: Cambridge University Press, 1978), chs 2 and 3.  
 Skinner, B. F.: *About Behaviorism* (New York: Knopf, 1974).  
 Zuriff, G. E.: *Behaviorism: A Conceptual Reconstruction* (New York: Columbia University Press, 1985).

EDWARD ERWIN

**belief** The ground-level epistemological concepts of *truth*, *falsity* and *justification* apply primarily to beliefs, and only derivatively, if at all, to knowledge. Belief is central to epistemology.

#### THE STANDARD PICTURE

Plato, in the *Meno*, set the tone for subsequent discussion. There, and in the *Theaetetus*, he distinguishes knowledge, belief, opinion and judgement, and advances a conception of these states of mind according

to which they incorporate a pair of components, one intentional or representational, one causal. In F. P. Ramsey's phrase, a belief is viewed as "a map of neighbouring space by which we steer" (Ramsey, 1978: 134; Armstrong, 1973: 3).

This dual-component picture of belief constitutes what has become the standard picture. It was embellished, for instance, by Hume, who regarded beliefs as "ideas" that, supplemented by a particular "sentiment or feeling", come to serve as guides to behaviour. This conception depicts beliefs as conscious episodes. Although we can be conscious of our beliefs, our being conscious of them appears inessential to their psychological role.

The standard picture is reflected in the notion that beliefs are to be located among the *propositional attitudes*, states of mind comprising (1) propositional contents paired with (2) attitudes towards those contents. These include, in addition to beliefs, desires, wishes, intentions, fears, doubts and hopes. Objects of belief, then, are taken to be propositions expressible sententially. What Agathon believes in believing that rhubarb is poisonous is expressed by the sentence "Rhubarb is poisonous". To some, this suggests that belief-states themselves possess a sentence-like structure, that they are maps in some literal sense. Jerry Fodor (1981), for one, takes beliefs to be internal sentences, inscriptions, "realized" in the head, that make up a "language of thought". Others, suspicious of the notion of a language of thought, nevertheless agree that beliefs possess a "logical form" mirroring that of sentences.

This is one way of filling out the standard picture, but not the only way. Robert Stalnaker (1987, chs 2, 3) argues that the thesis that beliefs are sententially characterizable does not entail that beliefs *themselves* must have a sentential structure. Stalnaker holds that beliefs are attitudes directed, not towards sentence-like propositions, but towards the world. Belief-states are best-characterized by sets of "possible worlds", alternative "ways the world might be". The content of Agathon's belief that rhubarb is poisonous might be represented as a set of

possible worlds, those in which rhubarb is poisonous. Constituents of possible worlds are possible objects and events, rather than concepts or descriptions of objects and events, and logical relations among propositions believed are grounded in relations among sets of possible worlds.

A conception of this sort might account for our ambivalence about beliefs ascribable to creatures lacking in linguistic abilities. Spot, we say, believes there is a squirrel in the tree. But can we be confident that we have captured the content of Spot's belief? Beliefs apparently owe their character, in part, to relations they bear to other beliefs. Your believing that this is a *tree*, for instance, might be thought to require that you believe many other things: that trees are living things, that trees have leaves, that if you set fire to this tree it will burn, and so on. But how many of these background beliefs could we plausibly credit to Spot? Donald Davidson (1984, essay 11) invokes considerations of this sort en route to an argument that the possession of beliefs requires a facility with language.

On Stalnaker's view, however, in ascribing a belief about trees to Spot we need suppose only that Spot has *some* mechanism for dividing alternative situations into those featuring and those lacking trees. The mechanism might be a crude one – cruder, certainly, than the mechanisms underlying the abilities of adult human beings. Nevertheless, Spot's belief concerns trees in part because the proposition believed partitions the set of possible worlds relevant to the explanation of Spot's capacities at the same place the proposition believed by ordinary human beings does.

#### REALISM, ANTI-REALISM, AND ELIMINATIVISM

Fodor and Stalnaker defend "realist" versions of the standard picture. Both take beliefs to be internal states causally implicated in behaviour. Others deflate the standard picture, imagining that beliefs (and other propositional attitudes) possess only an attenuated kind of reality. Quine (1960, § 45), for instance, links the ascription of attitudes to the

translation of utterances. Just as, for Quine, there is no translation-independent fact of the matter as to what a given sentence means, there is no non-contextual fact of the matter as to what a given agent believes. Beliefs, like meanings, exhibit “indeterminacy”. Davidson (1984, essays 9–12) echoes this sentiment. We can concoct distinct, but equally apt, schemes of belief-ascription for any agent. These schemes will depend partly on us, the ascribers, since they hinge on correlations between an agent’s utterances and sentences in our language. It makes no sense to talk of meanings *or* beliefs independently of a particular linguistic context.

A different anti-realist tack is taken by Daniel Dennett (1987, chs 2, 3), who defends an “instrumentalist” conception of belief. We have a practical interest in regarding certain “systems” – people, animals, machines, committees – as rational, as registering, on the whole, what is true and as reasoning in accord with appropriate rational norms. In so doing, we take up the “intentional stance”. We are, as a result, in a position to make sense of and, within limits, to predict the behaviour of the systems in question. The practical success of this enterprise, however, does not depend on its yielding true descriptions of states and goings-on inside agents. More recently (1991), Dennett has endorsed a “mild realism” about belief. We ascribe beliefs to creatures on the basis of “real patterns” the creatures’ behaviour exhibits. Belief ascriptions are true, not because they pick out definite internal states, but because they encompass patterns of behaviour. These patterns stem, no doubt, from internal states, but it would be a mistake to identify beliefs with these states.

It is but a short step from anti-realism to out-and-out scepticism about belief. Stephen Stich (1983), Patricia Churchland (1986) and Paul Churchland (1981) have argued that the concept of belief belongs to “folk psychology”. Folk psychology includes a theory of intelligent behaviour we deploy in interacting with reasoning agents. Beliefs are among the theoretical entities postulated by this theory. Were we to abandon the theory, we would thereby “eliminate” its postulated

entities. Beliefs would go the way of caloric, phlogiston and the ether – theoretical items dropped from our inventory of constituents of the world when theories in which they figured were replaced by better theories. Eliminativists contend that recent advances in psychology and neuroscience make it clear that folk psychology is being eclipsed by better theories, and, on that basis, it is reasonable to deny the existence of beliefs.

Critics respond that this claim is self-defeating. If eliminativists were right, then, by their own lights, it would be impossible to believe what they say – though, of course, it would be equally impossible to believe it false. Whatever the merits of eliminativism, such criticism is ill-considered. An assertion can be true but pragmatically unstable. I cannot coherently claim always to lie, yet, for all that, always lie. As things stand, eliminativists are obliged to articulate their position using concepts that have their home in the folk theory they attack. Presumably a replacement theory would bring with it a stock of concepts that would enable the point to be made without involving reference to beliefs.

Eliminativism has not been widely embraced. It is unclear whether theoretical advances touted by eliminativists would eradicate belief. Psychology might replace the concept of belief with something finer-grained; “belief” could turn out to designate a *range* of states. Were that to happen, however, we might reasonably regard it as a discovery about the deeper character of belief, not its elimination. Were neuroscience to abandon reference to beliefs, we should have no more reason to doubt their existence than we have reason to doubt the reality of trees, rocks, and solid surfaces because basic physics is silent about such things.

#### PROBABILITY THEORIES

Belief has often been depicted as all-or-nothing. Agathon believes that rhubarb is poisonous or he does not. But it is sometimes urged that belief admits of degrees. Agathon believes that rhubarb is poisonous, but he believes still more firmly that rhubarb is a

vegetable. According to Ramsey (1978), beliefs, so regarded, can usefully be treated as *probabilities*. Agathon assigns a certain probability to “rhubarb is poisonous” and a certain, presumably higher, probability to “rhubarb is a vegetable”. One advantage of this perspective is that it allows belief to be quantified and thereby raises the prospect of a rigorous science of behaviour grounded in agents’ beliefs and desires. If we can represent agents as harbouring certain probabilities and simultaneously assign numerical values to their desires, we would be in a position to account for their preferences for various courses of action, hence their behaviour – assuming that agents endeavour to do what they most prefer. The perspective is nowadays associated with decision-theory.

Decision-theory provides a framework for representing rational action. It can also, as Richard Jeffrey (1983, appendix) has shown, accommodate action that is less than perfectly rational. On the whole, however, decision-theory has the flavour of a normative rather than a descriptive theory of behaviour. This has led some to dismiss it as irrelevant to psychological explanation. Others, most notably Davidson (1984, essays 9, 10, 13), have argued that psychology includes an irreducibly normative component. In ascribing beliefs and other propositional attitudes to agents, we are obliged to exercise *charity* – obliged, that is, to assign beliefs and desires in a way that optimizes the rationality of those agents. It is, on Davidson’s view, a “synthetic a priori” truth about agents that, on the whole, they satisfy the canons of decision-theory.

This sanguine view of human rationality has been challenged by empirical work in the psychology of decision-making. Daniel Kahneman and Amos Tversky, for instance, have argued that agents reason using heuristics, rules of thumb that approximate decision-theoretic norms. Reliance on such heuristics results in judgemental biases (Kahneman et al., 1982). More dramatic deviations from rational norms lead to questions of intelligibility. Could agents *believe*, for instance, as some apparently seem to believe, that they are dead? How states of

mind of such agents are best-described is a delicate matter that remains largely inconclusive (Davies and Coltheart, 2000).

#### VOLUNTARISM

The notion that believing is subject to norms suggests that agents are in a position to decide on their beliefs. That seems unlikely. Acquiring a belief resembles catching cold. You might be able voluntarily to bring it about that you hold a certain belief (as Pascal recommends in the case of the belief that God exists), just as you might voluntarily bring it about that you catch cold. But *believing* appears not to be voluntary. Not everyone agrees. Descartes, for instance, writes as though believing depends on some prior act of will. In Meditation IV, he says that “when something is proposed to us by the intellect” we are free “either to affirm or deny” that something.

Descartes represents a minority view. Although what we believe is a function of our other states of mind, it is doubtful that belief essentially involves willing. Indeed, Bernard Williams (1978, ch. 6) has argued that believing is *necessarily* non-voluntary. In the Cartesian spirit, however, some philosophers advocate a two-tiered conception of belief. Thus, Keith Lehrer (1990, ch. 2) distinguishes believing and accepting. We *believe* propositions for a variety of reasons. Agathon might believe that rhubarb is poisonous because he was hypnotised or as a result of wishful thinking. His *accepting* that rhubarb is poisonous, however, requires that he do so because he has grounds for thinking it true. Acceptance is governed by epistemic norms, is at least partly subject to voluntary control, and has functional affinities to belief. Perhaps this would satisfy Descartes. Still, the notion of acceptance, like that of degrees of belief, merely extends the standard picture, and does not replace it.

#### BELIEFS AS DISPOSITIONS

Beliefs make their presence felt in agents’ psychological economies. The ontology of



belief is dispositional. What you do, or would do, depends in part on what you believe. Just as a vase can be fragile without ever shattering, so an agent can harbor a belief without ever acting on that belief. A belief can fail to be manifested when its manifestation is blocked or inhibited, or because an occasion for its manifestation never arises. Beliefs as dispositions must be distinguished from disposition for beliefs (Audi, 1994) – a distinction that would be difficult to make out if you thought dispositions were analyzable counterfactually (Martin 1994). Before reading this sentence, you might have been disposed to believe that the Moon is larger than a peanut but, because the thought never occurred to you, never formed the belief. Such cases are to be distinguished from those in which you harbour an unmanifested belief.

#### PROBLEMS WITH THE STANDARD PICTURE

The standard picture is not without its critics. Gilbert Ryle (1949, ch. 5) regards beliefs as “tendencies” to say and do various things. If Agathon believes that rhubarb is poisonous, he will be disposed to assent to the sentence “rhubarb is poisonous”, to refuse to eat rhubarb, and so on. In this respect beliefs resemble dispositions like fragility. If a glass is fragile, it will, under the right conditions, break when struck by a solid object. Ryle is often accused of reducing beliefs and other states of mind to behaviour, but this is misleading. A glass’s being fragile is not its breaking, and Agathon’s believing that rhubarb is poisonous is not his saying so. Ryle is explicit on the point: What Agathon does given that he believes rhubarb is poisonous will depend on *what else* Agathon believes, what he wants, and so on. If Agathon *wants* to be poisoned, for instance, and believes this is rhubarb, he might well eat it. Functionalists (like Fodor) have made much of this idea.

In one important respect, however, Ryle’s conception of belief clashes with the standard picture. The latter depicts the connection between belief and behaviour as causal. Ryle,

in contrast, argues that mental states like belief are not causes of behaviour any more than a glass’s fragility – its tendency to break – is a cause of its breaking. Beliefs are dispositional states *triggered* causally. But, because these states are characterised by reference to behaviour, their connection with behaviour is conceptual, not causal. Explaining Agathon’s behaviour by reference to Agathon’s belief about rhubarb is like explaining a move in chess by describing it as “taking a pawn”, and not like explaining its starting to rain by reference to falling temperature.

This non-causal conception of psychological explanation sets Ryle off from his functionalist successors. As Davidson (1980) has pointed out, however, any causal relation can be described in such a way that cause and effect are conceptually linked. The falling temperature can be described as the cause of its raining. Thus: The cause of the rain is what caused it to rain. The claim is unilluminating, perhaps, but neither false nor evidence for the absence of a straightforward causal relation.

Ironically, Davidson’s own theory of belief has recently come under attack as being at odds with causal theories. Beliefs are distinguished by their content: what is believed. Many philosophers have argued that content is not “in the head” (Putnam, 1981, ch. 1; Burge, 1986). What your beliefs concern – their content – depends on your worldly circumstances. But if belief content is “broad”, if it includes factors outside the head, in what sense could beliefs be said to cause behaviour? A given state might cause you to behave in a particular way, and that state might be a belief that rhubarb is poisonous. But, if its having this content depends on external circumstances, how could its *content* figure in causing you to do what you do?

The standard picture is usually invoked in attempts to analyze knowledge: *S* knows that *p* only if (1) *S* believes that *p*, (2) *p* is true, (3) *S*’s believing that *p* satisfies some further condition, *C*. A state of knowing is a state of believing that happens both to be true and to satisfy condition *C*. Just as a belief’s being true is irrelevant to the role it plays in an

agent's psychological economy, so its amounting to knowledge could make no difference. Timothy Williamson (2000) has argued that knowledge is both unanalysable and importantly distinct from mere belief. States of knowledge exhibit their own distinctive dispositional profiles. Traditional attempts to analyse knowledge in terms of justified true belief have it backwards: justified beliefs are those grounded in agents' knowledge.

Such considerations have shaken philosophers' confidence in their deployment of the standard picture of belief, without inspiring its wholesale abandonment. Indeed, the standard picture has proved remarkably resilient. It continues to inform our psychological and epistemological speculations. Its replacement, or out-and-out elimination, although conceivable, is not yet on the horizon.

## BIBLIOGRAPHY

- Armstrong, D. M.: *Belief, Truth, and Knowledge* (Cambridge: Cambridge University Press, 1973).  
 Audi, R.: "Dispositional Beliefs and Dispositions to Believe," *Nous* 28 (1994), 419–34.  
 Burge, T.: "Individualism and Psychology," *Philosophical Review* 45 (1986): 3–45.  
 Churchland, P. M.: "Eliminative Materialism and the Propositional Attitudes," *Journal of Philosophy* 78 (1981), 67–90.  
 Churchland, P. S.: *Neurophilosophy* (Cambridge, MA: MIT Press, 1986).  
 Davies, M. and Coltheart, M.: "Introduction: Pathologies of Belief," *Mind and Language* 15 (2000), 1–46.  
 Davidson, D.: "Mental Events," in his *Essays on Actions and Events* (Oxford: Clarendon Press, 1980).  
 Davidson, D.: *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984).  
 Dennett, D. C.: *The Intentional Stance* (Cambridge, MA: MIT Press, 1987).  
 Dennett, D. C.: "Real Patterns," *Journal of Philosophy* 88 (1991), 27–51.  
 Fodor, J. A.: "Propositional Attitudes," in his *Representations* (Brighton: Harvester, 1981).  
 Jeffrey, R. C.: *The Logic of Decision*, 2nd edn (Chicago, IL: University of Chicago Press, 1983).  
 Kahneman, D., Slovic, P. and Tversky, A., eds: *Judgment under Uncertainty: Heuristics and Biases* (Cambridge: Cambridge University Press, 1982).

- Lehrer, K.: *Theory of Knowledge* (Boulder, Colo.: Westview Press, 1990).  
 Martin, C. B.: "Dispositions and Conditionals," *The Philosophical Quarterly* 44 (1994), 1–8.  
 Putnam, H.: *Reason, Truth, and History* (Cambridge: Cambridge University Press, 1981).  
 Quine, W. V.: *Word and Object* (Cambridge, MA: MIT Press, 1960).  
 Ramsey, F. P.: *Foundations: Essays in Philosophy, Logic, Mathematics, and Economics*, ed. D. H. Mellor (London: Routledge and Kegan Paul, 1978).  
 Ryle, G.: *The Concept of Mind* (London: Hutchinson, 1949).  
 Stalnaker, R. C.: *Inquiry* (Cambridge, MA: Bradford Books/MIT Press, 1987).  
 Stich, S. P.: *From Folk Psychology to Cognitive Science: The Case against Belief* (Cambridge, MA: Bradford Books/MIT Press, 1983).  
 Williams, B.: *Descartes: The Project of Pure Enquiry* (Harmondsworth: Pelican Books, 1978).  
 Williamson, T.: *Knowledge and Its Limits* (Oxford: Oxford University Press, 2000).

JOHN HEIL

**belief in and belief that** Traditionally, belief has been of epistemological interest in its *propositional* guise: S believes that *p*, where *p* is a proposition towards which an agent, S, exhibits an attitude of acceptance. Not all belief is of this sort. If I trust what you say, I believe *you*. And someone may believe in Mrs Thatcher, or in a free-market economy, or in God. It is sometimes supposed that all belief is "reducible" to propositional belief, belief-*that*. Thus, my believing *you* might be thought a matter of my believing, perhaps, *that* what you say is true; and your belief in free markets or in God, a matter of your believing *that* free-market economies are desirable or *that* God exists.

It is doubtful, however, that non-propositional believings can, in every case, be reduced in this way. Debate on this point has tended to focus on an apparent distinction between belief-that and belief-in, and the application of this distinction to belief in God (see Swinburne, 1981). Some philosophers have followed Aquinas (see AQUINAS) (in his *Summa Theologiae*) in supposing that to believe in God is simply to believe that certain truths hold: that God exists, that he is benevolent, etc. Others (e.g. Hick, 1957)

argue that belief-in is a distinctive attitude, one that includes essentially an element of *trust*. More commonly, belief-in has been taken to involve a combination of propositional belief together with some further attitude.

H. H. Price (1969) defends the claim that there are different sorts of belief-in, some, but not all, reducible to beliefs-that. If you believe in God, you believe that God exists, that God is good, etc. But, according to Price, your belief involves, in addition, a certain complex pro-attitude toward its object. One might attempt to analyse this further attitude in terms of additional beliefs-that: S believes in *x* just in case (1) S believes that *x* exists (and perhaps holds further factual beliefs about *x*); (2) S believes that *x* is good or valuable in some respect; and (3) S believes that *x*'s being good or valuable in this respect itself a good thing. An analysis of this sort, however, fails adequately to capture the further *affective* component of belief-in. Thus, according to Price, if you believe in God, your belief is not merely that certain truths hold; you possess, in addition, an attitude of commitment and trust towards God.

Notoriously, belief-in outruns the evidence for the corresponding belief-that. Does this diminish its rationality? If belief-in presupposes belief-that, it might be thought that the evidential standards for the former must be at least as high as standards for the latter. And any additional pro-attitude might be thought to require a further layer of justification not required for cases of belief-that.

Some philosophers have argued that, at least for cases in which belief-in is synonymous with *faith* (or faith-in), evidential thresholds for constituent propositional beliefs are diminished (see e.g. Audi, 1990). You may reasonably have faith in God or Mrs Thatcher, even though beliefs about their respective attributes, were you to harbour them, would be evidentially substandard.

Belief-in may be, in general, less susceptible to alteration in the face of unfavourable evidence than belief-that. A believer who encounters evidence against God's existence may remain unshaken in his belief, in part because the evidence does not bear on his pro-attitude. So long as this is united with his

belief that God exists, the belief may survive epistemic buffeting – and reasonably so – in a way that an ordinary propositional belief-that would not.

See also RELIGIOUS BELIEF.

#### BIBLIOGRAPHY

- Audi, R.: "Faith, Belief, and Rationality," *Philosophical Perspectives* 5 (1990).  
 Aquinas: *Summa Theologiae*, trans. Fathers of the English Dominican Province, 5 vols (Westminster: Christian Classics, 1981).  
 Hick, J.: *Faith and Knowledge* (1957), 2nd edn (London: Macmillan, 1967).  
 Price, H. H.: *Belief* (London: George Allen and Unwin, 1965), ch. 2.9.  
 Swinburne, R.: *Faith and Reason* (Oxford: Clarendon Press, 1981).

JOHN HEIL

**Bergmann, Gustav (1906–87)** Bergmann, the last active member of the Vienna Circle (see VIENNA CIRCLE), retained his positivist heritage in denying any general problem of scepticism, but abandoned it in his commitment to the primacy of metaphysics. Thus "epistemology is merely the ontology of the knowing situation" (1964, p. 126), and the task is to show, ontologically, *how* and *what* we know and not *that* we know.

In taking commonsense (perception and knowledge of mind-independent objects, existence of other minds, discoverable regularities in nature) and the findings of science for granted, Bergmann showed the influence of Russell, Wittgenstein and especially Moore (see MOORE). His empiricism (see EMPIRICISM) was expressed in the proposition that all undefined descriptive signs in the ideal language must refer to simple entities (particulars, universals) with which we are directly acquainted in either sense experience or introspection. The ideal language must also reflect the commonsense division of propositions into contingent and necessary, a distinction the reality and importance of which Bergmann defended in his "Two cornerstones of empiricism" (1954, pp. 78–105),

the other cornerstone being just that of reduction to phenomenological simples.

Bergmann's greatest contributions to philosophy lie in philosophy of mind. Taking cues from Brentano and Meinong, he developed an account of intentionality designed in part to exhibit his disavowal of materialist theories of mind while maintaining his commitment to deterministic psychology and, more important, to show how knowledge of mind-independent objects is possible. In so doing, he hoped, in the spirit of Moore, definitively to show the way out of the Cartesian circle of ideas and thus ontologically to secure realism.

#### WRITINGS

*The Metaphysics of Logical Positivism* (Madison, WI: University of Wisconsin Press, 1954).

*Philosophy of Science* (Madison, WI: University of Wisconsin Press, 1957).

*Meaning and Existence* (Madison, WI: University of Wisconsin Press, 1959).

*Logic and Reality* (Madison, WI: University of Wisconsin Press, 1964).

*Realism: A Critique of Brentano and Meinong* (Madison, WI: University of Wisconsin Press, 1967).

*New Foundations of Ontology*, ed. William Heald (Madison, WI: University of Wisconsin Press, 1992).

LAIRD ADDIS

**Berkeley, George (1685–1753)** Berkeley was educated at Trinity College, Dublin, which elected him Junior Fellow in 1707 and with which he remained connected until appointed Dean of Derry in 1724. In 1734 he was appointed Bishop of Cloyne. His main philosophical works were *An Essay towards a New Theory of Vision* (1709), *A Treatise concerning the Principles of Human Knowledge* (1710) and *Three Dialogues between Hylas and Philonous* (1713), but *Alciphron* (1732), *Siris* (1744) and a number of shorter works are of philosophical interest. He was an important influence on Hume, Reid and Kant.

Berkeley's "immaterialism" is a metaphysical system opposed to the "modern philosophy" of Descartes, Locke and other Mechanists, whose account of matter as an independent

substance Berkeley thought set up a rival to God. His arguments have sweeping implications for epistemology. Hume describes them as "the best lessons of scepticism . . . either among the ancient or modern philosophers", but Berkeley's own claim is that they close the door to scepticism opened by "the materialists".

Berkeley's central aim is not, as popularly supposed, to cast doubt on the existence of physical objects. His system is structured round the principle that spirits are the only independent, active beings or, in the philosophical sense, "substances". Sensible things or bodies exist, but as inert beings dependent on a mind which perceives them (*Principles* I 7, 89, etc.). His argument radically reinterprets the traditional notion of substance. According to the logical side of the Aristotelian doctrine, independent substance and dependent accident exist on different ontological levels, indeed in categorically different senses of "exist". As the ultimate subjects of predication, substances are also the substrates of change: they are principles of activity, and change is understood in terms of their natures or essences. Despite a different notion of causal explanation, Mechanist theory of substance has a broadly similar shape. Berkeley's argument accordingly has two sides, corresponding to the logico-ontological independence of spirits and to their activity. The first involves a radical theory of perception, the second, an equally radical philosophy of physics.

#### SPIRITS AND IDEAS

Berkeleyan spirits, like Aristotelian substances, "exist in the first place and primary sense" (*Siris* 263), whereas the *esse* of sensible things is *percipi* (*Principles* I 4). The notion of substance is stripped away from the allegedly "groundless and unintelligible" theory of predication, and the "in" of "Accidents (or qualities) exist in substances" is reinterpreted as the "in" of "The colour I see exists in my mind" (*ibid.* I 49). Spirits are substances which "support" sensible qualities or ideas by perceiving them (*ibid.* I 7, 135). Talk of material substance "supporting" qualities is an empty metaphor (*ibid.* I 17). This identification of

“sensible qualities” with “ideas” becomes understandable if we recognize that, although Berkeley often uses “idea” interchangeably with “sensation” (ibid. I 4), he does not mean by it a state or modification of the mind, but something more in accordance with Descartes’ traditional account of the idea of the sun as “the sun itself as it exists in the understanding”, i.e. as it is experienced or conceived of. Since (Berkeley claims) sensible qualities are ideas in this sense, and “for an idea to exist in an unperceiving thing is a manifest contradiction”, it follows that “there is not any other substance than spirit” (ibid. I 7).

Berkeley is denying, as self-contradictory and absurd, that behind bodies as they appear to the senses lie bodies as they are in themselves, independent “unthinking” substances: he saw this precisely as not denying the existence of bodies, but as making them more immediately accessible to us. The crux of the argument is his rejection of the notion of a quality which is not an idea, i.e. which is not sense-relative. There cannot be “real” and independent qualities which are *like* ideas, since “an idea can be like nothing but another idea” (ibid. I 8). Moreover, our belief that we can conceive of bodies as they exist absolutely, independently of perception, is an illusion engendered by the possibility of imagining something without imagining someone by to perceive it: we forget that we are imagining it as we ourselves would perceive it (ibid. I 22f; *Dialogues* I 200). Yet these general arguments depend on the support of a broad attack on the doctrine according to which “primary” mechanical qualities qualify things as they are in themselves, while “secondary” qualities, such as colours and smells, exist (unless as “bare powers”) only “in the mind” as a product of the primary qualities of the minute parts (cf. Locke, *Essay* 2.8.7–17).

#### PRIMARY AND SECONDARY QUALITIES, SIGHT AND TOUCH, AND ABSTRACTION

One of Berkeley’s arguments against the distinction between primary and secondary qualities is that it is impossible to form an idea of an extended, moving body without

giving it some colour or other mind-dependent quality: what is inseparable from the mind-dependent must itself be mind-dependent (*Principles* I 10). Yet this argument seems weak in the case of qualities perceived by both sight and touch, since what is perceivable by either sense alone must be separable from the other. More fundamental, therefore, is Berkeley’s denial that there are such ideas of two senses. In the *New Theory of Vision* he draws this conclusion from a brilliant discussion of a series of problems in optics, including the “Molyneux problem” (see MOLYNEUX’S PROBLEM). Molyneux and Locke both argued that a man born blind and made to see would be unable to tell distance and depth by sight, since “he has not yet attained the Experience, that what affects his touch so or so, must affect his sight so or so” (Locke *Essay* 2.9.8). Berkeley argues that the only connections between objects of sight and touch are such contingent correlations. The former are signs of the latter, but in themselves numerically and qualitatively distinct (*New Theory* 47–157).

A further range of arguments is directed towards showing that determinate size, shape and motion is sense-relative. Behind these arguments lies the thought that “external” extension, supposed the common object of (say) the naked eye and a view through a microscope, has to be conceived of as infinitely divisible and so, having an infinite number of parts, as of no determinate size or shape (*Principles* I 11, 47; *Dialogues* I 184–193). Berkeley holds that, in contrast to this absurdity, any perceived extension is determinate, composed of a finite number of *minima sensibilia* (*New Theory* 80–5).

Berkeley attributes many errors to a common source: the assumption that what may be marked off by language can also be both abstracted in thought and separated in reality. The existence of sensible things is falsely supposed separable from their being perceived, primary from secondary qualities, and extension from both sight and touch. The Introduction of the *Principles* is accordingly devoted to an attack on Locke’s theory of abstract ideas. Although Locke is misrepresented, Berkeley’s own account of abstract



thought (in fact, much like Locke's) is an important statement of the nominalist view that in reasoning we hold particulars in mind while focusing on those aspects of them marked by the language we employ.

#### KNOWLEDGE OF SPIRIT, SPIRIT'S ACTIVITY AND THE LAWS OF PHYSICS

Berkeley reserves the term "idea" for objects of "perception", i.e. sense and imagination. Spirit is known (in one's own case) "immediately, or intuitively", "by reflection" (*Dialogues* III 231f). Despite the word "reflection", his model for this immediate self-consciousness was not Locke's "inner sense" so much as Cartesian "pure intellect", a term he occasionally uses (ibid. I 193f; *De Motu* 53; *Siris* 303). By its means we attain a "notion" of spirit which we can employ in our thought about God and other finite spirits (*Principles* I 135–40). No idea could represent spirit because ideas are self-evidently passive (we do not perceive power), whereas spirit is an active substance. It is to our consciousness of our own volition that we owe our notion of causal agency. It is therefore a contradiction that any sensible or unthinking thing should be a true cause (ibid. I 27f). Berkeley is here adapting the Cartesian doctrine, developed by Malebranche, that matter is the passive object of God's will.

These materials supply an argument for God's existence and a conception of physical laws as strictly contingent. Since all ideas require a cause, and ideas of sense are not caused by us, there must be some other spirit which causes them in us. Since they occur in an admirable and useful "train or series", they are evidently the work of a wise and benevolent "governing spirit whose will constitutes the Laws of Nature" (ibid. I 30–2). The scientist's task is not, therefore, to penetrate to the unknown essences of things, knowledge of which would make the world intelligible. There are no such essences behind our contingent generalizations. Science progresses purely through the drawing of analogies between phenomena, which it is the function

of such theoretical terms as "gravity" and "attraction" to mark (ibid. I 101–9).

#### REAL THINGS, DIVINE IDEAS AND SCEPTICISM

"The ideas imprinted on the senses by the Author of Nature are called *real things*." With this interpretation of ordinary language Berkeley pre-empts the criticism that, together with mind-independent objects, he has done away with the distinction between reality, and illusion or fiction. The involuntary, distinct and regularly connected ideas of sense are those which matter for the purposes of action (ibid. I 30–6). An objection from lively dreams is met with the principle of Berkeley's reductionism: the realist necessarily concedes all that is needed, since he too must "distinguish *things* from *chimeras* . . . by some perceived difference" (*Dialogues* III 235). The objection that Berkeleyan "real things" cease to exist when not perceived leads, in the *Principles*, to the suggestion that God may perceive them continuously (I 45). In the *Dialogues*, moreover, it is argued that the involuntariness of ideas of sense implies that they ("or their archetypes") have an existence distinct from our minds: causal independence implies ontological independence. Since they (or things like them) are ideas, they must exist in another mind, the one that exhibits them to us (II 211–16). We may conclude that "they exist, during the intervals between the times of my perceiving them" (ibid. III 230f).

Critics have complained that with this conception of a dependable, archetypal order of nature, distanced from finite perceivers, as also in his account of our inference to other minds (see OTHER MINDS), Berkeley leaves no less room than the realist for scepticism. Yet he never argues simply that "materialism" leaves room for doubt, but that it postulates what is contradictory, indeterminate or otherwise unintelligible. It is these disadvantages of matter, none of which (he argues) attach to spirit and its ideas, which in his view encourage scepticism (ibid. III 231–4).

See also IDEALISM; LOCKE; PHENOMENALISM; PRIMARY AND SECONDARY QUALITIES.

#### WRITINGS

- An Essay towards a New Theory of Vision* (1709),  
*A Treatise Concerning the Principles of Human Knowledge* (1710) and *Three Dialogues between Hylas and Philonous* (1713), in *Berkeley: Philosophical Works*, ed. M. R. Ayers (London: Dent, 1975).  
*Alciphron or the Minute Philosopher* (1732) and  
*Siris: A Chain of Philosophical Reflections and Enquiries Concerning the Virtues of Tar-water* (1744), in his *Works*, eds A. A. Luce and T. E. Jessop (London: Nelson, 1948–51).

#### BIBLIOGRAPHY

- Atherton, M.: *Berkeley's Revolution in Vision* (Ithaca, NY: Cornell University Press, 1990).  
Dancy, J.: *Berkeley: An Introduction* (Oxford: Blackwell, 1987).  
Foster, J. and Robinson, H., eds: *Essays on Berkeley* (Oxford: Clarendon Press, 1985).  
Grayling, A. C.: *Berkeley: The Central Arguments* (London: Routledge, 1986).  
Pappas, G. S.: *Berkeley's Thought* (Ithaca, NY: Cornell University Press, 2000).  
Pitcher, G.: *Berkeley* (London: Routledge and Kegan Paul, 1977).  
Sosa, E., ed.: *Essays on the Philosophy of George Berkeley* (Dordrecht: Reidel, 1987).  
Stoneham, T.: *Berkeley's World* (Oxford: Oxford University Press, 2002).  
Tipton, I. C.: *Berkeley: The Philosophy of Immaterialism* (London: Methuen, 1974).  
Turbayne, C., ed.: *Berkeley: Critical and Interpretative Essays* (Minneapolis, MN: University of Minnesota Press, 1982).  
Winkler, K. P.: *Berkeley: An Interpretation* (Oxford: Clarendon Press, 1989).

MICHAEL AYERS

**Blanshard, Brand (1892–1987)** American philosopher, who taught at the University of Michigan, Swarthmore College and Yale. Blanshard attended the University of Michigan but left at the end of his junior year upon being awarded a Rhodes Scholarship to Merton College, Oxford. His studies were

interrupted by the First World War but, having taken an MA at Columbia, where he studied with Dewey, he returned to Oxford to finish his work for the graduate BSc degree. He then went to Harvard where he completed a PhD thesis under the direction of C. I. Lewis (see LEWIS). However, in spite of his work with Dewey (see DEWEY) and Lewis, it was Oxford that decisively shaped Blanshard's ideas. Blanshard thought that the pragmatism (see PRAGMATISM) of Dewey and Lewis was challenging but untenable. Later in his career, he became an uncompromising foe of positivism and linguistic philosophy. He remained a traditionalist, never deviating from the idealism (see IDEALISM) he learned from H. H. Joachim and, particularly, F. H. Bradley. In epistemology, he became British Idealism's last living representative.

Blanshard's most significant epistemological work is *The Nature of Thought* (1939). It contains trenchant criticisms of both empiricism (see EMPIRICISM) and the correspondence theory of truth (see TRUTH). Blanshard is careful to distinguish between questions about the test of truth (i.e. justification) and questions about its nature but, unlike some contemporary philosophers, is certain that both must be explicated in terms of coherence. His argument, which he develops with considerable force, is that explicating justification and truth in disparate terms leads inevitably to scepticism. What is less clear is whether Blanshard himself has a satisfactory solution to this problem for he, too, seems to decouple the test of truth from its nature. As the test of truth, coherence is coherence with "the system of present knowledge", whereas truth in its nature is coherence with "a system complete and all-inclusive" (1939, vol. 2, p. 269). Why, then, suppose that a judgement's coherence with our current beliefs tells us anything about its relations to an ideally completed system? Blanshard replies that, indeed, no judgement in our current system is wholly true: this is the familiar idealist doctrine of degrees of truth which he argues, plausibly, follows necessarily from his holistic conceptions of meaning and

justification. However, the thought that current judgements are at least partially true depends on the claim that “thought is related to reality as the partial to the perfect fulfilment of a purpose” (ibid., p. 262). But Blanshard’s only justification for this teleological postulate is that, without it, scepticism results.

Unlike that of some idealists, Blanshard’s writing is admirably clear; and although his arguments and positions were going out of fashion when he first presented them, they will be found sympathetic by many contemporary philosophers, at least in their critical aspect. If Blanshard is neglected today, it is somewhat unjustly.

#### WRITINGS

*The Nature of Thought*, 2 vols (London: Allen and Unwin, 1939).

*Reason and Analysis* (La Salle, IL: Open Court; London: Allen and Unwin, 1962).

MICHAEL WILLIAMS

**Brentano, Franz (1838–1917)** German psychologist and philosopher. Brentano’s principal contributions to the theory of knowledge derive from the account of conscious states that he sets forth in his *Psychology from an Empirical Standpoint*, first published in 1874.

Our conscious states, according to Brentano, are all objects of “inner perception”. Every such state is such that, for the person who is in that state, it is *evident* to that person that he or she is in that state. (Brentano does not mean that each of our conscious states is an object of an *act* of perception; hence the doctrine does not lead to an infinite regress.) The expression “evident”, for Brentano, is purely epistemic; his use of “evident” is equivalent to the epistemic use of “certain”.

He holds that there are two types of conscious state – those that are “physical” and those that are “intentional”. His use of “physical” in this context is somewhat misleading, for he takes it to mean the the same

as “sensory”. A “*physical*”, or *sensory*, state is a sensation or sense-impression – a qualitative individual composed of parts that are spatially related to each other. *Intentional* states (e.g. believing, considering, hoping, desiring) are characterized by the facts that (1) they are “directed upon objects”, (2) objects may be “directed upon” without existing (e.g. we may fear things that do not exist), and (3) such states are not sensory. There is no sensation, no sensory individual, that can be identified with any particular intentional attitude.

Following Leibniz (see LEIBNIZ), Brentano distinguishes two types of certainty (see CERTAINTY): the certainty we can have with respect to the existence of our conscious states; and that a priori certainty that may be directed upon necessary truths. These two types of certainty may be combined in a significant way. At a given moment I may be certain, on the basis of inner perception, that there is believing, desiring, hoping and fearing; and I may also be certain a priori that there cannot be believing, desiring, hoping and fearing unless there is a substance that believes, desires, hopes and fears. In such a case, it will be certain for me (Brentano says that I will “perceive”) that there is a *substance* that believes, desires, hopes and fears. It is also axiomatic, Brentano says, that, if one is certain that a substance of a certain sort exists, then one is identical with that substance.

Brentano makes use of only two purely epistemic concepts: that of *being certain*, or *evident*, and that of *being probable*. If a given hypothesis is *probable*, in the epistemic sense, for a particular person, then that person can be *certain* that that hypothesis is probable for him. Making use of the principles of probability, one may calculate the probability that a given hypothesis has on one’s evidence base.

But if our evidence-base is composed only of necessary truths and the facts of inner perception, then it is difficult to see how it could provide justification for any contingent truths other than those that pertain to states of consciousness. How could such an evidence-base even lend *probability* to the

hypothesis that there is a world of external physical things? Brentano was reluctant to concede that his theory of knowledge might have such sceptical consequences. (In his theological writings, he attempts to prove a priori the existence of a personal God and he concludes that we have “a probability approaching certainty” that such a being exists. But, unlike Descartes, he does not attempt to base his theory of knowledge upon such a conclusion.)

Perhaps the most significant thing about Brentano’s theory of knowledge is the general problem that he leaves us with. If (a) the knowledge that we have is based upon necessary truths and certain facts about our conscious states, and if (b) application of the principles of probability to this evidence base does not provide probability for our common sense beliefs, then (c) it would seem to be questionable whether we can have *any* justification for such beliefs.

One solution was suggested by Alexius Meinong, who had studied with Brentano in Vienna: those intentional attitudes that we ordinarily call “perceiving” and “remembering” provide “presumptive evidence” – that is to say, *prima facie* evidence – for their intentional objects. For example, believing that one is looking at a group of people tends to justify the belief that there is a group of people that one is looking at. How, then, are we to distinguish merely “*prima facie*” justification from the real thing? This type of solution would seem to call for principles that specify, by reference to further facts of inner perception, the conditions under which merely *prima facie* justification may become real justification.

See also PRIMA FACIE REASONS.

#### WRITINGS

*The True and the Evident* (London: Routledge and Kegan Paul, 1966).  
*Sensory and Noetic Consciousness* (London: Routledge and Kegan Paul, 1981).

#### BIBLIOGRAPHY

McAlister, L. L., ed.: *The Philosophy of Brentano* (London: Duckworth, 1976).  
 Poli, R.: *The Brentano Puzzle* (Aldershot: Ashgate, 1998).

R. M. CHISHOLM

**burden of proof** The strength or weight of argument required by one side to convince the other side. The concept of burden of proof marks a balance between the competing points of view in a critical discussion, or in other adversarial types of dialogue. As more weight is gained by the argumentation of one side, its point of view is justified more strongly, and the burden of proof passes to the other side. As one side rises, the other falls, and the burden of proof passes accordingly.

Recognition and use of burden of proof can be a powerful factor in reasoned persuasion. There are two main uses for the notion. The first is when, having gained a temporary advantage, we announce that the burden of proof now lies with the other side, and simply wait to see what, if anything, they produce; if nothing emerges we claim the victory, even though our own reasoning may have been far from conclusive. The second is where we attempt to claim that our own view enjoys some antecedent presumption in its favour, so that the burden of proof lies *initially* with our opponents. The dictates of common sense are often held to enjoy this privileged position.

See also CIRCULAR REASONING; INFORMAL FALLACIES.

#### BIBLIOGRAPHY

Brown, R.: “The Burden of Proof,” *American Philosophical Quarterly* 7 (1970), 74–82.  
 Sproule, J. M.: “The Psychological Burden of Proof,” *Communication Monographs* 43 (1976), 115–29.  
 Walton, D. N.: “Burden of Proof,” *Argumentation* 2 (1988), 233–54.

DOUGLAS N. WALTON

## C

---

**Carnap, Rudolf (1891–1970)** Carnap was born in Ronsdorf, Germany, and died in Los Angeles. Although he was for a time a member of the Vienna Circle of logical positivists (see VIENNA CIRCLE), he later became a leader of the movement known as logical empiricism (see LOGICAL EMPIRICISM).

Carnap's doctoral dissertation, *Der Raum* (1921), exhibited a neo-Kantian orientation, but shortly after completing it he became a thoroughgoing empiricist. Impressed by the method of analysis advocated in Bertrand Russell (1914) he undertook the project that culminated in his *Der logische Aufbau der Welt* (1928), which he saw as a precise execution of the sort of logical construction Russell recommended (see LOGICAL CONSTRUCTION). Strongly influenced by Ernst Mach, he began using sensations as a basis; later (before publication), because of considerations of Gestalt psychology, he changed the basis to total momentary experiences. In this work he also showed how a physicalistic basis could be adopted; in fact, he subsequently adopted it himself. He held a longstanding commitment to his "principle of tolerance" – that the choice of a language is a matter of efficiency, not a question of ontology. In *Meaning and Necessity* (1956a) this theme is elaborated, with emphasis upon the meaninglessness of such ontological issues.

In keeping with his empiricist outlook (see EMPIRICISM), Carnap adopted a verifiability criterion of meaning (see VERIFICATIONISM). In early days he demanded conclusive verifiability, but later relaxed the requirement to testability, and then to confirmability to some degree (1936–7, 1956b). This development paralleled his realization that scientific knowledge could not be grounded in certainty.

Carnap's chief interests came to be centred, in the early 1940s, on probability and degree

of confirmation, an area that occupied him for the rest of his life. He distinguished two concepts of probability, degree of confirmation and relative frequency (see PROBABILITY, THEORIES OF); his work focused almost exclusively on the former. He constructed a monumental system of inductive logic based on the notion of probability as a logical relation between evidence and hypothesis (1950), which he intended as the foundation of a formalization of confirmation in science. He refined and extended this work until his death in 1970. One of the most valuable aspects is his "clarification of the explicandum" (1950, chs I, II, IV). The major difficulty with his system is the inescapable need for a priori probability measures on factual propositions. He concludes that the basic principles of inductive logic must be based on "inductive intuition" (Schilpp, 1963, p. 978).

In addition to the above-mentioned work, Carnap made contributions of first importance to logic, semantics, and the foundations of mathematics. He also contributed significantly to the unity of science movement.

See also EXISTENCE; NATURAL SCIENCE; PROTOCOL SENTENCES; SCHLICK; SOLIPSISM.

### WRITINGS

*Der Raum* (Jena: Universität Jena, 1921).

"Testability and Meaning," *Philosophy of Science* 3 (1936), 419–71; 4 (1937), 1–40.

"Empiricism, Semantics, and Ontology," in his *Meaning and Necessity*, 2nd edn (Chicago, IL: University of Chicago Press, 1956a), 205–21.

"The Methodological Character of Theoretical Concepts," *Minnesota Studies in the Philosophy of Science*, vol. 1, ed. H. Feigl and M. Scriven (Minneapolis, MN: University of Minnesota Press, 1956b), 38–76.



*The Logical Foundations of Probability* (Chicago, IL: University of Chicago Press, 1950; 2nd edn, with new preface, 1962).

*Der logische Aufbau der Welt* (Berlin: 1928), trans. R. A. George as *The Logical Structure of the World* (Berkeley and Los Angeles, CA: University of California Press, 1967).

# BIBLIOGRAPHY

Goodman, N.: *The Structure of Appearance* (Cambridge, MA: Harvard University Press, 1951).

Russell, B.: *Our Knowledge of the External World* (London: Allen and Unwin, 1914).

Schilpp, P. A., ed.: *The Philosophy of Rudolf Carnap* (La Salle, IL: Open Court, 1963).

WESLEY C. SALMON

**Cartesianism** The name given to the philosophical movement inaugurated by René Descartes (after “Cartesius”, the Latin version of his name) (see DESCARTES). The main features of Cartesianism are (1) the use of methodical doubt as a tool for testing beliefs and reaching certainty; (2) a metaphysical system which starts from the subject’s indubitable awareness of his own existence; (3) a theory of “clear and distinct ideas” based on the innate concepts and propositions implanted in the soul by God (these include the ideas of mathematics, which Descartes takes to be the fundamental building blocks of science); (4) the theory now known as “dualism” – that there are two fundamentally incompatible kinds of substance in the universe, mind (or thinking substance) and matter (or extended substance). A corollary of this last theory is that human beings are radically heterogeneous beings, composed of an unextended, immaterial consciousness united to a piece of purely physical machinery – the body. Another key element in Cartesian dualism is the claim that the mind has perfect and transparent awareness of its own nature or essence.

JOHN COTTINGHAM

**causal theories in epistemology** What makes a belief *justified* and what makes a

true belief *knowledge*? It is natural to think that whether a belief deserves one of these appraisals depends on what caused the subject to have the belief. In recent decades a number of epistemologists have pursued this plausible idea with a variety of specific proposals. Let us look first at some proposed causal criteria for knowledge and then at one for justification.

# KNOWLEDGE

Some causal theories of knowledge have it that a true belief that *p* is knowledge just in case it has the right sort of causal connection to the fact that *p*. Such a criterion can be applied only to cases where the fact that *p* is a sort that can enter into causal relations; this seems to exclude mathematical and other necessary facts and perhaps any fact expressed by a universal generalization; and proponents of this sort of criterion have usually supposed that it is limited to perceptual knowledge of particular facts about the subject’s environment.

For example, Armstrong (1973, ch. 12) proposed that a belief of the form “This (perceived) object is *F*” is (non-inferential) knowledge if and only if the belief is a completely reliable sign that the perceived object is *F*; that is, the fact that the object is *F* contributed to causing the belief and its doing so depended on properties of the believer such that the laws of nature dictate that, for any subject *x* and perceived object *y*, if *x* has those properties and believes that *y* is *F*, then *y* is *F* (see ARMSTRONG). (Dretske (1981) offers a rather similar account, in terms of the beliefs being caused by a signal received by the perceiver that carries the information that the object is *F*.)

This sort of condition fails, however, to be sufficient for non-inferential perceptual knowledge because it is compatible with the beliefs being unjustified, and an unjustified belief cannot be knowledge. For example, suppose that your mechanisms for colour perception are working well, but you have been given good reason to think otherwise, to think, say, that magenta things look

chartreuse to you and chartreuse things look magenta. If you fail to heed these reasons you have for thinking that your colour perception is awry and believe of a thing that looks magenta to you that it is magenta, your belief will fail to be justified and will therefore fail to be knowledge, even though it is caused by the thing's being magenta in such a way as to be a completely reliable sign (or to carry the information) that the thing is magenta.

One could fend off this sort of counterexample by simply adding to the causal condition the requirement that the belief be justified. But this enriched condition would still be insufficient. Suppose, for example, that in an experiment you are given a drug that in nearly all people (but not in you, as it happens) causes the aforementioned aberration in colour perception. The experimenter tells you that you've taken such a drug but then says, "No, wait a minute, the pill you took was just a placebo." But suppose further that this last thing the experimenter tells you is false. Her telling you this gives you justification for believing of a thing that looks magenta to you that it is magenta, but a fact about this justification that is unknown to you (that the experimenter's last statement was false) makes it the case that your true belief is not knowledge even though it satisfies Armstrong's causal condition.

Goldman (1986, ch. 3) has proposed an importantly different sort of causal criterion, namely, that a true belief is knowledge if it is produced by a type of process that is "globally" and "locally" reliable. It is globally reliable if its propensity to cause true beliefs is sufficiently high. Local reliability has to do with whether the process would have produced a similar but false belief in certain counter-factual situations alternative to the actual situation. This way of marking off true beliefs that are knowledge does not require the fact believed to be causally related to the belief, and so it could in principle apply to knowledge of any kind of truth.

Goldman requires the global reliability of the belief-producing process for the justification of a belief; he requires it also for knowledge because justification is required for knowledge. What he requires for knowledge but

does not require for justification is local reliability. His idea is that a justified true belief is knowledge if the type of process that produced it would *not* have produced it in any relevant counterfactual situation in which it is false (see RELEVANT ALTERNATIVES).

What makes an alternative situation relevant? Goldman does not try to formulate a criterion of relevance, but in giving examples of what he takes to be relevant alternative situations he makes remarks that suggest one. Suppose, he says, that a parent takes a child's temperature with a thermometer that the parent selected at random from several lying in the medicine cabinet; only the particular thermometer chosen was in good working order; it correctly shows the child's temperature to be normal but if it had been abnormal then any of the other thermometers would have erroneously shown it to be normal. The parent's actual true belief is caused by a globally reliable process but, Goldman says (1986, p. 45), because it was "just luck" that the parent happened to select a good thermometer "we would not say that the parent knows that the child's temperature is normal". Goldman gives another example:

Suppose Sam spots Judy across the street and correctly believes that it is Judy. If it were Judy's twin sister, Trudy, he would mistake her for Judy. Does Sam know that it is Judy? As long as *there is a serious possibility* [emphasis added] that the person across the street might have been Trudy rather than Judy... we would deny that Sam knows.

(1986, p. 46)

Goldman suggests that the reason for denying knowledge in the thermometer example is that it was "just luck" that the parent did not pick a non-working thermometer and in the twins example the reason is that there was "a serious possibility" that it might have been the other twin Sam saw. This suggests the following criterion of relevance: an alternative situation, where the same belief is produced in the same way but is false, is relevant just in case at some point before the actual belief was caused the chance of that situation's having come about instead of the actual situation was too high;

it was too much a matter of luck that it didn't come about.

This would mean that the proposed criterion of knowledge is the following: a justified true belief that *p* is knowledge just in case there is no alternative non-*p* situation in which the subject is similarly caused to believe that *p* and which is such that at some point in the actual world there was a serious chance that that situation might occur instead of the actual one.

This avoids the sorts of counterexamples we gave for the causal criteria we discussed earlier, but it is vulnerable to ones of a different sort. Suppose you stand on the mainland looking over the water at an island, on which are several structures that look (from at least some point of view) as would barns. You happen to be looking at one of these that is in fact a barn and your belief to that effect is justified, given how it looks to you and the fact that you have no reason to think otherwise. But suppose that the great majority of the barn-looking structures on the island are not real barns but fakes. Finally, suppose that from any viewpoint on the mainland all of the island's fake barns are obscured by trees and that circumstances made it very unlikely that you would have got to a viewpoint not on the mainland. Here, it seems, your justified true belief that you are looking at a barn is not knowledge, despite the fact that there was not a serious chance that there would have developed an alternative situation where you are similarly caused to have a false belief that you are looking at a barn.

That example shows that the "local reliability" of the belief-producing process, on the "serious chance" explication of what makes an alternative relevant, is not sufficient to make a justified true belief knowledge. Another example will show that it is also not necessary. Suppose I am justified in believing the truth that Cornell defeated Brown in their basketball game last night by hearing it so reported by a radio newscaster, and there is nothing at all untoward in the way the newscaster came to say what he did. But suppose further that at the same time, unknown to me, on the other local station a newscaster reads from mistyped copy and says that Brown defeated

Cornell. Since I pretty much randomly chose which local station to listen to, the probability that I would end up with a similarly caused but false belief about the outcome of the Cornell–Brown game was about one-half, a serious chance. Yet surely I know the outcome of the game on the basis of hearing the non-defective newscast as well as I ever know such a thing on such a basis.

These examples make it seem likely that, if there is a criterion for what makes an alternative situation relevant that will save Goldman's claim about local reliability and knowledge, it will not be simple.

#### JUSTIFICATION

The interesting thesis that counts as a causal theory of justification (in the meaning of "causal theory" intended here) is the following: a belief is justified just in case it was produced by a type of process that is "globally" reliable, that is, its propensity to produce true beliefs – which can be defined (to a good enough approximation) as the proportion of the beliefs it produces (or would produce were it used as much as opportunity allows) that are true – is sufficiently great (*see* RELIABILISM).

This proposal will be adequately specified only when we are told (1) how much of the causal history of a belief counts as part of the process that produced it, (2) which of the many types to which the process belongs is *the* type for purposes of assessing its reliability, and (3) relative to what world or worlds is the reliability of the process type to be assessed (the actual world, the closest worlds containing the case being considered, or something else?). Let us look at the answers suggested by Goldman, the leading proponent of a reliabilist account of justification.

(1) Goldman (1979, 1986) takes the relevant belief producing process to include only the proximate causes internal to the believer. So, for instance, when recently I believed that the telephone was ringing the process that produced the belief, for purposes of assessing reliability, includes just the causal chain of neural events from the stimulus in my ears inward and other concurrent brain states on

which the production of the belief depended; it does not include any events in the telephone, or the sound waves travelling between it and my ears, or any earlier decisions I made that were responsible for my being within hearing distance of the telephone at that time. It does seem intuitively plausible that the facts on which the justification of a belief depends should be restricted to internal ones proximate to the belief. Why? Goldman doesn't tell us. One answer that some philosophers might give is that it is because a belief's being justified at a given time can depend only on facts directly accessible to the believer's awareness at that time (for, if a believer ought to hold only beliefs that are justified, she can tell at any given time what beliefs would then be justified for her) (*see* EXTERNALISM/INTERNALISM). But this cannot be Goldman's answer because he wishes to include in the relevant process neural events that are not directly accessible to consciousness.

(2) Once the reliabilist has told us how to delimit the process producing a belief, he needs to tell us which of the many types to which it belongs is the relevant type. Consider, for example, the process that produces your current belief that you see a book before you. One very broad type to which that process belongs would be specified by "coming to a belief as to something one perceives as a result of activation of the nerve endings in some of one's sense-organs". A narrower type to which that same process belongs would be specified by "coming to a belief as to what one sees as a result of activation of the nerve endings in one's retinas". A still narrower type would be given by inserting in the last specification a description of a particular pattern of activation of the retina's receptor cells. Which of these or other types to which the token process belongs is the relevant type for determining whether the type of process that produced your belief is reliable?

If we select a type that is too broad, we will classify as having the same degree of justification various beliefs that intuitively seem to have different degrees of justification. Thus the broadest type we specified for your belief

that you see a book before you applies also to perceptual beliefs where the object seen is far away and seen only briefly through fog, and intuitively the latter sort of belief is less justified. On the other hand, if we are allowed to select a type that is as narrow as we please, then we make it out that an obviously unjustified but true belief is produced by a reliable type of process. For example, suppose I see a blurred shape through the fog far off in a field and unjustifiedly, but correctly, believe that it is a sheep; if we include enough details about my retinal images in specifying the type of the visual process that produced that belief, we can specify a type that is likely to have only that one instance and is therefore 100 per cent reliable. Goldman conjectures (1986, p. 50) that the relevant process type is "the narrowest type that is *causally operative* in producing the belief token in question". Presumably, a feature of the process producing a belief was causally operative in producing it just in case *some* alternative feature is such that, had the process had that feature instead, it would not have led to that belief. (We need to say "some" here rather than "any", because, for example, when I see an oak tree the particular "oakish" shape of my retinal images is clearly causally operative in producing my belief that I see a tree even though there are alternative shapes, for example, "sprucish" ones, that would have produced the same belief.) (*See* RELIABILISM.)

(3) Should the justification of a belief in a hypothetical, non-actual example turn on the reliability of the belief-producing process in the possible world of the example? That leads to the implausible result that in a world run by a Cartesian demon – a powerful being who causes the other inhabitants of the world to have rich and coherent sets of perceptual and memory impressions that are all illusory – the perceptual and memory beliefs of the other inhabitants are all unjustified, for they are produced by processes that are, in that world, quite unreliable. If we say instead that it is the reliability of the processes in the actual world that matters, we get the equally undesirable result that if the actual world is a demon world then our perceptual and memory beliefs are all unjustified.

Goldman's solution (1986, p. 107) is that the reliability of the process types is to be gauged by their performance in "normal" worlds, that is, worlds consistent with "our general beliefs about the world . . . about the sorts of objects, events and changes that occur in it". This gives the intuitively right results for the problem cases just considered, but it implies an implausible relativity regarding justification. If there are people whose general beliefs about the world are very different from mine, then there may, on this account, be beliefs that *I* can correctly regard as justified (ones produced by processes that are reliable in what I take to be normal worlds) but that *they* can correctly regard as not justified.

However these questions about the specifics are dealt with, there are reasons for questioning the basic idea that the criterion for a beliefs being justified is its being produced by a reliable process.

Doubt about the *sufficiency* of the reliabilist criterion is prompted by a sort of example that Goldman himself uses for another purpose. Suppose that being in brain-state B always causes one to believe that one is in brain-state B. Here the reliability of the belief-producing process is perfect. But "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" (Goldman, 1979, p. 6). Doubt about the *necessity* of the condition arises from the possibility that one might know that one has strong justification for a certain belief and yet that knowledge is not what actually prompts one to believe. For example, I might be well aware that, having read the weather bureau's forecast that it will be much hotter tomorrow, I have ample reason to be confident that it will be hotter tomorrow, but I irrationally refuse to believe it until my Aunt Hattie tells me that she feels in her joints that it will be hotter tomorrow. Here what prompts me to believe does not justify my belief, but my belief is nevertheless justified by my knowledge of the weather bureau's prediction and of its evidential force: I can cite it to refute any suggestion that I ought not to be holding the belief. Indeed,

given my justification and that there's nothing untoward about the weather bureau's prediction, my belief, if true, can be counted knowledge. This sort of example raises doubt whether any causal condition, be it a reliable process or something else, is necessary for either justification or knowledge.

*See also* EXTERNALISM/INTERNALISM; MEMORY; NATURALIZED EPISTEMOLOGY; PERCEPTUAL KNOWLEDGE; RELIABILISM; and RECENT WORK ON THE INTERNALISM – EXTERNALISM CONTROVERSY in Part I.

#### BIBLIOGRAPHY

- Armstrong, D. M.: *Belief, Truth and Knowledge* (London: Cambridge University Press, 1973).  
 Dretske, F.: "Conclusive Reasons," *Australasian Journal of Philosophy* 49 (1971), 1–22.  
 Dretske, F.: *Knowledge and the Flow of Information* (Cambridge, MA: MIT Press, 1981).  
 Dretske, F. and Enc, B.: "Causal Theories of Knowledge," *Midwest Studies in Philosophy* 9 (1984), 517–28.  
 Ginet, C.: "Contra Reliabilism," *Monist* 68 (1985), 175–88.  
 Ginet, C.: "Justification: It Need Not Cause but It Must Be Accessible," *Journal of Philosophical Research* 15 (1990), 93–107.  
 Ginet, C.: "The Fourth Condition," in *Philosophical Analysis*, ed. D. F. Austin (Amsterdam: Kluwer, 1988), 105–17.  
 Goldman, A. I.: "Discrimination and Perceptual Knowledge," *Journal of Philosophy* 73 (1976), 771–91.  
 Goldman, A. I.: "What Is Justified Belief?," in *Justification and Knowledge*, ed. G. Pappas (Dordrecht: Reidel, 1979), 1–23.  
 Goldman, A. I.: *Epistemology and Cognition* (Cambridge, MA: Harvard University Press, 1986).

CARL GINET

**certainty** Issues surrounding certainty are inextricably connected with those concerning scepticism (*see* SCEPTICISM). For many sceptics have traditionally held that knowledge requires certainty, and, of course, they claim that certain knowledge is not possible. In part, in order to avoid scepticism, the anti-sceptics have generally held that knowledge



does not require certainty (*see* PRAGMATISM; Lehrer, 1974; Dewey, 1960). A few anti-sceptics have held, *with* the sceptics, that knowledge does require certainty but, *against* the sceptics, that certainty is possible (Moore, 1959; Klein, 1981, 1990). The task here is to provide a characterization of certainty which would be acceptable to both the sceptics and the anti-sceptics. For such an agreement is a pre-condition of an interesting debate between them.

It seems clear that certainty is a property that can be ascribed to either a person or a belief. We can say that a *person*, *S*, is certain; or we can say that a *proposition*, *p*, is certain. The two uses can be connected by saying that *S* has the right to be certain just in case *p* is sufficiently warranted (Ayer, 1956). I will follow the lead of most *philosophers* who have taken the second sense, the sense in which a proposition is said to be certain, as the important one to be investigated by epistemology. An exception is Unger who defends scepticism by arguing that psychological certainty is not possible (Unger, 1975).

In defining certainty, it is crucial to note that the term has both an absolute and relative sense. Very roughly, one can say that a proposition is absolutely certain just in case there is no proposition more warranted than it (Chisholm, 1977). But we also commonly say that one proposition is more certain than another, implying that the second one, though less certain, is still certain.

Now some philosophers, notably Unger (1975), have argued that the absolute sense is the only sense, and that the relative sense is only apparent. Even if those arguments are unconvincing (*see* Cargile, 1972; Klein, 1981), what remains clear is that there is an absolute sense and it is that sense which is crucial to the issues surrounding scepticism.

So let us suppose that the interesting question is this: What makes a belief or proposition absolutely certain?

There are several ways of approaching an answer to that question. Some, like Russell, will take a belief to be certain just in case there is no *logical* possibility that our belief is false (Russell, 1922). On this definition propositions about physical objects (objects occupying

space) cannot be certain. However, that characterization of certainty should be rejected precisely because it makes the question of the existence of absolutely certain empirical propositions uninteresting. For it concedes to the sceptic the impossibility of certainty about physical objects too easily (Ayer, 1956; Moore, 1959). Thus, this approach would not be acceptable to the anti-sceptics.

Other philosophers have suggested that the role that a belief plays within our set of actual beliefs makes a belief certain. For example, Wittgenstein has suggested that a belief is certain just in case it can be appealed to in order to justify other beliefs but stands in no need of justification itself (Wittgenstein, 1969) (*see* WITTGENSTEIN). Thus, the question of the existence of beliefs which are certain can be answered by merely inspecting our practices to determine whether there are any beliefs which play the specified role. This approach would not be acceptable to the sceptics. For it, too, makes the question of the existence of absolutely certain beliefs uninteresting. The issue is not whether there are beliefs which play such a role, but whether there are any beliefs which should play that role. Perhaps our practices cannot be defended.

Let us return to the suggested, rough characterization of absolute certainty given above, namely that a belief, *p*, is certain just in case there is no belief which is more warranted than *p*. Although it does delineate a necessary condition of absolute certainty and it is preferable to the Wittgensteinian approach mentioned above, it does not capture the full sense of "absolute certainty". The sceptics would argue that it is not strong enough. For, according to this rough characterization, a belief could be absolutely certain and yet there could be good grounds for doubting it – just as long as there were equally good grounds for doubting every proposition that was equally warranted. In addition, to say that a belief is certain is to say, in part, that we have a guarantee of its truth. There is no such guarantee provided by this rough characterization.

A Cartesian characterization of the concept of absolute certainty seems more promising. Roughly, this approach is that a

proposition, *p*, is certain for S just in case S is warranted in believing that *p* and there are absolutely no grounds whatsoever for doubting it. Now one could characterize those grounds in a variety of ways (Firth, 1976; Miller, 1978; Klein, 1981, 1990). For example, a ground, *g*, for making *p* doubtful for S could be such that (a) S is not warranted in denying *g* and:

- (b1) if *g* is added to S's beliefs, the negation of *p* is warranted; or
- (b2) if *g* is added to S's beliefs, *p* is no longer warranted; or
- (b3) if *g* is added to S's beliefs, *p* becomes less warranted (even if only slightly so).

Although there is a guarantee of sorts of *p*'s truth contained in (b1) and (b2), those notions of grounds for doubt do not seem to capture a basic feature of *absolute* certainty delineated in the rough account given above. For a proposition, *p*, could be immune to grounds for doubt, *g*, in those two senses and yet another proposition would be "more certain" if there were no grounds for doubt like those specified in (b3). So, only (b3) can succeed in providing *part* of the required guarantee of *p*'s truth.

An account like that contained in (b3) can provide only part of the guarantee because it is only a *subjective* guarantee of *p*'s truth. S's belief system would contain adequate grounds for assuring S that *p* is true because S's belief system would warrant the denial of every proposition that would lower the warrant of *p*. But S's belief system might contain false beliefs and still be immune to doubt in this sense. Indeed, *p* itself could be certain and false in this subjective sense.

An *objective* guarantee is needed as well. We can capture such *objective* immunity to doubt by requiring, roughly, that there be no *true* proposition such that if it is added to S's beliefs, the result is a reduction in the warrant for *p* (even if only very slight). I say "roughly" because there will be the problem of so-called "misleading defeaters". That is, there will be true propositions which if added to S's beliefs result in lowering the warrant of *p* because they render evident some false proposition

which actually reduces the warrant of *p*. It is debatable whether misleading defeaters provide genuine grounds for doubt. But this is a minor difficulty which can be overcome (Klein, 1981, esp. pp. 148–56). What is crucial to note is that given this characterization of objective immunity to doubt, it follows trivially that *p* is true and that there is a set of true propositions in S's belief set which warrant *p* and which are themselves objectively immune to doubt.

Thus, we can say that a belief that *p* is absolutely certain just in case it is *subjectively* and *objectively* immune to doubt. In other words a proposition, *p*, is absolutely certain for S if and only if (1) *p* is warranted for S and (2) S is warranted in denying every proposition, *g*, such that if *g* is added to S's beliefs, the warrant for *p* is reduced (even if only very slightly) and (3) there is no true proposition, *d*, such that if *d* is added to S's beliefs the warrant for *p* is reduced (even if only very slightly).

This is an account of absolute certainty which captures what is demanded by the sceptic. If a proposition is certain in this sense, it is indubitable and guaranteed both subjectively and objectively to be true. In addition, such a characterization of certainty does not automatically lead to scepticism (see Klein, 1981, 1990). Thus, this is an account of certainty that satisfies the task at hand, namely to find an account of certainty that provides the precondition for a debate between the sceptic and anti-sceptic.

#### BIBLIOGRAPHY

- Ayer, A. J.: *The Problem of Knowledge* (London: Macmillan, 1956).
- Cargile, J.: "In Reply to 'A Defense of Scepticism,'" *Philosophical Review* 81 (1972), 229–36.
- Chisholm, R.: *Theory of Knowledge*, 2nd edn (Englewood Cliffs, NJ: Prentice-Hall, 1977).
- Dewey, J.: *The Quest for Certainty* (New York: Putnam, 1960).
- Firth, R.: "The Anatomy of Certainty," *Philosophical Review* 76 (1976), 3–27.
- Klein, P.: *Certainty: A Refutation of Scepticism* (Minneapolis, MN: University of Minnesota Press, 1981).

- Klein, P.: "Epistemic Compatibilism and Canonical Beliefs," in *Doubting: Contemporary Perspectives on Scepticism*, ed. M. D. Roth and G. Ross (Boston, MA: Kluwer, 1990), 99–117.
- Klein, P.: "Immune Belief Systems," *Philosophical Topics* 14 (1986), 259–80.
- Lehrer, K.: *Knowledge* (Oxford: Oxford University Press, 1974).
- Miller, R.: "Absolute certainty," *Mind* 87 (1978), 46–65.
- Moore, G. E.: *Philosophical Papers* (New York: Macmillan, 1959).
- Russell, B.: *Our Knowledge of the External World* (London: Allen and Unwin, 1922).
- Unger, P.: *Ignorance: A Case for Scepticism* (Oxford: Oxford University Press, 1975).
- Wittgenstein, L.: *On Certainty* (Oxford: Blackwell, 1969).

PETER D. KLEIN

**Chisholm, Roderick (1916–99)** Chisholm was an American philosopher who has been influential in a number of different areas of philosophy, including epistemology, metaphysics, and ethics. He was an undergraduate at Brown University and then a graduate student at Harvard in 1938–42. After finishing his studies, he served in the military as a clinical psychologist. He then returned to Brown University in 1947, where he remained on the faculty until his retirement.

In 1942 Chisholm published in *Mind* his first paper on epistemology, "The Problem of the Speckled Hen." Since then, he has addressed every major problem in epistemology. The most important of his writings on epistemology are *Perceiving* (1957), *The Foundations of Knowing* (1982), and, most famously, the 1966, 1977 and 1989 editions of *Theory of Knowledge*. The result of all this work is an epistemological system whose scope and subtlety are unsurpassed in the twentieth century.

At the base of Chisholm's system is a basic notion of justification, which he uses to give definitions of various terms of epistemic appraisal. For example, in the third edition of *Theory of Knowledge* (hereafter TK3), he defines "beyond reasonable doubt" as follows:  $p$  is beyond reasonable doubt for an individual  $S =_{df}$   $S$  is more justified in believing  $p$  than in withholding judgement on  $p$ . In addition,

he defines when  $p$  is certain (nothing is more justified for  $S$  to believe), when  $p$  is evident ( $S$  is as justified in believing  $p$  as in withholding judgement on what is counterbalanced), when  $p$  is epistemically in the clear ( $S$  is as justified in believing  $p$  as in withholding judgement on  $p$ ), when  $p$  is probable ( $S$  is more justified in believing  $p$  than in disbelieving  $p$ ), and when  $p$  is counterbalanced ( $S$  is as justified in believing  $p$  as in disbelieving  $p$ , and vice versa).

Chisholm then uses these terms of epistemic appraisal to formulate a number of epistemic principles. The principles are expressed as conditionals, whose antecedents describe sufficient logical conditions for the application of these terms of appraisal. In the most straightforward case, a principle will assert that if certain non-epistemic conditions are satisfied (e.g. conditions about what  $S$  is experiencing, believing, etc.), then a proposition  $p$  has a certain epistemic status for  $S$  (e.g. it is evident or beyond reasonable doubt for  $S$ ).

The definitions above tell us to understand this status in terms of an undefined notion of justification. Thus, Chisholm's project in epistemology can be seen as the counterpart of a project in ethics that seeks to describe various sets of non-moral conditions that are sufficient to make an action morally right. The ethicists who try to carry out this project are at odds with utilitarians, since for utilitarians there is but one source of moral obligation – namely, utility. These non-utilitarian ethicists insist that there are other sources as well, ones that aren't directly concerned with the maximization of happiness. Equality and fairness, for example, are among the usual candidates. Analogously, Chisholm insists that there is more than one source of epistemic justification.

The principal sources of empirical justification in Chisholm's system are self-presentation (of certain kinds of psychological states – e.g. thinkings, desirings, intendings and sensings), perception, memory, belief coupled with a lack of negative coherence, and, finally, positive coherence among propositions with some antecedent positive epistemic status. Corresponding to each of these sources, Chisholm proposes an epistemic principle describing

the conditions under which the source produces justification. For example, his principle for self-presenting psychological states is as follows (TK3, 19): “If the property of being *F* is self-presenting, if *S* is *F*, and if *S* believes himself to be *F*, then it is certain that he is *F*.”

Chisholm has a reputation as one of leading foundationalists of the twentieth century (see FOUNDATIONALISM), but as the above list of sources makes plain, Chisholm is also a coherentist (see COHERENTISM). However, unlike a pure coherentist, he doesn’t think that positive coherence relations are the only source of empirical justification. Indeed, he doesn’t think that positive coherence is capable of generating justification for propositions that have nothing else to recommend them. It cannot create justification *ex nihilo*. On the other hand, it can ratchet justification up a notch for propositions that already have some other source of justification.

A key to understanding Chisholm’s general approach to epistemology is to understand the metaphysical status of his principles. And for this, it is necessary to understand the status of the basic notion of justification that he uses to define his terms of epistemic appraisal. According to Chisholm, the basic notion is one that we bring to epistemology. It is only because we have a pre-philosophical idea of justification that we are able to identify instances of beliefs that are clearly justified, and it is this, he says, that allows the epistemological project to get off the ground. Thus, he is a particularist when it comes to matters of epistemological method. He begins by examining particular instances of beliefs that he takes to be justified, and then out of these instances he tries to abstract out general conditions of justification, which he expresses in the form of epistemic principles.

There is also an important presupposition that shapes how Chisholm conceives this basic notion of justification. He presupposes that we can improve and correct our beliefs by reflection, eliminating those that are unjustified and adding others that are justified (TK3, 1 and 5). This presupposition acts as a constraint when we try to use particular instances of justified belief to formulate general conditions of justification. It forces us to

look for conditions to which we have reflective access, since otherwise there would be no reason to think that we could eliminate unjustified beliefs and add justified ones simply by being reflective. In effect, this is to say that it forces us to be internalists (see EXTERNALISM/INTERNALISM).

The basic notion of justification that we bring to epistemology is vague, says Chisholm, but he doesn’t think it has to remain vague. On the contrary, he thinks that in the process of formulating epistemic principles, the notion becomes less and less vague, until eventually we are in a position to give a precise characterization of it. The characterization he gives is in terms of ethical requirements on our believings and withholdings. For example, to say that *S* is more justified in believing *p* than withholding on *p* (i.e. to say that *p* is beyond reasonable doubt for *S*) is to say that *S* is required to prefer the former over the latter (TK3, 59).

In turn, Chisholm conceives these ethical requirements as supervening on non-normative states (see EPISTEMIC SUPERVENIENCE). Specifically, they supervene on our conscious states (TK3, 60). As such, a proposition could not have an epistemic status different from the one it does have for us without our conscious states being different.

So, this is how Chisholm understands the metaphysical status of his principles: he thinks of them as necessary truths, and the truths they express are ultimately ones about the relationship between our conscious states at a time and ethical requirements on our believings and withholdings.

This view, when combined with his particularism, yields a position that Chisholm calls (TK2, 124–34) ‘critical cognitivism’ (see CRITICAL COGNITIVISM), which he sees as an alternative to scepticism, intuitionism and reductionism in epistemology (and in ethics as well). Unlike the sceptic, Chisholm begins with the presupposition that some of our beliefs are justified and indeed that some constitute knowledge. Unlike the intuitionist, he denies that there is some special faculty that tells which beliefs are justified (just as he denies that there is some special faculty that tells which actions are good). And unlike

the reductionist, he denies that the truths of epistemology can be reduced to empirical truths. Instead, the relationship is one of supervenience.

See also COMMONSENSISM AND CRITICAL COGNITIVISM; CRITERIA AND KNOWLEDGE; ETHICS AND EPISTEMOLOGY; MORAL EPISTEMOLOGY; PHENOMENALISM; PROBLEM OF THE CRITERION; REID.

#### WRITINGS

- "The Problem of the Speckled Hen," *Mind*, 51 (1942), 71–9.  
*Perceiving: A Philosophical Study* (Ithaca, NY: Cornell University Press, 1957).  
*The Foundations of Knowing* (Minneapolis, MN: University of Minnesota Press, 1982).  
*Theory of Knowledge* (Englewood Cliffs, NJ: Prentice-Hall, 1966, 1977 and 1989).

RICHARD FOLEY

**circular reasoning** Circular reasoning is very important and characteristic of all kinds of everyday argumentation where feedback is used. So it is often quite correct and useful – not fallacious, as traditionally portrayed in the logic textbooks. Studying circular reasoning, for example, is very important for artificial intelligence, e.g. in expert systems. Circular reasoning can be used fallaciously, however, in arguments which require the use of premisses that can be shown to be better established than the conclusion to be proved. The requirement here is one of evidential priority (see INFORMAL FALLACIES: *Arguing in a circle*). Arguing in a circle becomes a fallacy of *petitio principii* or begging the question where an attempt is made to evade the burden of proving one of the premisses of an argument by basing it on the prior acceptance of the conclusion to be proved. (See Walton 1987, pp. 124–44.) So the fallacy of begging the question is a systematic tactic to evade fulfilment of a legitimate burden of proof (see BURDEN OF PROOF) by the proponent of an argument in dialogue by using a circular structure of

argument to block the further progress of dialogue and, in particular, to undermine the capability of the respondent, to whom the argument was directed, to ask legitimate critical questions in reply.

#### BIBLIOGRAPHY

- Basu, D. K.: "A Question of Begging," *Informal Logic* 8 (1986), 19–26.  
 Mackenzie, J. D.: "Begging the Question in Dialogue," *Australasian Journal of Philosophy* 62 (1984), 174–81.  
 Walton, D. N.: *Informal Logic: A Handbook for Critical Argumentation* (Cambridge: Cambridge University Press, 1989).  
 Walton, D. N.: *Begging the Question* (New York: Greenwood Press, 1991).  
 Walton, D. N. and Batten, L. M.: "Games, Graphs and Circular Arguments," *Logique et Analyse* 106 (1984), 133–64.

DOUGLAS N. WALTON

**cogito** The name given to Descartes' famous dictum "I am thinking, therefore I exist" (*je pense, donc je suis*; Latin *cogito ergo sum*). The French phrase appears in the *Discourse on the Method* (1637); the Latin formulation in the *Principles of Philosophy* (1644). The term "the Cogito" is commonly used by commentators to refer not just to the dictum, but to the whole process of reasoning whereby the Cartesian meditator becomes aware of the indubitable existence of the thinking subject. The definitive account of that process occurs in the Second Meditation, where Descartes says that "I am, I exist (*sum, existo*) is necessarily true as often as it is put forward by me or conceived in my mind".

See also DESCARTES; SELF-KNOWLEDGE AND SELF-IDENTITY.

#### BIBLIOGRAPHY

- Hintikka, J.: "Cogito Ergo Sum: Inference or Performance?," *Philosophical Review* 72 (1964), 3–32; reprinted in *Descartes: A Collection of Critical Essays*, ed. W. Doney (London: Macmillan, 1968).



Williams, B.: "The Certainty of the Cogito," in *Descartes: A Collection of Critical Essays*, ed. W. Doney (London: Macmillan, 1968).

JOHN COTTINGHAM

**coherentism** Coherence is a major player in the theatre of knowledge. There are coherence theories of belief, truth and justification. These combine in various ways to yield theories of knowledge. We shall proceed from belief through justification to truth. Coherence theories of belief are concerned with the content of beliefs. Consider a belief you now have, the belief that you are reading a page in a book. So what makes that belief the belief that it is? What makes it the belief that you are reading a page in a book rather than the belief that you have a centaur in the garden?

One answer is that the belief has a coherent place or role in a system of beliefs. Perception has an influence on belief. You respond to sensory stimuli by believing that you are reading a page in a book rather than believing that you have a centaur in the garden. Belief has an influence on action. You will act differently if you believe that you are reading a page than if you believe something about a centaur. Perception and action underdetermine the content of belief, however. The same stimuli may produce various beliefs and various beliefs may produce the same action. The role that gives the belief the content it has is the role it plays in a network of relations to other beliefs, the role in inference and implication, for example. I infer different things from believing that I am reading a page in a book than from any other belief, just as I infer that belief from different things than I infer other beliefs from.

The input of perception and the output of action supplement the central role of the systematic relations the belief has to other beliefs, but it is the systematic relations that give the belief the specific content it has. They are the fundamental sources of the content of beliefs. That is how coherence comes in. A belief has the content that it does because of the way in which it coheres within a system of beliefs (Rosenberg, 1980). We might distinguish weak coherence theories

of the content of beliefs from strong coherence theories. Weak coherence theories affirm that coherence is one determinant of the content of belief. Strong coherence theories of the content of belief affirm that coherence is the sole determinant of the content of belief.

When we turn from belief to justification, we confront a similar group of coherence theories. What makes one belief justified and another not? The answer is the way it coheres with the background system of beliefs. Again there is a distinction between weak and strong theories of coherence. Weak theories tell us that the way in which a belief coheres with a background system of beliefs is one determinant of justification, other typical determinants being perception, memory and intuition. Strong theories, by contrast, tell us that justification is solely a matter of how a belief coheres with a system of beliefs. There is, however, another distinction that cuts across the distinction between weak and strong coherence theories of justification. It is the distinction between positive and negative coherence theories (Pollock, 1986). A positive coherence theory tells us that if a belief coheres with a background system of belief, then the belief is justified. A negative coherence theory tells us that if a belief fails to cohere with a background system of beliefs, then the belief is not justified. We might put this by saying that, according to a positive coherence theory, coherence has the power to produce justification, while according to a negative coherence theory, coherence has only the power to nullify justification.

A strong coherence theory of justification is a combination of a positive and a negative theory which tells us that a belief is justified if and only if it coheres with a background system of beliefs.

Let us illustrate the foregoing distinctions with an example. Coherence theories of justification and knowledge have most often been rejected as being unable to deal with perceptual knowledge (Audi, 1988; Pollock, 1986), and, therefore, it will be most appropriate to consider a perceptual example which will serve as a kind of crucial test. Suppose that a person, call her Trust, works with a scientific instrument that has a gauge for measuring

the temperature of liquid in a container. The gauge is marked in degrees. She looks at the gauge and sees that the reading is 105 degrees. What is she justified in believing and why? Is she, for example, justified in believing that the liquid in the container is 105 degrees? Clearly, that depends on her background beliefs. A weak coherence theorist might argue that, though her belief that she sees the shape 105 is immediately justified as direct sensory evidence without appeal to a background system, the belief that the liquid in the container is 105 degrees results from coherence with a background system of beliefs affirming that the shape 105 is a reading of 105 degrees on a gauge that measures the temperature of the liquid in the container. This sort of weak coherentism combines coherence with direct perceptual evidence, the foundation of justification, to account for justification of our beliefs.

A strong coherence theory would go beyond the claim of the weak coherence theory to affirm that the justification of all beliefs, including the belief that one sees the shape 105, or even the more cautious belief that one sees a shape, results from coherence with a background system. One may argue for this strong coherence theory in a number of different ways. One line of argument would be appeal to the coherence theory of the content of belief. If the content of the perceptual belief results from the relations of the belief to other beliefs in a system of beliefs, then one may argue that the justification of the perceptual beliefs also results from the relations of the belief to other beliefs in the system. One may, however, argue for the strong coherence theory without assuming the coherence theory of the content of beliefs. It may be that some beliefs have the content that they do atomistically but that our justification for believing them is the result of coherence. Consider the very cautious belief that I see a shape. How could the justification for that belief be the result of coherence with a background system of beliefs? What might the background system tell us that would justify that belief? Our background system contains a simple and primal theory about our relationship to the world. To come to the

specific point at issue, we believe that we can tell a shape when we see one, that we are trustworthy about such simple matters as whether we see a shape before us or not. We may, with experience, come to believe that sometimes we think we see a shape before us when there is nothing there at all, when we see an after-image, for example, and so we are not perfect, not beyond deception, yet we are trustworthy for the most part. Moreover, when Trust sees the shape 105, she believes that the circumstances are not those that are deceptive about whether she sees that shape. The light is good, the numeral shapes are large, readily discernible, and so forth. These are beliefs that Trust has that tell her that her belief that she sees a shape is justified. Her belief that she sees a shape is justified because of the way it is supported by her other beliefs. It coheres with those beliefs, and so she is justified.

There are various ways of understanding the nature of this support or coherence. One way is to view Trust as inferring that her belief is true from the other beliefs. The inference might be construed as an inference to the best explanation (*see INFERENCE TO THE BEST EXPLANATION*) (Harman, 1973; Goldman, 1988; Lycan, 1988). Given her background beliefs, the best explanation Trust has for the existence of her belief that she sees a shape is that she does see a shape. Thus, we might think of coherence as inference to the best explanation based on a background system of beliefs. Since we are not aware of such inferences for the most part, the inferences must be interpreted as unconscious inferences, as information processing, based on or accessing the background system. One might object to such an account on the grounds that not all justifying inference is explanatory and, consequently, be led to a more general account of coherence as successful competition based on a background system (BonJour, 1985; Lehrer, 1990). The belief that one sees a shape competes with the claim that one does not, with the claim that one is deceived, and other sceptical objections. The background system of belief informs one that one is trustworthy and enables one to meet the objections. A belief coheres with

a background system just in case it enables one to meet the sceptical objections and in that way justifies one in the belief. This is a standard strong coherence theory of justification (Lehrer, 1990).

It is easy to illustrate the relationship between positive and negative coherence theories in terms of the standard coherence theory. If some objection to a belief cannot be met in terms of the background system of beliefs of a person, then the person is not justified in that belief. So, to return to Trust, suppose that she has been told that a warning light has been installed on her gauge to tell her when it is not functioning properly and that when the red light is on, the gauge is malfunctioning. Suppose that when she sees the reading of 105, she also sees that the red light is on. Imagine, finally, that this is the first time the red light has been on, and, after years of working with the gauge, Trust, who has always placed her trust in the gauge, believes what the gauge tells her, that the liquid in the container is at 105 degrees. Though she believes what she reads, her belief that the liquid in the container is at 105 degrees is not a justified belief because it fails to cohere with her background belief that the gauge is malfunctioning. Thus, the negative coherence theory tells us that she is not justified in her belief about the temperature of the contents in the container. By contrast, when the red light is not illuminated and the background system of Trust tells her that under such conditions that gauge is a trustworthy indicator of the temperature of the liquid in the container, then she is justified. The positive coherence theory tells us that she is justified in her belief because her belief coheres with her background system.

The foregoing sketch and illustration of coherence theories of justification have a common feature, namely, that they are what are called internalistic theories of justification (*see* EXTERNALISM/INTERNALISM). They are theories affirming that coherence is a matter of internal relations among beliefs and that justification is a matter of coherence. If, then, justification is solely a matter of internal relations between beliefs, we are left with the

possibility that the internal relations might fail to correspond with any external reality. How, one might object, can a completely internal subjective notion of justification bridge the gap between mere true belief, which might be no more than a lucky guess, and knowledge, which must be grounded in some connection between internal subjective conditions and external objective realities?

The answer is that it cannot and that something more than justified true belief is required for knowledge. This result has, however, been established quite apart from consideration of coherence theories of justification. What is required may be put by saying that the justification one has must be undefeated by errors in the background system of belief. A justification is undefeated by errors just in case any correction of such errors in the background system of belief would sustain the justification of the belief on the basis of the corrected system. So knowledge, on this sort of positive coherence theory, is true belief that coheres with the background belief system and corrected versions of that system. In short, knowledge is true belief plus justification resulting from coherence and undefeated by error (Lehrer, 1990). The connection between internal subjective conditions of belief and external objective realities results from the required correctness of our beliefs about the relations between those conditions and realities. In the example of Trust, she believes that her internal subjective conditions of sensory experience and perceptual belief are connected with the external objective reality of the temperature of the liquid in the container in a trustworthy manner. This background belief is essential to the justification of her belief that the temperature of the liquid in the container is 105 degrees, and the correctness of that background belief is essential to the justification remaining undefeated. So our background system of beliefs contains a simple theory about our relation to the external world which justifies certain of our beliefs that cohere with that system. For such justification to convert to knowledge, that theory must be sufficiently free from error so that the coherence is sustained in corrected versions of our background system

of beliefs. The correctness of the simple background theory provides the connection between the internal conditions and external realities.

The coherence theory of truth arises naturally out of a problem raised by the coherence theory of justification. The problem is that anyone seeking to determine whether she has knowledge is confined to the search for coherence among her beliefs. The sensory experiences she has are mute until they are represented in the form of some perceptual belief. Beliefs are the engine that pulls the train of justification. But what assurance do we have that our justification is based on true beliefs? What justification do we have that any of our justifications are undefeated? The fear that we might have none, that our beliefs might be the artifact of some deceptive demon or scientist, leads to the quest to reduce truth to some form, perhaps an idealized form, of justification (Rescher, 1973; Rosenberg, 1980). That would close the threatening sceptical gap between justification and truth. Suppose that a belief is true if and only if it is ideally justified for some person. For such a person there would be no gap between justification and truth or between justification and undefeated justification. Truth would be coherence with some ideal background system of beliefs, perhaps one expressing a consensus among belief systems or some convergence toward consensus. Such a view is theoretically attractive for the reduction it promises, but it appears open to profound objections. One is that there is a consensus that we can all be wrong about at least some matters, for example, about the origins of the universe. If there is a consensus that we can all be wrong about something, then the consensual belief system rejects the equation of truth with consensus. Consequently, the equation of truth with coherence with a consensual belief system is itself incoherent.

Coherence theories of the content of our beliefs and the justification of our beliefs themselves cohere with our background systems but coherence theories of truth do not. A defender of coherentism must accept the logical gap between justified belief and truth, but she may believe that her capacities suffice

to close the gap to yield knowledge. That view is, at any rate, a coherent one.

See also CONCEPTS; GETTIER PROBLEM; TRUTH; and FOUNDATIONS AND COHERENCE in Part I.

#### BIBLIOGRAPHY

- Audi, R.: *Belief, Justification and Knowledge* (Belmont, CA: Wadsworth, 1988).  
 Bender, J., ed.: *The Current State of the Coherence Theory* (Dordrecht: Kluwer, 1989).  
 Bonjour, L.: *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1985).  
 Chisholm, R. M.: *Theory of Knowledge*, 3rd edn (Englewood Cliffs, NJ: Prentice-Hall, 1989).  
 Goldman, A.: *Empirical Knowledge* (Berkeley, CA: University of California Press, 1988).  
 Harman, G.: *Thought* (Princeton, NJ: Princeton University Press, 1973).  
 Lehrer, K.: *Theory of Knowledge* (Boulder, Colo.: Westview, 1990).  
 Lycan, W. G.: *Judgement and Justification* (New York: Cambridge University Press, 1988).  
 Pollock, J.: *Contemporary Theories of Knowledge* (Totowa, NJ: Rowman and Littlefield, 1986).  
 Rescher, N.: *The Coherence Theory of Truth* (Oxford: Oxford University Press, 1973).  
 Rosenberg, J.: *One World and Our Knowledge of It* (Dordrecht: Reidel, 1980).

KEITH LEHRER

**collective belief** In everyday speech we often refer to the beliefs of a group of people. We say such things as "The union believes that a strike would succeed", and "In the opinion of the government war is inevitable". How are such claims to be interpreted? Do they imply the existence of a "group mind"?

Many assume that a simple summative (or aggregative) analysis is correct. That is, they assume that a group believes that *p* if and only if all or at least most members of the group personally believe that *p* (see e.g. Quinton, 1975, pp. 9, 17).

A related analysis which has been considered is this: a group believes that *p* if and only if all or most members believe that *p* and this is common knowledge in the group (Gilbert, 1987, 1989; see also Bach and

Harnish, 1979, p. 270). ("Common knowledge" is a technical term from Lewis, 1969.)

Neither of these analyses copes well with cases such as the following. A committee has to reach a view on some matter, for instance, whether taxation should be increased. Only one individual believes that taxation should be increased, but he is feared by the others and they "go along" with him, voting in favour of increasing taxes. It seems that we can now say that in the committee's view taxation should be increased. If this is so, then an analysis of group belief in terms of what most members believe cannot be correct for all cases.

An alternative analysis has been proposed according to which, roughly, a group believes something if and only if the members understand that they are jointly committed to uphold the belief in question as a body. Thus when speaking as members of a committee individuals must express the committee view rather than their personal belief. (For details, see Gilbert, 1987, 1989.)

In so far as group belief in this sense is a common phenomenon, empirical investigation will need special sensitivity to determine whether an individual is expressing a group or a personal belief on a particular occasion, and the epistemology of group beliefs in this sense will require special investigation.

See also BELIEF.

#### BIBLIOGRAPHY

- Bach, K. and Harnish, R. M.: *Linguistic Communication and Speech Acts* (Cambridge, MA: MIT Press, 1979).  
 Gilbert, M.: "Modelling Collective Belief," *Synthese* 73 (1987), 185–204.  
 Gilbert, M.: *On Social Facts* (London and New York: Routledge, 1989).  
 Lewis, O. K.: *Convention* (Cambridge, MA: Harvard University Press, 1969).  
 Quinton, A.: "Social objects," *Proceedings of the Aristotelian Society* 75 (1975), 1–27.

MARGARET GILBERT

#### **commonsensism and critical cognitivism**

Commonsensism is the view that we know

most, if not all, of those things which ordinary people think they know and that any satisfactory epistemological theory must be adequate to the fact that we do know such things. Defenders of commonsense, such as Thomas Reid (see REID) and G. E. Moore (see MOORE), maintain, for example, that we know that there are material or physical objects having shape and size in three dimensions, that there are other people who think and feel and have bodies, and that we and other things, such as the earth, have existed for many years. Commonsensism claims it is more reasonable for us to hold these particular commonsense beliefs than any epistemological theory which implies that we do not know such things. In this respect, commonsensism is incompatible with, and rejects, various forms of scepticism (see SCEPTICISM) and certain traditional forms of empiricism (see EMPIRICISM), such as those represented by Berkeley (see BERKELEY) and Hume (see HUME). In this spirit, Moore writes:

There is no reason why we should not, in this respect, make our philosophical opinions agree with what we necessarily believe at other times. There is no reason why I should not confidently assert that I do really know some external facts, although I cannot prove this assertion except by simply assuming that I do. I am, in fact, as certain of this as anything; and as reasonably certain of it.

(Moore, 1922, p. 162)

While commonsensism claims we can be confident that any theory is mistaken that implies that we do not know that there are material objects or facts about other minds, philosophers such as Reid and Moore are also astute critics, carefully pointing out mistaken assumptions which generate sceptical conclusions. Reid's criticism of the empiricist theory of ideas is an especially noteworthy example.

Some philosophers claim that we know things about physical objects and other minds, but attempt to reduce, translate or analyse statements about the "external world" and other minds into statements of a "less problematic" variety. For example, some claim that we do know that there are tables and chairs,



but the statement "I am seated at a table" can be translated or reduced to statements about one's present sense experience or about "permanent possibilities of sensations". This is the approach of Berkeley and the phenomenologists (see PHENOMENALISM). Similarly, some who concede that one can know, for example, that "Smith believes it will rain" attempt to reduce this statement into statements about Smith's having certain dispositions to behave in various ways (see RYLE; BEHAVIOURISM). Commonsensism rejects these forms of "reductionism" on the ground that they do not capture the meaning of what it is we think we know when we claim to know that there are tables or that someone believes it will rain. It may ask precisely *what* statements about our sense experiences or dispositions to behave express what we think we know about tables or someone's believing that it will rain. Moore writes:

Some philosophers seem to . . . use such expressions as, e.g., "The earth has existed for many years past", as if they expressed something which they really believed, when in fact they believe that every proposition, which such an expression would *ordinarily* be understood to express, is, at least partially, false . . . I wish, therefore, to make it quite plain that I was not using the expressions . . . in any such subtle manner. I meant by each of them precisely what every reader, in reading them, will have understood me to mean.

(1959, p. 36)

Though commonsensism holds that a satisfactory epistemological theory must be adequate to the fact that we know certain things about physical objects, other minds, and the past, it is not committed to any particular account of how we know such things or even to the view that we can formulate a satisfactory account of how we know such things. The defender of commonsense holds that he knows or is justified in believing certain things quite independently of his being able to say how he knows or his knowing some criterion of knowledge or justification. Moore expresses this view when he writes, "We are all, I think, in this strange position that we do *know* many things, with regard to which we

*know* further that we must have evidence for them, and yet we do not know *how* we know them" (1959, p. 44).

Some defenders of commonsense have tried to formulate "marks" of commonsense beliefs. Reid, for example, suggests that a commonsense belief is one (1) which is universally held by mankind, (2) whose acceptance is reflected in the common structure of all languages, (3) whose contradictory is not merely false, but absurd, and (4) that is irresistible, so that even those who question them are compelled to believe them when engaging in the practical affairs of life. But whatever the merits of Reid's attempts to identify the range of commonsense beliefs, it would be a mistake to take him to hold that we must first identify some mark or criterion of commonsense beliefs before they are evident to us.

"Critical cognitivism" is a term coined by Roderick Chisholm (see CHISHOLM) to refer to a certain approach to the problems of epistemology. According to Chisholm (1977, 1982), we may distinguish two sorts of epistemological questions. The first sort of question may be put, "*What* do we know?" or "What is the *extent* of our knowledge?" The second may be put, "How are we to decide, in any particular case, *whether* we know?" or "What are the *criteria* of knowing?" Philosophers who assume that they have an answer to the second sort of question and then try to work out an answer to the first, Chisholm calls "methodists". Those who assume that they have an answer to the first sort of question and then try to answer the second, he calls "particularists". Chisholm takes Locke and Hume to be methodists and Reid and Moore to be particularists. Critical cognitivism assumes that one can identify particular instances of knowledge, and given answers to the first sort of question develop an answer to the second, that one can formulate criteria of knowledge. (For further discussion of particularism and methodism, see PROBLEM OF THE CRITERION.)

Chisholm's version of critical cognitivism exemplifies the two features of commonsensism mentioned above. He assumes that we know certain facts about the material objects around

us, other minds, and the past, and that any satisfactory account of the criteria of knowledge must be adequate to this fact. Like commonsensism, it rejects certain forms of scepticism, traditional empiricism and reductionism. The critical cognitivist and defender of commonsense hold that scepticism and traditional empiricism “rest upon some premiss which is, beyond comparison, less certain than is the proposition which it is designed to attack” (Moore, 1922, p. 228). Critical cognitivism, however, goes beyond commonsensism in (a) offering a positive account of how we know commonsense propositions, and (b) holding that there are just four sources of our knowledge: external perception, memory, self-awareness or inner consciousness, and reason. It asserts (1) that we do know facts about the external world, other minds, and the past, (2) that we have no other sources of knowledge, and, therefore, (3) our knowledge of the external world, other minds, and the past is yielded by these four sources.

Critical cognitivists such as Reid and Chisholm hold that one sort of fact can function as a “sign” or evidence for a different sort of fact. For example, certain psychological facts about one’s sensations or the way one is “appeared to” can, under certain conditions, function as evidence for beliefs about external, physical objects. Critical cognitivism also holds that knowledge yielded by these four traditional sources can function as signs or evidence for certain “problematic” kinds of claims, such as those about ethical values and other minds. For example, Reid tells us “certain features of the countenance, sounds of the voice, and gestures of the body, indicate certain thoughts and dispositions of the mind” (Reid, 1785, Essay VI, ch. V). According to Reid, our knowledge of certain nonmental facts about bodies, gestures and voices can, under certain conditions, confer warrant upon our beliefs about other minds. Similarly, Chisholm writes, “My own feeling is a sign of the evil nature of ingratitude, and so it could be said to confer evidence upon the statement that ingratitude is evil” (Chisholm, 1977, p. 126). According to Chisholm, certain non-ethical facts can, under certain conditions,

confer warrant upon our ethical beliefs. These facts about our emotional experiences are known through inner consciousness and, though they are not themselves ethical facts, they can confer warrant on ethical beliefs.

The critical cognitivist claims that there are *only* four sources of knowledge. What is the status of this claim? If he assumes, *ab initio*, that there are just four sources, then it appears that, like the methodist, he is assuming some general principle about how we are to decide whether we know. “If some putative instance of knowledge can’t be accounted for in terms of just these four sources, then it isn’t knowledge.” However, there are two alternatives for the critical cognitivist. First, he might simply hold that we have no particular knowledge of any other source, that we have no experience, for example, of the operation of a faculty of moral or religious intuition. Alternatively, we may take the critical cognitivist not to assume from the start that there are just four sources, but rather to hold that since we can account for everything we ordinarily think we know in terms of these four sources, we have no reason to recognize any other source. Whether this conclusion is justified depends, of course, on whether these sources can accommodate everything it is reasonable for us to think we know. This is a point with which any particularist or critical cognitivist will agree.

See also CHISHOLM; MOORE; PROBLEM OF THE CRITERION; REID.

#### BIBLIOGRAPHY

- Chisholm, R. M.: *The Foundations of Knowing* (Minneapolis, MN: University of Minnesota Press, 1982).  
 Chisholm, R. M.: *Theory of Knowledge*, 2nd edn (Englewood Cliffs, NJ: Prentice-Hall, 1977).  
 Lemos, N.: *Common Sense: A Contemporary Defense* (Cambridge: Cambridge University Press, 2004).  
 Madden, E. H.: “The Metaphilosophy of Commonsense,” *American Philosophical Quarterly* 20 (1983), 23–36.  
 Moore, G. E.: *Philosophical Papers* (London: George Allen and Unwin, 1959).

- Moore, G. E.: *Philosophical Studies* (London: Routledge and Kegan Paul, 1922).
- Reid, T.: *Essays on the Intellectual Powers of Man* (1785), ed. B. Brody (Cambridge, MA: MIT Press, 1969).
- Reid, T.: *Thomas Reid's Inquiry and Essays*, ed. K. Lehrer and R.E. Beanblossom (Indianapolis, IN: Bobbs-Merrill, 1975).
- Sosa, E.: "The foundations of foundationalism," *Nous* 14 (1980), 547–64.

NOAH H. LEMOS

**concepts** Mental states have contents: a belief may have the content that I will catch the train, a hope may have the content that the prime minister will resign. A concept is something which is capable of being a constituent of such contents. More specifically, a concept is a way of thinking of something – a particular object, or property, or relation, or some other entity.

Several different concepts may each be ways of thinking of the same object. A person may think of himself in the first-person way, or think of himself as the spouse of Mary Smith, or as the person located in a certain room now. More generally, a concept *c* is distinct from a concept *d* if it is possible for a person rationally to believe "*c* is such-and-such" without believing "*d* is such-and-such". As words can be combined to form structured sentences, concepts have also been conceived as combinable into structured complex contents. When these complex contents are expressed in English by "that . . ." clauses, as in our opening examples, they will be capable of being true or false, depending on the way the world is.

Concepts are to be distinguished from stereotypes and from conceptions. The stereotypical spy may be a middle-level official down on his luck and in need of money. None the less we can come to learn that Anthony Blunt, art historian and Surveyor of the Queen's Pictures, is a spy; we can come to believe that something falls under a concept while positively disbelieving that the same thing falls under the stereotype associated with the concept. Similarly, a person's conception of a just arrangement for resolving disputes may involve something like contemporary Western

legal systems. But whether or not it would be correct, it is quite intelligible for someone to reject this conception by arguing that it does not adequately provide for the elements of fairness and respect which are required by the concept of justice.

A theory of a particular concept must be distinguished from a theory of the object or objects it picks out. The theory of the concept is part of the theory of thought and epistemology; a theory of the object or objects is part of metaphysics and ontology. Some figures in the history of philosophy – and perhaps even some of our contemporaries – are open to the accusation of not having fully respected the distinction between the two kinds of theory. Descartes appears to have moved from facts about the indubitability of the thought "I think", containing the first-person way of thinking, to conclusions about the non-material nature of the object he himself was. But though the goals of a theory of concepts and a theory of objects are distinct, each theory is required to have an adequate account of its relation to the other theory. A theory of concepts is unacceptable if it gives no account of how the concept is capable of picking out the objects it evidently does pick out. A theory of objects is unacceptable if it makes it impossible to understand how we could have concepts of those objects.

A fundamental question for philosophy is: what individuates a given concept – that is, what makes it the one it is, rather than any other concept? One answer, which has been developed in great detail, is that it is impossible to give a non-trivial answer to this question (Schiffer, 1987). An alternative approach, favoured by the present author, addresses the question by starting from the idea that a concept is individuated by the condition which must be satisfied if a thinker is to possess that concept and to be capable of having beliefs and other attitudes whose contents contain it as a constituent. So, to take a simple case, one could propose that the logical concept *and* is individuated by this condition: it is the unique concept *C* to possess which a thinker has to find these forms of inference compelling, without basing them on any further inference or information:

from any two premisses A and B, ACB can be inferred; and from any premiss ACB, each of A and B can be inferred. Again, a relatively observational concept such as *round* can be individuated in part by stating that the thinker finds specified contents containing it compelling when he has certain kinds of perception, and in part by relating those judgements containing the concept and which are not based on perception to those judgements that are. A statement which individuates a concept by saying what is required for a thinker to possess it can be described as giving the *possession condition* for the concept.

A possession condition for a particular concept may actually make use of that concept. The possession condition for *and* does so. We can also expect to use relatively observational concepts in specifying the kind of experiences which have to be mentioned in the possession conditions for relatively observational concepts. What we must avoid is mention of the concept in question as such within the content of the attitudes attributed to the thinker in the possession condition. Otherwise we would be presupposing possession of the concept in an account which was meant to elucidate its possession. In talking of what the thinker finds compelling, the possession conditions can also respect an insight of the later Wittgenstein (see WITTGENSTEIN): that a thinker's mastery of a concept is inextricably tied to how he finds it natural to go on in new cases in applying the concept.

Sometimes a family of concepts has this property: it is not possible to master any one of the members of the family without mastering the others. Two of the families which plausibly have this status are these: the family consisting of some simple concepts *0*, *1*, *2*, . . . of the natural numbers and the corresponding concepts of numerical quantifiers *there are 0 so-and-so's*, *there is 1 so-and-so*, . . . ; and the family consisting of the concepts *belief* and *desire*. Such families have come to be known as "local holisms". A local holism (see HOLISM) does not prevent the individuation of a concept by its possession condition. Rather, it demands that all the concepts in the family be individuated simultaneously. So one

would say something of this form: *belief* and *desire* form the unique pair of concepts *C1* and *C2* such that for a thinker to possess them is to meet such-and-such condition involving the thinker, *C1* and *C2*. For these and other possession conditions to individuate properly, it is necessary that there be some ranking of the concepts treated. The possession conditions for concepts higher in the ranking must presuppose only possession of concepts at the same or lower levels in the ranking.

A possession condition may in various ways make a thinker's possession of a particular concept dependent upon his relations to his environment. Many possession conditions will mention the links between a concept and the thinker's perceptual experience. Perceptual experience represents the world as being a certain way. It is arguable that the only satisfactory explanation of what it is for perceptual experience to represent the world in a particular way must refer to the complex relations of the experience to the subject's environment. If this is so, then mention of such experiences in a possession condition will make possession of that concept dependent in part upon the environmental relations of the thinker. Burge (1979) has also argued from intuitions about particular examples that, even though the thinker's non-environmental properties and relations remain constant, the conceptual content of his mental state can vary if the thinker's social environment is varied. A possession condition which properly individuates such a concept must take into account the thinker's social relations, in particular his linguistic relations.

Concepts have a normative dimension, a fact strongly emphasized by Kripke. For any judgement whose content involves a given concept, there is a *correctness condition* for that judgement, a condition which is dependent in part upon the identity of the concept. The normative character of concepts also extends into the territory of a thinker's reasons for making judgements. A thinker's visual perception can give him good reason for judging "*That man is bald*"; it does not by itself give him good reason for judging "*Rostropovich is bald*", even if the man he sees is Rostropovich. All these normative connections must be

explained by a theory of concepts. One approach to these matters is to look to the possession condition for a concept, and consider how the referent of the concept is fixed from it, together with the world. One proposal is that the referent of the concept is that object (or property, or function . . .) which makes the practices of judgement and inference mentioned in the possession condition always lead to true judgements and truth-preserving inferences. This proposal would explain why certain reasons are necessarily good reasons for judging given contents. Provided the possession condition permits us to say what it is about a thinker's previous judgements that makes it the case that he is employing one concept rather than another, this proposal would also have another virtue. It would allow us to say how the correctness condition is determined for a judgement in which the concept is applied to a newly encountered object. The judgement is correct if the new object has the property which in fact makes the judgemental practices mentioned in the possession condition yield true judgements, or truth-preserving inferences.

See also COHERENTISM; LINGUISTIC UNDERSTANDING.

#### BIBLIOGRAPHY

- Burge, T.: "Individualism and the Mental," *Midwest Studies in Philosophy* 4 (1979), 73–121.  
 Frege, G.: "The Thought," in *Logical Investigations*, ed. P. Geach (Oxford: Blackwell, 1977).  
 Kripke, S.: *Wittgenstein on Rules and Private Language* (Oxford: Blackwell, 1982).  
 Peacocke, C.: "What Are Concepts?," *Midwest Studies in Philosophy* 14 (1989), 1–28.  
 Schiffer, S.: *Remnants of Meaning* (Cambridge, MA: MIT Press, 1987).

CHRISTOPHER PEACOCKE

**Continental epistemology** For the purposes of this entry, Continental epistemology will be defined as the cluster of problems concerning knowledge in French and German philosophy of the nineteenth and twentieth centuries. This time frame can be justified on

the grounds that with Hegel a decidedly different orientation to knowledge began to emerge.

Problems of knowledge and truth are as central to the continental tradition as they are to the Anglo-American tradition, but because they start from a different place different sorts of questions arise. This different starting point is Hegel's perception of the historical dimension of knowledge (see HEGEL). Hegel was influenced by Kant's critique of the limits of reason and his recognition of the subjective input into knowledge (see KANT). But Hegel drew two conclusions from these claims that Kant did not draw: the fundamental limit to reason is its embeddedness in a historical context, which further suggests that the categories of interpretation Kant identified must themselves be understood as historically situated and thus limited. It is this insight into the historical nature of reason, knowledge and even truth that might be said to begin a different trajectory of development in epistemology on the European continent.

The claim of historicity follows simply from the fact that there is a subjective component to knowledge and that subjects are historical creatures, incapable of fully transcending their historical location. For Hegel, a particular, such as a singular historical moment, cannot contain the universal, though it exemplifies a moment of the universal. This means that a belief can be both partial and true simultaneously, and that truth needs to be understood as indexed to the perspective of a history and a culture. Furthermore, on Hegel's view the locus of belief formation cannot be meaningfully understood as an isolated individual. The individual, as a particular, is also an exemplification of a universal, the spirit of an age, for example. The individual's beliefs must therefore be understood within this context. These two shifts – from the ahistorical to the historical and from the individual to the culture as a whole – produced a turn in the way in which epistemology has been done in France and Germany since Hegel's time. Neither claim has been superseded, and the continental schools of thought all develop further and in diverse ways these fundamental insights.



Where for Descartes the problem was how to get to the Real beyond mere sensory ideas, for Hegel the problem was how to get to knowledge through historical location. And whereas Descartes posited a unified, self-knowing subject essentially separate from reality and linked to it only by potentially deceiving sensory organs, Hegel's subject is in part an effect of the Real, as a partial representative of the Absolute Spirit. So the question whether the subject can pass over an interminable gulf between its consciousness and the object world does not arise; for Hegel subject and object, particular and universal, and appearance and essence are ontologically connected at their core. The essence of a thing shines forth in its appearance, just as the universal manifests itself only in the particular. The problem of scepticism (*see* SCEPTICISM) thus never arises.

The key epistemic problem for Hegel is not how can we know anything at all, but how can we move from a partial and limited historical perspective to absolute knowledge. This move is necessary because it is only from the perspective of absolute knowledge that we can judge the validity of claims to partial truth, since the criterion of truth is ultimately the universal, or the ability of a particular to be subsumed within the universal synthesis of all truths. Hegel's answer here further reveals the differences between continental and Anglo-American epistemology. His solution is not a methodology or set of procedures which a subject can utilize, since no methodology can permit one to become ahistorical or perspectiveless, but rather history itself – the dialectical movement of the rational Real in history through social revolutions, philosophical reflection, and the subsequent development of consciousness. The endpoint and highest level of this consciousness is the attainment of absolute knowledge, defined as an all-inclusive synthesis of the whole.

Marxist epistemology (*see* MARXISM) offers a materialist account of historical perspective, in which the spirit of an age expressed in knowledge arises fundamentally from economic practices. And where Hegel historicizes Kant's categories, Marx politicizes

them, by connecting them to class interests. His concept of ideology (*see* IDEOLOGY) offers an analysis of how power relations affect belief formation as well as the constitution of identities and subjective experience. But like Hegel, Marx posits an end to history: through concrete political struggle we can exert control over the conditions of knowledge production and achieve an absolute synthesis involving political, epistemic and historical dimensions simultaneously.

In Nietzsche (*see* NIETZSCHE) one gets Hegelian perspectivism for the first time without an absolute end and without a dialectic progression. Knowledge for Nietzsche is again material, but now based on desire and bodily needs more than social labour. Perspectives are to be judged not from their relation to the absolute but on the basis of their effects in a specific era. The possibility of any truth beyond such a local, pragmatic one becomes a problem in Nietzsche, since neither a noumenal realm nor an historical synthesis exists to provide an absolute criterion of adjudication for competing truth claims: what get called truths are simply beliefs that have been held for so long that we have forgotten their genealogy. In this Nietzsche reverses the Enlightenment dictum that truth is the way to liberation by suggesting that truth claims, in so far as they are considered absolute and transhistorical, shut down the possibilities for debate and conceptual progress and thus cause rather than alleviate backwardness and unnecessary misery. Nietzsche moves back and forth without resolution between the positing of transhistorical truth claims, such as his claim about the will to power, and a kind of epistemic nihilism that calls into question not only the possibility of truth but the need and desire for it as well. But perhaps most importantly, Nietzsche introduces the notion that truth is a kind of human practice, a move in a game whose rules are contingent rather than necessary. The evaluation of truth claims should be based on their strategic effects, not their ability to represent a reality conceived of as separate and autonomous of human influence.

In the twentieth century five major orientations to epistemology developed in different

ways from this point: phenomenology, critical theory, hermeneutics, post-structuralism, and feminism. Phenomenology (*see* PHENOMENOLOGY) in some ways returned to the Cartesian project of bracketing prejudgements to gain a direct awareness of being. But the objectivist conception of the Real found in Descartes is rejected as ontic and replaced with an ontological conception which conceives of the object in its relation to consciousness, thus rejecting a subject/object split. The object of knowledge is not a thing in itself but the thing for-me and my lived experience (*see* IN ITSELF/FOR ITSELF). Sartre offers a notion of the world as a system of objects where meanings and values are organized in light of the individual's project. Existentialist philosophy in general asserted the reality of the particular over the universal, which meant that the criterion of truth no longer required the capacity to be subsumed within a universal. The problem of subjectivism (*see* SUBJECTIVISM) this might seem to entail was avoided by reorienting the ontology of truth from the ontic to the ontological: from the subjectless object to lived experience and a Sartrean world. Subjectivism could further be avoided by a collective notion of the subject, as Sartre tried to develop in his later work (*see* SARTRE).

Philosophical hermeneutics (*see* HERMENEUTICS) can be viewed as an epistemological fellow-traveller of phenomenology to the extent it continues the focus on the subjective input into knowledge and seeks to break down the subject/object separation found in both Cartesianism and positivism. Gadamer's hermeneutics (*see* GADAMER), for example, explores the way in which a knower's "horizon", made up of their prejudgements, historical location and conceptual tradition, both limits and enables understanding. Truth, on his view, is a kind of emergent property of a dialogic interaction between reader and text, knower and object, each of which come to the interaction with a kind of horizon. Truth occurs when the interaction produces a fusion. One primary effect of this is to reorient the notion of truth as a property of propositional content to an event with temporal specificity.

The Frankfurt School also offered a critique of the ontology of the knowing process found in Cartesian-based epistemologies, but Adorno and Horkheimer's main target was what they called instrumental reason and the Enlightenment belief in the necessary connection between knowledge and freedom. Instrumental reason allows only for means-end calculations and posits itself as politically neutral yet the best route to liberation. Adorno and Horkheimer's major insight was to see this objectivist stance as founded in the desire for mastery and implicated in the practices of domination within the West. For Horkheimer, instead of a distanced, disinterested position over the object, which simply serves to conceal the values and interests of the knower, we should see ourselves as embedded within social locations and understand reality as the product of an interaction between society and nature. For Adorno, what is needed urgently from within the dialectic of Enlightenment reason is an immanent, negative critique which will disrupt the systematically circular confirmations of instrumental reason, rearticulate and reinvigorate Enlightenment ideals, and thus advance the dialectic. Their concept of critical theory was a continuation of Kant's immanent critique of reason, but this time through revealing the artificially constrained concept of reason under commodity production and the ideological and non-rational elements at work in justification strategies in this era.

Habermas's project as the heir apparent of critical theory (*see* HABERMAS) has been to foreground questions of epistemology raised by the movement's work, in particular, the nature of the epistemic basis needed to justify the claims of the critical theorists themselves. He has rejected the possibility of an interest-free reason, but argued that interests can themselves be rationally evaluated. But his main work has been to suggest a new communication-based model for epistemic evaluation. Inquiry occurs in an intersubjective context in which the goal is better understood as mutual agreement than knowledge of an object. The pragmatic aspects of this context, such as the power relations among the participants, are salient with respect to the

epistemic validity of its outcome. Unequal power relations can inhibit participants' input into the discussion, thus distorting the resulting conclusions. Habermas's analysis here effectively integrates epistemological and political considerations. This provides a new perspective on Marx's concept of ideology and a new proposal about how we can enact the Enlightenment project of achieving liberation through knowledge.

If the above three movements have in common an attempt to develop a new synthesis of subject and object, post-structuralism returns us to a Nietzschean scepticism towards the possibility of liberation through reason. The term post-structuralism refers to a cluster of writings in reaction against (but in some ways a development of) the structuralist work of Saussure and Lévi-Strauss. Where phenomenology was concerned with what occurs in experience, structuralism was interested in the conditions which make that experience possible. These conditions were theorized as existing in the structural relations between constituents of experience rather than in relations of reference between those constituents and an outside world. Post-structuralism continues this turn away from reference as the prime criterion of truth but rejects the positing of universal and ahistorical structures. Though very disparate, post-structuralists are probably united in the belief that knowledge systems are ultimately contingent and connected intrinsically to power relations and to desire. In this and in their conceptualization of truth as a kind of strategic, linguistic practice post-structuralists are influenced most importantly by Nietzsche.

For Derrida (*see* DERRIDA), Western epistemology is logocentric in the sense that it assumes the neutrality and transparency of knowledge and of language. Language works precisely because it is not transparent, because it represses those elements which make meaning possible. Derrida's main project has been to reveal the ways in which knowledge is dependent on its other. Beliefs are justified or epistemically validated not because they correspond to a transcendental signified or a reality outside or behind the text, but

through the always plural and shifting, deferred play of interrelationships between textual elements. Justification is therefore inherently unstable and ultimately undecidable. Foucault (*see* FOUCAULT) has similarly been concerned to reveal the "positive (or constitutive) unconscious of knowledge", though he theorizes this primarily in terms of desire and power rather than the endless deferral of signification. Foucault emphasizes much more than Derrida the level of materiality at play in the web of elements which produce knowledge, a materiality involving bodies and pleasures and non-linguistic practices as well as linguistic ones.

For both Derrida and Foucault, the outcome of knowledge production is contingent and even arbitrary in an epistemic sense. On Derrida's view, the binary terms through which meaning operates are necessarily in a relation of hierarchy but the choice of which term is dominant is essentially arbitrary. In a similar vein, Foucault sees regimes of truth as the product of as much historical accident as identifiable causes. Given the underdetermination of epistemic causes or reasons, belief formation, strategies of justification and epistemologies themselves must be explained in significant measure by reference to non-rational forces.

It is unclear how post-structuralism understands truth (*see* TRUTH). One possibility is to construct a meta-level analysis, in which justification works internal to a discourse, but where the post-structuralist account of discourses themselves have a different epistemic status by operating at a different level. But this position would seem to be self-refuting to the extent that it accords a degree of transparency to post-structuralist discourses that they maintain is impossible. Another option is to see truth as only a kind of validating but redundant attribution, similar to Tarski's view. But post-structuralists tend to see truth as more substantive, as marking a kind of conceptual domination, a claiming of hegemony, and effecting a curtailment of conflicting discursive moves. This suggests that truth is a kind of strategy within a language-game, and that as a criterion of justification the true is judged in terms of its strategic effects or

the configuration of possibilities a claim makes possible, rather than by its correspondence to a non-linguistic realm. This strategic account of truth is the position adopted most often by post-structuralists, in both giving an account of how truth operates in discourse and in explaining the status of their own claims.

French feminist philosophy has developed yet another approach to knowledge building from this point. Irigaray suggests that there is a relation of isomorphism between the validity conditions and a priori assumptions of Western knowledge and male subjectivity. The impulse to achieve non-fallible knowledge from a position which is itself conceived of as disinterested is a desire for a kind of mastery which is essentially phallic. And the post-structuralist critique of logocentrism has also been largely gender-blind, completely inattentive to the maleness of both the interests involved in and the linear style of knowledge production in the West. The power and desire post-structuralists take to be formative in knowledge is not sexually neutral but masculine.

Continental epistemology is thus a heterogeneous terrain, yet the following list of family resemblances within this terrain may prove useful:

1. There is a general rejection of the Cartesian framing of epistemology (see CARTESIANISM) as concerned with scepticism and with divining an accessible criterion of justification a believer can use to improve the epistemic status of their beliefs.
2. There is a turn away from focus on propositional content and the representational and referential status of claims to their interrelationships.
3. Connected to this, the ontology of knowing shifts from a separated subject and object, where the object world is set up as transcendent of subjectivity, to a "lived world", an intersubjective dialogic realm, or a discursively constructed world.
4. Assumptions about the power and neutrality of reason are replaced with a critique of reason as ideological, dominating and

limited and an attention to the unconscious, irrational forces at work in belief formation.

5. Truth is reconceptualized as perspectival, historical, indexed to a spatio-temporal context and plural.
6. A primary concern is with the politics of truth: what is the relationship between knowledge and liberation? What are the power relations of knowledge production in our society? And how are the marks of epistemic justification and ontologies of truth connected to masculinity?

See also DEATH OF EPISTEMOLOGY; EPOCHE; FEMINIST EPISTEMOLOGY; HEGEL; HISTORICISM; HUSSERL; IN ITSELF/FOR ITSELF; NIETZSCHE.

#### BIBLIOGRAPHY

(Asterisked titles contain further bibliographic information on this topic)

- Adorno, T. and Horkheimer, M.: *Dialectic of Enlightenment*, trans. J. Gunning (New York: Continuum, 1987).
- Baynes, K., Bohman, J. and McCarthy, T., eds: *After Philosophy: End or Transformation* (Cambridge, MA: MIT Press, 1987).\*
- Bernstein, R.: *Beyond Objectivism and Relativism* (Philadelphia, PA: University of Pennsylvania Press, 1983).
- Dreyfus, H. and Rabinow, P.: *Michel Foucault: Beyond Structuralism and Hermeneutics*, 2nd edn (Chicago, IL: University of Chicago Press, 1983).
- Grosz, E.: *Sexual Subversions: Three French Feminists* (Sydney: Allen and Unwin, 1989).\*
- Heidegger, M.: *Being and Time*, trans. J. Macquarrie (New York: Harper and Row, 1962).
- Horkheimer, M.: *Critical Theory*, trans. M. J. O'Connell (New York: Continuum, 1972).
- Irigaray, L.: *Speculum of the Other Woman*, trans. G. Gill (Ithaca, NY: Cornell University Press, 1985).
- Lyotard, J.-F.: *The Postmodern Condition: A Report on Knowledge*, trans. G. Bennington and B. Massumi (Minneapolis, MN: University of Minnesota Press, 1985).
- Martin Alcoff, L.: *Real Knowing: New Versions of Coherence Epistemology* (Ithaca, NY: Cornell University Press, 1996).
- Megill, A.: *Prophets of Extremity* (Berkeley, CA: University of California Press, 1985).

Rorty, R.: *Philosophy and the Mirror of Nature* (Princeton, NJ: Princeton University Press, 1979).

Rouse, J.: *Knowledge and Power* (Ithaca, NY: Cornell University Press, 1987).

Sartre, J.-P.: *Being and Nothingness*, trans. H. Barnes (New York: Simon and Schuster, 1956).

LINDA ALCOFF

**convention** There are two principal ways in which words or signs are given meaning, “by ostension” and “by convention”.

In *meaning by ostension* words or signs of a language are assigned entities as their meaning, e.g. mental entities such as ideas, experiences, concepts or non-mental entities such as concrete things, classes, functions, universals. *Generally speaking*, the entities assigned are so chosen that it will make sense to ask “Is *s* true (of the world)?” for sentences *s* in the language at issue, as when, for example, in “Cynthia’s ring is gold”, “Cynthia’s ring” is assigned Cynthia’s ring, and “is gold” is assigned a function from each and every thing in the world to *True* or *False* according as that thing is gold or not. Solely for our convenience here, let us call this sort of truth *O-truth*.

In *meaning by convention* words or signs of a language have meaning by virtue of more or less explicit rules for the use of the words or signs in relation to one another. The meaning of a word in this sense will not lie in an entity assigned to it, but will lie wholly in its rule-governed use with respect to other words or signs. *Generally speaking* (but not always), it is supposed that the language has a logical or logic-like syntax and that the rules for the use of the words or signs are rules for logical or logic-like operations (e.g. those generating formal-logical “deductions” from given sentences or formulas). Then, given such rules for a language, meanings for words are instituted “by convention” by specifying a subset *X* of the sentences or formulas in the language. Relative to the given rules, the meanings of the signs in the sentences or formulas in the set *X* are implicitly defined (see DEFINITION) through *X* and the rules.

*Example:* Language consists of signs *x*, *y*, 0, 1, 2, = as a binary relation sign, + as a binary operation sign. The formulas will all be of the form  $\_ + \_ = \_$ , or  $\_ + \_ = \_ + \_$ , or  $\_ = \_ + \_$  where we distribute *x*, *y*, 0, 1, 2 in all possible ways over the  $\_$ ’s. The rules are:

- (1) Wherever *x* occurs, you may substitute 0, 1, or 2.
- (2) If  $W = Z$  occurs, you may substitute *Z* for *W* whenever *W* occurs in the position  $\_ =$  or  $\_ = \_$ .

Here are the formulas *X* that, together with the rules, implicitly define the signs *x*, *y*, 0, 1, 2, +, = :  $x + 0 = x$ ,  $x + y = y + x$ . So what, for example, does 0 mean? It means, among other things, that  $x + 0 = x$ ,  $0 + 0 = 0$ ,  $1 + 0 = 1$ ,  $2 + 0 = 2$ ,  $0 + x = x$ ,  $0 + 1 = 1$ , . . . (see AXIOMATIZATION, AXIOMATICS).

Since the subset of formulas or sentences *X* which are settled on do implicitly define the words or signs composing them, it is said that they are true by convention – since the *X* determine the meanings of the words or signs composing them, we could say that they are true by virtue of the meanings of words or signs composing them. It is also usually said that the formulas or sentences derived from the *X* by exercise of the given rules for the uses of signs are also true by convention. Solely for our convenience here, let us call this sort of truth *C-truth*.

We have formal-logical syntaxes rich enough to enable us to formal-logically regiment the languages of the sciences. At least for the thus regimented languages of the natural sciences, we would expect most of the words occurring in a scientific theory to have two components of meaning, one determined by ostension, the other by convention. Thus some of the sentences (statements, propositions, theorems, etc.) of such a theory may be true by convention (*C-true*), some may be *O-true*, some *O-false*. However, on some views, some of the words or signs composing the language of our scientific theory may not have ostensive meanings, e.g. this is sometimes held of the peculiarly logical words such as “not”, “or”, “all”, and sometimes it is also held of mathematical words, and



also of “theoretical terms” rather than “observational terms” when it is thought that sense-experience is the only source of O-meaning and that we need expressions in scientific theory (theoretical terms) that cannot be given such an observational meaning.

Since sentences whose words have C-meaning but no O-meaning might logically imply sentences whose terms do have an O-meaning (as when only observational terms are given an O-meaning), and since, say, the implicit and explicit definitions in a theory may be variously chosen so that sentences are true by convention on one choice, but not on another, and yet the same sentences have O-truth under either choice, and since it is not inconceivable that some subject matters, such as logic and even mathematics may be only a matter of C-truth, there is much philosophical difficulty in the exact characterization of C-truth, O-truth, and the relation between them.

#### BIBLIOGRAPHY

- Quine, W. V.: “Truth by Convention,” in his *The Ways of Paradox* (Cambridge, MA: Harvard University Press, 1976), ch. 9.
- Quine, W. V.: “Two Dogmas of Empiricism,” in his *From a Logical Point of View* (Cambridge, MA: Harvard University Press, 1953), ch. 2.
- Russell, B.: *Our Knowledge of the External World* (London: Allen and Unwin, 1980), chs 4, 5.
- Torretti, R.: *Philosophy of Geometry from Riemann to Poincaré* (Dordrecht: Reidel, 1978), ch. 4, esp. 4.4.

ROBERT S. TRAGESSER

**conversational implicature** An inference which one may legitimately draw upon hearing an assertion, given that one is entitled to assume certain things about the agent making the assertion: for instance, that the agent is honest, co-operative and (at least otherwise) rational. Suppose I asked you where my umbrella was and you answered, “It could be in the bedroom cupboard”. As long as I can legitimately make these assumptions about you, I am entitled

to infer that you do not know where my umbrella is. Although you have not actually said that you do not know where my umbrella is, your assertion nevertheless carries this implicature, for if you did know where it was, then – given that you were being honest and co-operative – you would have simply told me.

The standard picture of how conversational implicature functions is due to H. P. Grice (1975) (see GRICE). Grice identified a number of conversational maxims which govern the generation of conversational implicatures, and which all fall under a general principle of conversational exchange which he refers to as the “Co-operation Principle”. He describes this principle as follows: “Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged” (Grice, 1975: 45). The maxims which fall under this general principle include, for example, the super-maxim of quality, “Try to make your contribution one that is true”, of which there are two sub-maxims, “Do not say what you believe to be false”, and “Do not say that for which you lack adequate evidence” (ibid.).

The primary importance of this pragmatic story that Grice offers is that it enables us to explain why certain assertions may well be true even though they seem improper. The reason for this is that if an assertion generates a false conversational implicature, then it will be judged to be inappropriate even if it is true. To return to the example given a moment ago, suppose in response to the question of where my umbrella is you merely say that it could be in the bedroom cupboard even though you know full well that it's in the hallway. What you say will be judged inappropriate even if it is true because it generates the false conversational implicature that you don't know where my umbrella is. Given that, in this way, assertions can be true but nevertheless inappropriate, it follows that one cannot simply infer from the impropriety of an assertion that what is asserted is false.

This point is relevant to epistemology in a number of ways. Consider, for example, its

relevance to the epistemic externalism/internalism distinction (see EXTERNALISM/INTERNALISM). A standard requirement placed on knowledge possession by the internalist is that knowing agents must be able to offer good reasons in favour of their beliefs, while the typical externalist, in contrast, will allow that at least sometimes agents can have knowledge even while lacking grounds of this sort, just so long as they meet other epistemic conditions such as a reliability condition. With this contrast in mind, it will be inevitable that externalists will sometimes ascribe knowledge to agents who, since they lack good grounds which they can offer in favour of their belief, would be unable appropriately to ascribe that knowledge to themselves since the assertion would offend against one of the conversational sub-maxims. This is because in making such an assertion the agent would be generating the false conversational implicature that she is able to offer grounds in favour of her belief. The externalist could thus, in principle at least, use the Gricean account of conversational implicature in order to explain away the apparent falsehood of claims to know where one lacks grounds – an appearance which, if not accounted for, would provide strong *prima facie* grounds for internalism.

A second epistemological application of conversational implicature can be found in recent work on scepticism (see SCEPTICISM), especially as regards the response to scepticism offered by attributor contextualism (see CONTEXTUALISM in Part I). In essence, the attributor contextualist maintains that “knows” is a *context-sensitive* term: An assertion of the form “S knows that *p*” can be true when uttered in one conversational context where the epistemic standards in play are loose, yet false when uttered in a second conversational context where the epistemic standards in play are strict. Appealing to the context-sensitivity of “know”, argues as follows: In contexts in which sceptical hypotheses are salient, the standards of knowledge rise to a level that is difficult to meet. Thus, we are disinclined to attribute knowledge in such contexts. However, in everyday contexts in which our standards of knowledge are more or less loose, we are no longer

reluctant to ascribe knowledge. In this way, both sceptical and anti-sceptical intuitions are accounted for.

Conversational implicature poses a potential challenge to this picture. An alternative, non-contextualist explanation of our readiness or reluctance to ascribe knowledge, depending on which context we are in, could go as follows: An assertion which ascribes knowledge in a sceptical conversational context generates *different* conversational implicatures from an assertion which ascribes that same knowledge in a non-sceptical conversational context. According to this alternative approach, it could be that both assertions are true. However, the assertion in the sceptical conversational context *seems* false because it generates false conversational implicatures, unlike that same assertion when made in an everyday conversational context. (See DeRose, 2002, for a defence of contextualism in the light of a proposal of this sort.)

See also GRICE.

#### BIBLIOGRAPHY

- DeRose, K.: “Knowledge, Assertion, and Context,” *Philosophical Review* 111 (2002), 167–203.  
 Grice, H. P.: “Logic and Conversation,” in *Syntax and Semantics: Speech Acts*, ed. P. Cole and J. L. Morgan (New York: Academic Press, 1975).

DUNCAN PRITCHARD

**criteria and knowledge** Except for alleged cases of things that are evident for one just by being true, it has often been thought, anything that is known must satisfy certain criteria as well as being true (see CRITERION, CANON). These criteria are general principles specifying what sorts of considerations *C* will make a proposition *p* evident to us.

Traditional suggestions include: (a) if a proposition *p*, e.g. that  $2 + 2 = 4$ , is clearly and distinctly conceived, then *p* is evident (see DESCARTES; HUME); or simply, (b) if we can’t conceive *p* to be false, then *p* is evident; or (c) whatever we are immediately conscious of in thought or experience, e.g. that we seem to

see red, is evident. These might be criteria whereby putative self-evident truths, e.g. that one clearly and distinctly conceives  $p$ , “transmit” the status as evident they already have for one without criteria to other propositions like  $p$ . Alternatively, they might be criteria whereby epistemic status, e.g.  $p$ ’s being evident, is “originally created” by purely non-epistemic considerations, e.g. facts about how  $p$  is conceived which are neither self-evident nor already criterially evident.

However it is “originally created”, presumably epistemic status, including degrees of warranted acceptance or probability, can be “transmitted” deductively from premisses to conclusions. Criteria then must say when and to what degree, e.g.,  $p$  and  $q$  is warranted, given the epistemic considerations that  $p$  is warranted and so is  $q$ . (Must the logical connection itself be evident?) It is usually thought warrant can also be “transmitted” inductively, as when evidence that observed type A things have regularly been F warrants acceptance, in the absence of undermining (overriding) evidence, of an unobserved A as F. Such warrant is defeasible (see *DEFEASIBILITY*). Thus, despite regular observations of black crows, thinking an unobserved crow black might not be very warranted if there recently have been radiation changes potentially affecting bird colour.

Traditional criteria do not seem to make evident propositions about anything beyond our own thoughts, experiences and necessary truths, to which deductive or inductive criteria may be applied. Moreover, arguably, inductive criteria, including criteria warranting the best explanations of data, never make things evident or warrant their acceptance enough to count as knowledge.

Contemporary philosophers, however, have defended criteria whereby, e.g., considerations concerning a person’s facial expression, may (defeasibly) make her pain or anguish evident (Lycan, 1971). More often, they have argued for criteria whereby some propositions about perceived reality can be made evident by sense experience itself or by evident propositions about it. For instance, in the absence of relevant evidence that perception is currently unreliable, it is evident we

actually see a pink square *if* we have the sense experience of seeming to see a pink square (Pollock, 1986), or *if* it is *evident* we have such experience, or *if* in sense experience we spontaneously think we see a pink square. The experiential consideration allegedly can be enough to make reality evident, albeit defeasibly. It can do this on its own, and doesn’t need support from further considerations such as the absence of undermining evidence or inductive evidence for a general link between experience and reality. Of course, there can be undermining evidence. So we need criteria that determine when evidence undermines and ceases to undermine.

Warrant might also be *increased* rather than just ‘passed on’. The coherence (see *COHERENTISM*) of probable propositions with other probable propositions might (defeasibly) make them all more evident (Firth, 1964). Thus even if seeming to see a chair initially made a chair’s presence only probable, its presence might eventually become evident by cohering with claims about chair perception in other cases (Chisholm, 1989). The latter may be warranted in turn by “memory” and “introspection” criteria, as often suggested, whereby recalling or introspecting  $p$  defeasibly warrants  $p$ ’s acceptance. Some philosophers argue further that coherence doesn’t just increase warrant, and defend an overall coherence criterion: excluding perhaps initial warrant for propositions concerning our beliefs and their logical interrelations, what warrants any proposition to any degree for us is its coherence with the most coherent system of belief available (BonJour, 1985).

Contemporary epistemologists thus suggest the traditional picture of criteria may need alteration in three ways. Additional evidence may subject even our most basic judgements to rational correction, though they count as evident on the basis of our criteria. Warrant may be transmitted other than through deductive and inductive relations between propositions. Transmission criteria might not simply “pass” evidence on linearly from a foundation of highly evident “premisses” to “conclusions” that are never more evident.

Criteria then standardly take the form: “if C, then (in the absence of undermining

evidence) *p* is evident or warranted to degree *d*". Arguably, criteria don't play much role initially in forming our beliefs. (*But see* Pollock, 1986.) For them to be the standards of epistemic status for us, however, it's typically thought criterial considerations must be ones in the light of which we can at least check, and perhaps correct, our judgement (*see* EXTERNALISM/INTERNALISM). Traditionally, epistemologists have therefore thought criterial considerations must be at least discoverable through reflection or introspection and thus ultimately concern internal factors about our conception, thoughts or experience. However, others think objective checks must be publicly recognizable checks (*see* PRIVATE LANGUAGE ARGUMENT; PROBLEM OF RULE-FOLLOWING) and thus that criterial considerations must ultimately concern public factors, e.g. that standard conditions (daylight, eyes open, etc.) for reliable perceptual reports obtain.

What makes criteria correct? For many epistemologists, their correctness is an irreducible necessary truth, a matter of brute metaphysics or of our linguistic conventions, concerning epistemic status and the considerations that determine it. Others object that it remains mysterious why particular considerations are criterial unless notions of the evident or warranted or correct are further defined in non-epistemic terms. Criteria might be definable, for example, as principles reflecting our deepest self-critical thoughts about what considerations yield truth, or as norms of thought that practical rationality demands we adopt if we are to be effective agents. However, many will *further* object that satisfying criteria *must* yield truth or be prone to. They insist that necessarily (1) whatever is warranted has an objectively good chance of truth and (2) whatever is evident is true or – to allow for the defeasibly evident – almost invariably true. Epistemic notions allegedly lose their point unless they somehow measure a proposition's actual prospects for truth for us.

Against (1) and (2), a common objection is that no considerations relevantly guarantee truth, even for the most part, or in the long run (*but see* Bonjour, 1985). This is not obvious

with traditional putative criterial considerations like clear and distinct conception or immediate awareness. But, critics argue, when talk of such considerations is unambiguously construed as talk of mental activity, and is not just synonymous with talk of clearly and distinctly or immediately knowing, there's no necessary connection between being criterially evident on the basis of such considerations and being true (Sellars, 1979). The mere coincidence in some cases that the proposition we conceive is true can't be what makes the proposition evident.

None the less, (1) and (2) might be necessary, while the correctness of putative criteria is a contingent fact: given various facts about us and our world, it is no coincidence that adhering to these criteria leads to truth, almost invariably or frequently. Given our need to survive with limited intellectual resources and time, perhaps it isn't surprising that in judging issues we only demand criterial considerations that are fallible, checkable, correctible and contingently lead to truth. However, specifying the relevant truth connections is highly problematic (*see* RELIABILISM). Moreover, reliability considerations now seem to be criterial for criteria although reliability, e.g. concerning perception, is not always accessible to introspection and reflection. Perhaps traditional accessibility requirements may be rejected. Possibly, instead, what makes a putative criterion correct can differ from the criterial considerations that make its correctness evident. Thus, there might be criteria for (defeasibly) identifying criteria, e.g. whether propositions "feel right", or are considered warranted, in "thought experiments" where we imagine various putative considerations present and absent. Later reflection and inquiry might reveal what makes them all correct, e.g. reliability, or being designed by God or nature for our reliable use, etc.

In any case, if criterial considerations do not guarantee truth, knowledge will require more than truth and satisfying even the most demanding criteria (*see* GETTIER PROBLEM). Whether we know we see a pink cube on a particular occasion may also require that there fortunately be no *discoverable* facts (e.g.

of our presence in a hologram gallery) to undermine the experiential basis for our judgement – or perhaps instead that it's no accident our judgement is true rather than merely probably true, given the criteria we adhere to and the circumstances (e.g. our presence in a normal room). Claims that truths which satisfy the relevant criteria are known can clearly be given many interpretations.

Many contemporary philosophers address these issues about criteria with untraditional approaches to meaning and truth. Pollock (1974), for example, argues that learning ordinary concepts like "bird" or "red" involves learning to make judgements with them in conditions, e.g. perceptual experiences, which warrant them, albeit defeasibly since we also learn to correct the judgements despite the presence of such conditions. These conditions are not logically necessary or sufficient for the truth of judgements. None the less the identity of our ordinary concepts makes the criteria we learn for making judgements necessarily correct. Although not all warranted assertions are true, there is no idea of their truth completely divorced from what undefeated criterial considerations allow us to assert. However, satisfying criteria still seems in some way compatible with future defeat, even frequent, and with not knowing, just as it was with error and, defeat in more traditional accounts.

By appealing to defeasibly warranting criteria then, it seems we can not *show* we know *p* rather than merely satisfy the criteria. Worse, critics argue, we can not even have knowledge by satisfying such criteria. Knowing *p* allegedly requires more. But what evidence, besides that entitling us to claim the currently undefeated satisfaction of criteria, could entitle us to claim more, e.g. that *p* won't be defeated? Yet knowers, at least on reflection, must be entitled to give assurances concerning these further conditions (Wright, 1984). Otherwise we wouldn't be interested in a concept of knowledge as opposed to the evident or warranted. These contentions might be disputed to save a role for defeasibly warranting criteria. But why bother? Why can't a pink cube manifest itself in visual experiences that are essentially different from

those where it merely appears present (McDowell, 1983)? We thereby know objective facts through experiences that are criterial for them and make them indefeasibly evident. However, to many, this requires a seamless, quite mystifying, fusion of appearance and reality. Alternatively, perhaps knowledge requires exercising an ability to judge accurately in specific relevant circumstances, but doesn't require criterial considerations which, as a matter of general principle, make propositions evident, even if only in the absence of undermining evidence or contingently, no matter what the context. Arguably, however, our position for giving relevant assurances doesn't improve with these new conditions for knowing.

It is difficult to formulate general principles determining when criterial warrant is and isn't undermined (Pollock, 1974). So one might think that warrant in general depends just on what is presupposed as true and relevant in a potentially shifting context of thought or conversation, not on general criteria. However, defenders of criteria may protest that coherence, *at least*, remains as a criterion applicable across contexts.

It is often felt that *p* can not be evident by satisfying criteria unless it is evident that (a) criterial considerations obtain, and also evident either that (b) the criteria have certain correctness-making features e.g. leading to truth, or just that (c) the criteria are correct. Otherwise any conformity to pertinent standards is in a relevant sense only accidental (BonJour, 1985). Yet vicious regress or circularity loom, unless (a)–(c) or supporting propositions are evident without criteria. At worst, as sceptics argue, nothing can be warranted; at best a consistent role for criteria is limited. A common reply is that being criterially warranted, by definition, just requires that adequate (checkable) criterial considerations in fact obtain, i.e. that (a)–(c) be true. There is no need to demand further cognitive achievements for which one or more of (a)–(c) must also be evident, e.g. actually checking that criterial considerations obtain, proving truth or likelihood of truth on the basis of these considerations, or proving warrant on their basis.



Even so, how *can* propositions stating which putative criteria are correct be warranted? Any proposal for criterial warrant invokes the classic sceptical charge of vicious regress or circularity (*see* PROBLEM OF THE CRITERION). Yet again, arguably, as with *p* above, correct criteria must in fact be satisfied, but this fact itself needn't be *already* warranted. So, one might argue there's no debilitating regress or circle of warrant, even when, as may happen with some criterion, its correctness is warranted ultimately only because it itself is satisfied (Van Cleve, 1979). Independent, ultimately non-criterial, evidence is not needed. None the less, suppose we argue that our criteria are correct, because, e.g. they lead to truth, are confirmed by thought experiments, or are clearly and distinctly conceived as correct, etc. However we develop our arguments, they won't persuade those who, doubting the criteria we conform to, doubt our premisses or their relevance. Dismissing our failure as merely conversational and irrelevant to our warrant, moreover, may strike sceptics and non-sceptics alike as question-begging or as arbitrarily altering what warrant requires. For the charge of ungrounded dogmatism to be inappropriate, more than the consistency of criterial warrant, including warrant about warrant, may be required, no matter what putative criteria we conform to.

*See also* CHISHOLM; COMMONSENSISM AND CRITICAL COGNITIVISM.

## BIBLIOGRAPHY

- BonJour, L.: *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1985).  
 Chisholm, R. M.: *Theory of Knowledge*, 3rd edn (Englewood Cliffs, NJ: Prentice-Hall, 1989).  
 Firth, R.: "Coherence, Certainty, and Epistemic Priority," *Journal of Philosophy* 61 (1964), 545–57.  
 Lycan, W. G.: "Non-inductive Evidence: Recent work on Wittgenstein's 'Criteria,'" *American Philosophical Quarterly* 8 (1971), 109–25.  
 McDowell, J.: "Criteria, Defeasibility, and Knowledge," *Proceedings of the British Academy* 68 (1983), 455–79.

- Pollock, J.: *Knowledge and Justification* (Princeton, NJ: Princeton University Press, 1974).  
 Pollock, J.: *Contemporary Theories of Knowledge* (Totowa, NJ: Rowman and Littlefield, 1986).  
 Sellars, W.: "More on Givenness and Explanatory Coherence," *Justification and Knowledge*, ed. G. Pappas (Dordrecht: Reidel, 1979).  
 Van Cleve, J.: "Foundationalism, Epistemic Principles, and the Cartesian Circle," *Philosophical Review* 88 (1979), 55–91.  
 Wright, C. J. G.: "Second Thoughts about Criteria," *Synthese* 58 (1984), 383–405.

BRUCE HUNTER

**criterion, canon** Except for alleged cases of self-evident truths, it is often thought that anything that is known must satisfy certain criteria or standards (Sextus Empiricus, 1933). These criteria are general principles specifying the sorts of considerations that will make a proposition evident or just make accepting it warranted to some degree. Common suggestions for this role include: if one clearly and distinctly conceives a proposition *p*, e.g. that  $2 + 2 = 4$ , *p* is evident: or, if *p* coheres with the bulk of one's beliefs, *p* is warranted. These might be criteria whereby putative self-evident truths, e.g. that one clearly and distinctly conceives *p*, "transmit" the status as evident they already have without criteria to other propositions like *p*, or they might be criteria whereby purely non-epistemic considerations, e.g. facts about logical connections or about conception that need not be already evident or warranted, originally "create" *p*'s epistemic status. If that in turn can be "transmitted" to other propositions, e.g. by deduction or induction, there will be criteria specifying when it is.

*See also* CRITERIA AND KNOWLEDGE; SEXTUS EMPIRICUS.

## BIBLIOGRAPHY

- Chisholm, R. M.: *Theory of Knowledge*, 3rd edn (Englewood Cliffs, NJ: Prentice-Hall, 1989), 6–7, 61–74.  
 Lycan, W. G.: "Non-inductive Evidence: Recent Work on Wittgenstein's 'Criteria,'" *American Philosophical Quarterly* 8 (1971), 109–25.

Pollock, J.: *Knowledge and Justification* (Princeton, NJ: Princeton University Press, 1974), 3–22, 33–49, 81–5, 255–64.

Sextus Empiricus: *Outlines of Empiricism* (Cambridge, MA: Harvard University Press, 1933; London: William Heinemann, 1933), 17, 125–203, 213–75.

Van Cleve, J.: “Foundationalism, Epistemic Principles, and the Cartesian Circle”, *Philosophical Review* 88 (1979), 55–91.

BRUCE HUNTER

**critical cognitivism** *see* COMMONSENSISM AND CRITICAL COGNITIVISM.

# D

---

**Davidson, Donald (1930–2003)** American philosopher. Davidson argues that if many of our beliefs cohere with many others, then many of our beliefs are true. This has, as Davidson emphasizes, striking consequences for traditional epistemological issues: if it is right, then the refutation of scepticism (see SCEPTICISM) requires only the weak premiss of belief coherence (see COHERENTISM).

To understand his argument for this conclusion, we must consider the notion of radical interpretation. A “radical interpreter” is defined to be someone who faces the problem of content attribution given only knowledge of the correlations between his informant’s local circumstances and the occasion sentences that the informant holds true (together with general principles of warranted deductive and non-demonstrative inference). Now Davidson takes it to be a necessary truth that any content-bearing state can be interpreted under these epistemological conditions, and argues from this that there is no way the radical interpreter can discover the speaker to be largely wrong about the world. The argument is that the interpreter has no alternative but to interpret sentences held true according to the events and objects in the outside world that cause the sentence to be held true. In effect, Davidson holds that the radical interpreter’s strategy must be first to find out what causes the informant to say what he does, and then to identify the truth conditions of the informant’s utterances (more or less comprehensively) with their causes. But, Davidson concludes, if the radical interpreter does proceed this way, he cannot but accept that, in general, the informant’s utterances about the world are true (by the interpreter’s lights).

Suppose that, for the sake of the argument, we assume that the radical interpreter’s

epistemological position does have a privileged metaphysical standing. It looks like all that follows is that the sentences that the informant holds true must be true by the interpreter’s lights. It is possible that what the informant says should be true by the interpreter’s lights may still generally be false. It looks like the possibility of a *folie à deux* is left open by the principle of charity (which is, of course, just what the sceptic always thought).

Davidson is quite aware of this problem; he offers the following argument in reply. It cannot be the rule that what the informant says, though true by the interpreter’s lights, is nevertheless false. 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 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 by the objective standards. So the sceptic is finally refuted.

Many philosophers find that the following line of argument provides a *reductio* of the notion that radical interpretation should proceed by the exercise of charity: if an omniscient being interprets my utterances (or beliefs) so that they come out true by its lights, then it will misinterpret me whenever I say (or believe) something false.

## WRITINGS

“On the Very Idea of a Conceptual Scheme,” in his *Inquiries into Truth and Interpretation* (Oxford: Oxford University Press. 1984).

"Empirical Content," in *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, ed. E. LePore (Oxford: Blackwell, 1986).

"A Coherence Theory of Truth and Knowledge," in *ibid.*

ERNEST LEPORE

**death of epistemology** Rumours about the death of epistemology began to circulate widely in the 1970s. Death notices appeared in such works as *Philosophy and the Mirror of Nature* (1979) by Richard Rorty (see RORTY) and Williams's *Groundless Belief* (1977). Of late, the rumours seem to have died down, but whether they will prove to have been exaggerated remains to be seen.

Arguments for the death of epistemology typically pass through three stages. At the first stage, the critic characterizes the task of epistemology by identifying the distinctive sorts of questions it deals with. At the second stage, he tries to isolate the theoretical ideas that make those questions possible. Finally, he tries to undermine those ideas. His conclusion is that, since the ideas in question are less than compelling, there is no pressing need to solve the problems they give rise to. Thus the death-of-epistemology theorist holds that there is no barrier in principle to epistemology's going the way of, say, demonology or judicial astrology. These disciplines too centred on questions that were once taken very seriously indeed, but as their presuppositions came to seem dubious, debating their problems came to seem pointless. Furthermore, some theorists hold that philosophy, as a distinctive, professionalized activity, revolves essentially around epistemological inquiry, so that speculation about the death of epistemology is apt to evolve into speculation about the death of philosophy generally.

Clearly, the death-of-epistemology theorist must hold that there is nothing special about philosophical problems. This is where philosophers who see little sense in talk of the death of epistemology disagree. For them, philosophical problems, including epistemological problems, are distinctive in that they are "natural" or "intuitive": that is to say,

they can be posed and understood taking for granted little or nothing in the way of contentious, theoretical ideas. Thus, unlike problems belonging to the particular sciences, they are "perennial" problems that could occur to more or less anyone, anytime and anywhere. But are the standard problems of epistemology really as "intuitive" as all that? Or if they have indeed come to seem so commonsensical, is this only because commonsense is a repository for ancient theory? These are the sorts of question that underlie speculation about epistemology's possible demise.

Because it revolves round questions like this, the death-of-epistemology movement is distinguished by its interest in what we may call "theoretical diagnosis": bringing to light the theoretical background to philosophical problems so as to argue that they cannot survive detachment from it. This explains the movement's interest in historical-explanatory accounts of the emergence of philosophical problems. If certain problems can be shown not to be perennial, but rather to have emerged at definite points in time, this is strongly suggestive of their dependence on some particular theoretical outlook; and if an account of that outlook makes intelligible the subsequent development of the discipline centred on those problems, that is evidence for its correctness. Still, the goal of theoretical diagnosis is to establish logical dependence, not just historical correlation. So although historical investigation into the roots and development of epistemology can provide valuable clues to the ideas that inform its problems, history cannot substitute for problem-analysis.

The death-of-epistemology movement has many sources: in the pragmatists, particularly James and Dewey, and in the writings of Wittgenstein, Quine, Sellars and Austin (see JAMES; DEWEY; WITTGENSTEIN; QUINE; SELLARS; AUSTIN). But the project of theoretical diagnosis must be distinguished from the "therapeutic" approach to philosophical problems that some names on this list might call to mind. The practitioner of theoretical diagnosis does not claim that the problems he analyses are "pseudo-problems", rooted in "conceptual confusion". Rather, he claims

that, while genuine, they are wholly internal to a particular intellectual project whose generally unacknowledged theoretical commitments he aims to isolate and criticize.

Turning to details, the task of epistemology, as these radical critics conceive it, is to determine the nature, scope and limits, indeed the very possibility of human knowledge. Since epistemology determines the extent to which knowledge is possible, it cannot itself take for granted the results of any particular forms of empirical inquiry. Thus epistemology purports to be a non-empirical discipline, the function of which is to sit in judgement on all particular discursive practices with a view to determining their cognitive status. The epistemologist (or, in the era of epistemologically centred philosophy, we might as well say "the philosopher") is someone professionally equipped to determine what forms of judgement are "scientific", "rational", "merely expressive", and so on. Epistemology is therefore fundamentally concerned with sceptical questions. Determining the scope and limits of human knowledge is a matter of showing where and when knowledge is possible. But there is a project called "showing that knowledge is possible" only because there are powerful arguments for the view that knowledge is impossible. Here the scepticism (*see* SCEPTICISM) in question is first and foremost radical scepticism, the thesis that with respect to this or that area of putative knowledge we are never so much as justified in believing one thing rather than another. The task of epistemology is thus to determine the extent to which it is possible to respond to the challenges posed by radically sceptical arguments by determining where we can and cannot have justifications for our beliefs. If it turns out that the prospects are more hopeful for some sorts of beliefs than for others, we shall have uncovered a difference in epistemological status. The "scope and limits" question and problem of radical scepticism are two sides of one coin.

This emphasis on scepticism as the fundamental problem of epistemology may strike some philosophers as misguided. Much recent work on the concept of knowledge, particularly that inspired by Gettier's demon-

stration of the insufficiency of the standard "justified true belief" analysis, has been carried on independently of any immediate concern with scepticism. I think it must be admitted that philosophers who envisage the death of epistemology tend to assume a somewhat dismissive attitude to work of this kind. In part, this is because they tend to be dubious about the possibility of stating precise necessary and sufficient conditions for the application of any concept. But the determining factor is their thought that only the centrality of the problem of radical scepticism can explain the importance for philosophy that, at least in the modern period, epistemology has taken on. Since radical scepticism concerns the very possibility of justification, for philosophers who put this problem first, questions about what special sorts of justification yield knowledge, or about whether knowledge might be explained in non-justificational terms, are of secondary importance. Whatever importance they have will have to derive in the end from connections, if any, with sceptical problems.

In light of this, the fundamental question for death-of-epistemology theorists becomes, "What are the essential theoretical presuppositions of arguments for radical scepticism?" Different theorists suggest different answers. Rorty (*see* RORTY) traces scepticism to the "representationalist" conception of belief and its close ally, the correspondence theory of truth. According to Rorty, if we think of beliefs as "representations" that aim to correspond with mind-independent "reality" (mind as the mirror of nature), we will always face insuperable problems when we try to assure ourselves that the proper alignment has been achieved. In Rorty's view, by switching to a more "pragmatic" or "behaviouristic" conception of beliefs as devices for coping with particular, concrete problems, we can put scepticism, hence the philosophical discipline that revolves around it, behind us once and for all.

Other theorists stress epistemological foundationalism (*see* FOUNDATIONALISM) as the essential background to traditional sceptical problems. There are reasons for preferring this approach. Arguments for epistemological



conclusions require at least one epistemological premiss. It is, therefore, not easy to see how metaphysical or semantic doctrines of the sort emphasized by Rorty could, by themselves, generate epistemological problems, such as radical scepticism. On the other hand the case for scepticism's essential dependence on foundationalist preconceptions is by no means easy to make. It has even been argued that this approach "gets things almost entirely upside down". The thought here is that foundationalism is an attempt to save knowledge from the sceptic, and is therefore a reaction to, rather than a presupposition of, the deepest and most intuitive arguments for scepticism. Challenges like this certainly need to be met by death-of-epistemology theorists, who have sometimes been too ready to take for obvious scepticism's dependence on foundationalist or other theoretical ideas. This reflects, perhaps, the dangers of taking one's cue from historical accounts of the development of sceptical problems. It may be that, in the heyday of foundationalism, sceptical arguments were typically presented within a foundationalist context. But the crucial question is not whether some sceptical arguments do take foundationalism for granted but whether there are any that do not. This issue – indeed, the general issue of whether scepticism is a truly intuitive problem – can only be resolved by detailed analysis of the possibilities and resources of sceptical argumentation.

Another question concerns why anti-foundationalism leads to the death of epistemology rather than a non-foundational, hence coherentist (see COHERENTISM), approach to knowledge and justification. It is true that death-of-epistemology theorists often characterize justification in terms of coherence. But their intention is to make a negative point. According to foundationalism, our beliefs fall naturally into broad epistemological categories that reflect objective, context-independent relations of epistemological priority. Thus, for example, experiential beliefs are thought to be naturally or intrinsically prior to beliefs about the external world, in the sense that any evidence we have for the latter must derive in the end from the former. This relation of epistemic priority is, so to say,

just a fact, Foundationalism is therefore committed to a strong form of realism (see REALISM) about epistemological facts and relations, call it "epistemological realism". For some anti-foundationalists, talk of coherence is just a way of rejecting this picture in favour of the view that justification is a matter of accommodating new beliefs to relevant background beliefs in contextually appropriate ways, there being no context-independent, purely epistemological restrictions on what sorts of beliefs can confer evidence on what others. If this is all that is meant, talk of coherence does not point to a theory of justification so much as to the deflationary view that justification is not the sort of thing we should expect to have theories about. There is, however, a stronger sense of "coherence" which does point in the direction of a genuine theory. This is the radically holistic account of justification, according to which inference depends on assessing our entire belief-system or "total view" in the light of abstract criteria of "coherence". But it is questionable whether this view, which seems to demand privileged knowledge of what we believe, is an alternative to foundationalism or just a variant form. Accordingly, it is possible that a truly uncompromising anti-foundationalism will prove as hostile to traditional coherence theories as to standard foundationalist positions, reinforcing the connection between the rejection of foundationalism and the death of epistemology.

The death-of-epistemology movement has some affinities with the call for a "naturalized" approach to knowledge. Quine argues that the time has come for us to abandon such traditional projects as refuting the sceptic by showing how empirical knowledge can be rationally reconstructed on a sensory basis, hence *justifying* empirical knowledge at large. We should concentrate instead on the more tractable problem of *explaining* how we "project our physics from our data", i.e. how retinal stimulations cause us to respond with increasingly complex sentences about events in our environment. Epistemology should be transformed into a branch of natural science, specifically experimental psychology. But though Quine presents this as a suggestion

about how to continue doing epistemology, to philosophers who think that the traditional questions still lack satisfactory answers, it looks more like abandoning epistemology in favour of another pursuit entirely. It is significant, therefore, that in subsequent writings Quine has been less dismissive of sceptical concerns. But if this is how "naturalized" epistemology develops, then for the death-of-epistemology theorist, its claims will open up a new field for theoretical diagnosis.

See also NATURALIZED EPISTEMOLOGY; REPRESENTATION; REPRESENTATIVE REALISM; TRUTH.

# BIBLIOGRAPHY

- Kaplan, M.: "Epistemology on Holiday," *Journal of Philosophy* (1991).
- Quine, W. V.: "Epistemology Naturalized," in his *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969).
- Quine, W. V.: "The Nature of Natural Knowledge," in S. Guttenplan, *Mind and Nature* (Oxford: Oxford University Press, 1975).
- Rorty, R.: *Philosophy and the Mirror of Nature* (Princeton, NJ: Princeton University Press, 1979).
- Rorty, R.: *Consequences of Pragmatism* (Minneapolis, MN: University of Minnesota Press, 1982).
- Rorty, R.: *Objectivity, Relativism and Truth*, (Cambridge: Cambridge University Press, 1991).
- Stroud, B.: "Skepticism and the Possibility of Knowledge," *Journal of Philosophy* (1984).
- Williams, M.: *Groundless Belief* (Oxford: Blackwell, 1977).
- Williams, M.: *Unnatural Doubts* (Oxford: Blackwell, 1992).

MICHAEL WILLIAMS

**defeasibility** The warrant a proposition *p* has for us on the basis of evidence *e* is *defeasible* when expanded or improved evidence could decrease *p*'s warrant. For example, "The next crow we see will be black" is less warranted when, despite evidence *e* that observed crows were black, we are told on usually reliable authority that there are many albino birds nearby, and have no evidence that this present testimony is an

exception. The warrant the proposition or belief has on evidence *e* is thus defeasible. Our actual warrant for a proposition or belief depends on our total evidence. In stock cases of defeated warrant, we don't "lose" *e*, but its import is undercut when it and the rest of our original total evidence is combined with *additional* evidence *e'*. With *holistic* warrant the situation seems different. For instance, the warrant *p* has through membership in the most coherent system of belief available is *defeasible* when there could be a more coherent system available which omits *p*. Much else may disappear from our original system or evidence.

Various philosophically interesting senses of "defeasible" correspond to different senses in which there "could" be defeating evidence, e.g. because it is logically possible that there is evidence that, combined with our present evidence, makes *p* less warranted, or because there actually are past or future events whose discovery, combined with our present evidence, would make *p* less warranted, etc. In any case, defeasible warrant can be contrasted with warrant that makes a proposition or belief *certain* for us. Thus, the nature of defeasible warrant and justification, as well as that of defeasible reasons (as opposed to conclusive reasons) for belief, is explored by recent epistemologists interested in developing accounts of knowledge according to which knowledge needn't require certainty but only a lesser grade of warrant or justification.

John Pollock and Roderick Chisholm, for example, have argued that not only can there be inductive evidence or inductive reasons that defeasibly warrant a belief; there can be non-inductive defeasible reasons for belief. Pollock argues, for example, that looking red is a non-inductive defeasible reason for thinking something is red. One issue, therefore, for defenders of defeasible warrant is to explain how evidence *e* can warrant a proposition or belief *p* for us if *e* doesn't entail the truth of *p* and there are no inductive reasons for thinking that *e* makes *p* likely to be true. Indeed, even in the case of putative inductive evidence *e* for *p*, defenders of defeasible warrant need to explain how evidence *e* can warrant *p* if it doesn't entail *p*. Is it a

necessary truth, perhaps a conceptual one, that *e* defeasibly warrants *p* for us, or is it a contingent truth – and, in either case, how and why?

Pollock has also influentially distinguished two types of defeaters. When *R* is a defeasible reason for believing *p*, *D* is a *rebutting* defeater for *R* as a reason for believing *p* when *D* is a reason for believing not-*P*, and *D* is an *undercutting* defeater when it is a reason to doubt or deny that *R* wouldn't be true unless *p* were true. For example, "crows have changed colour due to radiation" is a rebutting defeater for my past observations of black crows as a reason for believing the next crow I see will be black, but "there have been radiation effects on earth affecting the colour of some birds" is an undercutting defeater. Pollock also recognizes that, as well as defeaters, there can be defeater defeaters that restore the warrant our original reasons provide, and defeaters for defeater defeaters, and so on. One problem for defenders of defeasible warrant is to explain why *D* defeats *R* as a reason for believing *p*, rather than *R* defeating *D* as a reason for believing not-*p*, or as a reason for doubting whether *R* wouldn't be true unless *p* were true. This has also become a problem for researchers in logic, computing science, and artificial intelligence, who have recognized that much human reasoning we regard as good reasoning is defeasible, and so have made the study of defeasible reasoning (often under the label "non-monotonic logic"), as well as the study of belief revision and updating, central to their work.

The notion of defeasible warrant also plays a key role in so-called "defeasibility" solutions to the Gettier problem for the analysis of knowledge. According to such accounts, very roughly, knowledge requires true belief which is warranted, at least defeasibly, and remains warranted when our evidence is expanded to include all (relevant) truths and only truths, i.e. when any falsehoods are deleted from our evidence base and all (relevant) truths are added.

See also CERTAINTY; CRITERIA AND KNOWLEDGE; EVIDENCE; GETTIER PROBLEM; PRIMA FACIE REASONS.

## BIBLIOGRAPHY

- Chisholm, R. M.: *Theory of Knowledge*, 3rd edn (Englewood Cliffs, NJ: Prentice-Hall, 1989), 49–60.
- Firth, R.: "Anatomy of Certainty," *Philosophical Review* 76 (1967), 3–27.
- Harman, G.: *Change in View* (Cambridge, MA: MIT Press, 1986).
- Pollock, J., and Cruz, J.: *Contemporary Theories of Knowledge*, 2nd edn (Totowa, NJ: Rowman and Littlefield, 1999), chs 2, 5, and 7.
- Shope, R.: *The Analysis of Knowing: A Decade of Research* (Princeton, NJ: Princeton University Press, 1983), chs 2–4.

BRUCE HUNTER

**definition** Complex expressions either *reporting* or *instituting* equivalences among verbal or symbolic expressions, in form, definitions are either *explicit* or *implicit*. A definition that *institutes* explains how an expression *will be used* henceforth. A definition that *reports* explains how an expression *has been used*. An *explicit* definition explains, by means of words given in *use*, how an expression given *in mention* has been or will be used (see USE/MENTION). An *implicit* definition explains how an expression has been or will be used by using it, usually in conjunction with the use of other expressions.

Dictionary definitions are reportive and explicit. Symbols introduced in technical writings are usually institutive and explicit. When a word is learned in the context of its use, that context *in effect* provides a reportive, implicit definition. Formal, axiomatic systems, in which the meaning of each expression is gathered from its formal-logical relationships with the other expressions, provide institutive, implicit definitions.

## BIBLIOGRAPHY

- Frege, G.: *The Basic Laws of Arithmetic*, trans. and ed. M. Furth (Berkeley and Los Angeles, CA: University of California Press, 1967).
- Quine, W. V.: "Implicit Definition Sustained," in his *The Ways of Paradox* (Cambridge, MA: Harvard University Press, 1976), ch. 13.

Quine, W. V.: “Vagaries of Definition,” in his *The Ways of Paradox* (Cambridge, MA: Harvard University Press, 1976). ch. 7.

ROBERT S. TRAGESSER

**Derrida, Jacques (1930–2004)** French philosopher, born in Algiers. Derrida argues against foundational and essentialist metaphysics and epistemology. As a paradigm example, *Speech and Phenomena* examines Husserl’s attempt to ground knowledge on contents which can be present before a unitary self (*see* HUSSERL):

1. For Husserl, there is no consciousness of present contents as present except by reference to the just-past and the just-to-come. So “intentional contents” cannot be entirely present before the mind. The ultimate data, by Husserl’s own account, include always-non-present items, and are experienced that way.
2. Intentional contents, as repeatables, are what they are by virtue of other possible cases. But then an intentional object cannot be present to the self as such. Derrida thus supplements Sellars’s attack (*see* SELLARS) on sense-data (*see* SENSE-DATA) with an argument against the givenness of intentional contents generally. Meanings are essentially repeatables; what is fully present can only be particulars. So meanings cannot in principle be given as such. “Intelligible” data have the same difficulties in functioning cognitively as sense-data have.
3. The necessary “traces” of what cannot be given make thought and experience language-like, because their meaning is partly non-present. That is, no pure meaning can be present to consciousness as a foundation which grounds interpretation. Thus thought is language-like and itself subject to interpretation.

Derrida draws many Quinean consequences from the idea that thought is no more transparent than writing. Moreover, Derrida, emerging from a thorough grounding in Heidegger, Levinas and Hegel, is anti-essentialist

about the theory-builder, the self. The attack above on presence is also part of an attack on the unitary self which can be self-present in its given nature. If the subject is not given, but is rather on a par with other “posits”, then a basis for an epistemologically significant division between “inner” and “outer” disappears. Thus also the notion of the subjective as a realm of representations which must somehow be matched with the objective, evaporates. Derrida thus shows that a genuine anti-essentialism is also post-Cartesian.

*See also* NECESSARY/CONTINGENT; QUINE; SELF-KNOWLEDGE AND SELF-IDENTITY.

#### WRITINGS

*La Voix et le Phénomène* (Paris: Presses Universitaires de France, 1967); trans. D. Allsion as *Speech and Phenomena* (Evanston, IL: Northwestern University Press, 1973).

*De la Grammatologie* (Paris: Editions de Minuit, 1967); trans. G. Spivak as *Of Grammatology* (Baltimore, MD: Johns Hopkins University Press, 1976).

*Marges de la Philosophie* (Paris: Editions de Minuit, 1972); trans. A. Bass, *Margins of Philosophy* (Chicago, IL: University of Chicago Press, 1982).

*Limited Inc.* (Evanston, IL: Northwestern University Press, 1988).

SAMUEL C. WHEELER III

**Descartes, René (1596–1650)** French philosopher, scientist and mathematician. Descartes is often called the “father of modern philosophy”, on the grounds that he made epistemological questions the primary and central questions of the discipline. But this is misleading for several reasons. In the first place, Descartes’ conception of philosophy was very different from our own. The term “philosophy” in the seventeenth century was far more comprehensive than it is today, and embraced the whole of what we nowadays call natural science, including cosmology, and physics, as well as subjects like anatomy, optics and medicine. Descartes’ reputation as

a “philosopher” in his own time was based as much as anything on his contributions in these scientific areas. In the second place, even in those Cartesian writings that are philosophical in the modern academic sense, the epistemological concerns are rather different from the conceptual and linguistic inquiries that characterize present-day “theory of knowledge”. Descartes saw the need to base his scientific system on secure metaphysical foundations; by “metaphysics” he meant inquiries into “God and the soul and in general all the first things to be discovered by philosophizing” (letter to Mersenne of 11 November 1640). These foundational inquiries included, to be sure, questions about knowledge and certainty; but even here, Descartes is not *primarily* concerned with the criteria for knowledge claims, or with definitions of the epistemic concepts involved; his aim, rather, is to provide a unified framework for understanding the universe. In place of the fragmented scholastic world of separate disciplines, each with its own methods and standards of precision, he aimed to construct a coherent theory of the world and man’s place within it. And this project required him “once in the course of his life” systematically to test all his former beliefs, and to subject them to radical scrutiny, in order to see whether he could “establish anything at all in the sciences that was stable and likely to last” (AT VII 17; CSM II 12).

René Descartes was born in France on 31 March 1596, in the small town near Tours which now bears his name, and educated at the Jesuit college of La Flèche in Anjou. As a young man he was strongly influenced by the Dutchman Isaac Beeckman, who awakened his lifelong interest in mathematics – a discipline in which he discerned the precision, order and certainty which merited the title of *scientia* (Descartes’ term for systematic and reliable knowledge based on indubitable foundations). In 1628, Descartes emigrated to Holland where he was to live for most of the rest of his life. In 1633, he had ready a treatise on cosmology and physics, *Le Monde*; but he cautiously withdrew the work from publication when he heard of the condemnation of Galileo by the

Inquisition for rejecting (as Descartes himself did) the geocentric theory of the universe. But in 1637 Descartes decided to release for publication (in French) a sample of his scientific work, the *Optics, Meteorology and Geometry*, together with an autobiographical introduction entitled *Discourse on the Method of rightly conducting one’s reason and reaching the truth in the sciences*. Criticisms of his arguments led Descartes to compose his philosophical masterpiece, the *Meditations on First Philosophy*, published in Latin in 1641. In 1644 Descartes produced, in Latin, a mammoth compendium of his metaphysical and scientific views, the *Principles of Philosophy*, which he hoped would become a university textbook to rival the standard texts based on Aristotle. In 1649 he published *The Passions of the Soul*, a lengthy treatise on ethics and psychology. The same year he accepted an invitation to go to Stockholm to give philosophical instruction to Queen Christina of Sweden. He was required to provide tutorials at the royal palace at five o’clock in the morning, and the strain of this break in his habits (he had maintained the lifelong custom of lying in bed late into the morning) led to his catching pneumonia. He died on 11 February 1650, just short of his fifty-fourth birthday.

Descartes’ views on knowledge were conditioned by the time in which he lived, which had witnessed a gradual erosion of beliefs held for centuries and apparently based on straightforward observation and “common sense”. The most notable example of this was the long-held conviction, bolstered by the authority of the Church, that an immovable earth was the centre of the universe. Galileo’s discovery of the moons of Jupiter (made when Descartes was a nine-year-old schoolboy at La Flèche) was but one piece in a mounting pile of evidence suggesting that the traditional view was radically mistaken. Descartes became obsessed by the thought that no lasting progress could be made in the sciences unless a systematic method could be devised for sifting through our preconceived opinions and establishing which of them, if any, was reliable. “Suppose we had a basket full of apples and were worried that some of them were



rotten. How would be we proceed? Would we not begin by tipping the whole lot out and then pick up and put back only those we saw to be sound?" (AT VII 481; CSM II 324). Descartes' "method of doubt" involves a determined effort to test our preconceived opinions or "prejudices" (*praejudicia*) to the limit, by applying a series of deliberate sceptical techniques (often derived from classical arguments for doubt which had been revived in the sixteenth century). He points out first, that the senses (sight, hearing, touch, etc.) are often unreliable, and "it is prudent never to trust entirely those who have deceived us even once" (First Meditation); later, he cited such instances as the straight stick which looks bent in water, and the square tower which looks round from a distance. This argument from illusion (as it is called today) (see ARGUMENT FROM ILLUSION), has not, on the whole, impressed commentators; and some of Descartes' contemporaries pointed out that since such sensory errors come to light as a result of further sensory information, it cannot be right to cast wholesale doubt on the evidence of the senses. But Descartes himself regarded the argument from illusion as only the first stage in a softening up process which would "lead the mind away from the senses". He admits that there are some cases of sense-based belief about which doubt would be insane – "for example the belief that I am sitting here by the fire, wearing a winter dressing gown" (*ibid.*).

At this point, Descartes introduces a fresh reason for doubt – the celebrated "dreaming argument". "How often, asleep at night, am I convinced of just such familiar events, that I am here in my dressing gown, sitting by the fire, when in fact I am lying undressed in bed." Observing that there are "no conclusive signs" by which being awake can be distinguished from being asleep, Descartes proceeds, in effect, to mount a general doubt about whether we are justified in asserting the real extra-mental existence of any particular object which we appear to perceive via the senses. Critics of this argument have suggested that the very concept of dreaming is parasitic on the concept of waking life, so that, again, we have not been offered a *general*

reason for doubting the existence of external objects. Descartes' defenders, however, can plausibly reply that if in any particular instance the possibility that one is dreaming cannot be ruled out, the solitary doubter has no guarantee of the independent existence of any given object of perception. The conclusion which Descartes eventually draws is that any sciences which make existential assumptions (such as physics, astronomy and medicine) are potentially doubtful, and that only disciplines like arithmetic and geometry "which deal only with the simplest and most general things, regardless of whether they exist in nature or not" enjoy cast iron certainty (AT VII 20; CSM II 14).

Yet even this last certainty is undermined in Descartes' most radical argument for doubt – the deceiving God hypothesis: if, as I have been taught, there is an omnipotent being who created me, then "how do I know that he has not brought it about that I go wrong every time I add two and three or count the sides of a square?" (*ibid.*). There may, of course, be no God; but in that case, Descartes reasons, I owe my existence not to a divine creator but to some chance chain of imperfect causes. And in that case there is even less reason to suppose that my basic mathematical judgements are sound. By the end of the First Meditation, the meditator is "tumbling around" in a vortex of doubt. There is "not one of my former beliefs about which a doubt may not properly be raised"; and Descartes dramatizes this horror of extreme uncertainty by invoking a "supremely powerful and malicious demon" intent on deceiving me in any way he possibly can.

Despite the commonly employed label "Cartesian scepticism" it is important to realize that Descartes is in no sense a sceptic. The systematic doubt is merely a means to an end: the aim is to demolish in order to rebuild – to throw out the rubble and loose sand in order to reach a bedrock of certainty (AT VII 546; CSM II 373). That bedrock is reached in the Second Meditation in the famous *Cogito* argument: "let the demon deceive me as much as he can, he will never bring it about that I am nothing so long as I think I am something. So... I must finally conclude

that this proposition, *I am, I exist* is necessarily true whenever it is put forward by me or conceived in my mind” (AT VII 25; CSM II 17). As Descartes phrased it in the *Discourse*, “I am thinking therefore I exist” (*je pense donc je suis*) is “so firm and sure that the most extravagant suppositions of the sceptics were incapable of shaking it” (AT VI 32; CSM I 127). The most interesting epistemic feature of the Cogito argument is the way in which Descartes extrudes certainty from the very process of doubting: the act of casting doubt on the proposition that one is thinking confirms its truth, and this in turn unavoidably implies that there must be an existing subject. At least one existential truth, *I exist*, survives everything the sceptic can throw at it.

Descartes’ questioning of his previous beliefs is not as radical as is often supposed. In order to reach the certainty of the Cogito, he has to rely (as he later admitted) on an unquestioned underlying conceptual apparatus – for example, his grasp of what is meant by knowledge, or by doubt, and of the principle that “in order to think one must exist” (AT VIII 8; CSM I 196). It may be seen from this that the Cartesian project is not, as is sometimes suggested, “the validation of reason”; apart from the fleeting (and later to be retracted) suggestion in the First Meditation that even the fundamental truths of logic and mathematics might be unstable, there is never any attempt to start with a completely blank slate. If the doubt were as extreme as that, the very process of systematic meditation could never get off the ground in the first place. What Descartes aims to show, rather, is that there is an inescapable logical limit to scepticism about what exists: pushing such doubt to its limits shows that it is self-defeating. And once the existence of at least one item, the thinking self, has been arrived at, Descartes will attempt systematically to reconstruct a reliable body of knowledge. But here we come up against the most striking feature of the Cartesian system from an epistemological point of view: its radically subjective orientation. Descartes has to reconstruct knowledge “from the inside outwards” – from awareness of self to knowledge of the external world. And given the whole-

sale doubts he has raised about the latter, he can only reinstate it by relying on the resources of his own subjective consciousness. One such resource is the idea he finds within him of a supremely perfect being; and (by a complex and notoriously problematic causal argument) he reasons that this can only have been placed in his mind by a really existing perfect creator – God (Third Meditation). Once God’s existence is established by this route, Descartes can proceed to reinstate his former belief in an external world, reasoning that, since God has given him a powerful propensity to believe that many of his ideas have their source in real external objects, such objects must exist – otherwise the deity would be systematically deceiving him, which would be incompatible with divine perfection (Sixth Meditation).

Two important points need to be made about the general Cartesian approach to knowledge. The first is that when Descartes’ reconstruction project has been completed, the resulting edifice is very different from the “commonsense”, pre-philosophical world of the man of the senses. Physical objects exist – that much is guaranteed – but “they may not at all exist in the way that exactly corresponds to my sensory grasp of them – for in many cases the grasp of the senses is obscure and confused” (AT VII 80; CSM II 55). In order to achieve a reliable grasp of the nature of physical reality Descartes urges that we must systematically disregard the confused deliverances of the senses, and rely instead on the “clear and distinct” concepts of pure mathematics which God has implanted in our souls. And hence the resulting structure of Cartesian science sets out (not always convincingly) to reduce all physics to “what the geometers call *quantity*, and take as the object of their demonstrations, i.e. that to which every kind of division shape and motion is applicable” (*Principles of Philosophy*, Part II, article 64). The world of the senses, the qualitative world of smells and tastes and colours and sounds, is thus resolutely excluded from Cartesian science – an exclusion which remains to this day a problem for those who wish, as Descartes did, to achieve a systematic and unified understanding of reality.

The second point to be made about Descartes' system is that his foundational project, even when construed in the relatively modest way suggested above, cannot, it seems, entirely escape what has come to be known as the "bootstrap" problem. God, once his existence is established, functions in Descartes' system as an epistemic guarantor: "the certainty and truth of all knowledge depends uniquely on my awareness of the true God, to such an extent that I was incapable of perfect knowledge about anything until I became aware of him" (Fifth Meditation). But if this is the case, it is not easy to see how Descartes is in a position to establish the reliability of the knowledge needed to establish God's existence in the first place. This problem, which has come to be known as the "Cartesian Circle" was graphically highlighted by Descartes' contemporary Antoine Arnauld: "How do you avoid reasoning in a circle when you say that we are sure that what we clearly and distinctly perceive is true only because God exists, yet we are sure that God exists only because we clearly and distinctly perceive this?" Descartes wrestled with the challenge at length (notably in the Second and Fourth Replies to the *Objections* published with the *Meditations*); his contemporary critics were not satisfied by his answers, and from a present day standpoint it is probably fair to say that the consensus view of Cartesian-style "foundational epistemology" is that it is doomed to failure by its very ambitiousness. But even here the power of Descartes' thinking is manifest; for a great part of the history of the philosophy of our own century has been, in effect, a struggle to escape from Descartes' individualistic and autocentric perspective on the problems of knowledge and certainty – a perspective which, whether we like it or not, has become part of our conceptual heritage.

See also SCEPTICISM, CONTEMPORARY; SELF-KNOWLEDGE AND SELF-IDENTITY.

#### WRITINGS

'AT' refers, by volume and page number, to the standard Franco-Latin edition of Descartes:

*Oeuvres de Descartes* eds C. Adam and P. Tannery. 12 vols (Paris: 1887–1913); revised edn (Paris: Vrin/CNRS, 1964–76).

'CSM' refers by volume and page number to the standard English translation: *The Philosophical Writings of Descartes* eds J. Cottingham, R. Stoothoff and D. Murdoch, vols i and 2 (Cambridge: Cambridge University Press, 1975); for Descartes' philosophical correspondence, see vol. III ('CSMK') by the same translators and A. Kenny (Cambridge: Cambridge University Press, 1991).

#### BIBLIOGRAPHY

- Beyssade, J.-M.: *La Première Philosophie de Descartes* (Paris: Flammarion, 1979).
- Broughton, J.: *Descartes's Method of Doubt* (Princeton, NJ: Princeton University Press, 2002).
- Cottingham, J.: *Descartes* (Oxford: Blackwell, 1986).
- Curley, E. M.: *Descartes against the Sceptics* (Oxford: Blackwell, 1978).
- Frankfurt, H.: *Demons, Dreams and Madmen: The Defense of Reason in Descartes' Meditations* (New York: Bobbs-Merrill, 1970).
- Kenny, A.: *Descartes* (New York: Random House, 1968).
- Newman, L.: "Descartes on Unknown Faculties and Our Knowledge of the External World," *Philosophical Review* 103 (1994), 489–531.
- Newman, L.: "Rocking the Foundations of Cartesian Knowledge: Critical Notice of Janet Broughton, *Descartes's Method of Doubt*," *Philosophical Review* 113 (2004), 101–25.
- Popkin, R.: *The History of Skepticism from Erasmus to Descartes*, 4th edn (Berkeley, CA: University of California Press, 1979).
- Van Cleve, J. 1979: "Foundationalism, Epistemic Principles, and the Cartesian Circle," *Philosophical Review* 88 (1979), 55–91.
- Williams, B.: *Descartes: The Project of Pure Enquiry* (Harmondsworth: Penguin, 1978).
- Wilson, M.: *Descartes* (London: Routledge, 1978).

JOHN COTTINGHAM

**Dewey, John (1859–1952)** American philosopher and educationist. After beginning his long and prolific career as a neo-Hegelian and neo-Kantian. Dewey developed a type of pragmatism ("instrumentalism")

which incorporated in naturalized form many Hegelian and Kantian themes. He taught at Michigan, Chicago and finally at Columbia University.

Dewey so often expressed hostility towards epistemology that many commentators have supposed that his own philosophy lacked what could be properly called an epistemology. To be sure, Dewey did not treat epistemological issues in isolation from other problems. Such isolation, he believed, was partly responsible for the fundamental flaws in previous theories of knowledge. For Dewey, knowing can be understood only in the context of a “theory of inquiry” (i.e. a natural history of thinking as a life process). His focus was primarily on the *process* of knowing (Dewey, 1984, p. 131). In his view, such traditional philosophical tasks as analysis of objects of knowledge and of principles of epistemic justification can be accomplished only within a general theory of the activity of inquiry.

In the context of such a theory, it becomes apparent, Dewey argued, that all previous epistemologies have been examples of “The Spectator Theory of Knowledge” (i.e. the view that the knower is only passively related to the thing known; Dewey, 1984, p. 19). Central to Dewey’s account of the active process of inquiry is the “situation”. Influenced by both neo-Hegelianism and Peirce (see PEIRCE), Dewey considered inquiry a natural, dialectical process by which over time an “indeterminate situation” is made to become a “unified whole” (Dewey, 1986, p. 108).

Commentators have often charged that Dewey, confusing methodology with epistemology, presented an account of coming-to-know as an analysis of knowledge itself (Murphy, 1989, p. 203). However, Georges Dicker (Dicker, 1976, pp. 22–9) has persuasively argued that Dewey intended to show that there is no special act of knowing distinct from and brought about by the natural process of inquiry.

Critics of Dewey’s instrumentalism have further charged that he confused an account of the application of knowledge with a genuinely epistemological account of the possession of knowledge and consequently failed to provide the latter. But, in Dewey’s pragmatism,

to *have* knowledge is precisely to have the ability to anticipate the consequences of manipulating things in the world (Dewey, 1910, pp. 77–111). The metaphysics of Dewey’s theory of inquiry is crucial. He conceived of knowledge in terms of processes, abilities and dispositions, not in the traditional Spectator terms of mental acts and occurrences.

A third objection has frequently been made from Dewey’s time to our own. Commentators complain that Dewey rejected realism (see REALISM) in saying that “the object of knowledge” is the product of inquiry. Undeniably, Dewey sometimes lapsed into the idealistic idiom of his early work. In such lapses he neglected to make distinctions crucial to steering his epistemology safely through the tortuous channel between the Spectator Theory and idealism (see IDEALISM). For example, on occasion he forgot to distinguish between saying that inquiry produces *knowledge* of objects and saying that it produces the *objects* of knowledge (Dewey, 1984, p. 88). The former assertion is consistent with realism, while the latter entails idealism.

This longstanding misreading of Dewey as an idealist has recently been given added currency by philosophers who applaud the constructivism they believe is to be found in Dewey. Some philosophers, most prominently Richard Rorty (Rorty, 1979, pp. 6, 381) (see RORTY), compound the error by claiming Dewey as an ally in their total rejection of epistemology. It is hard to say whether the understanding of Dewey’s epistemology suffers more from attacks by his enemies or more from praise by his friends.

To understand Dewey’s original and powerful form of realism, it is necessary to disentangle what changes in the process of inquiry from what does not change. As H. S. Thayer (1990, pp. 447–8) points out, those readers still in the grip of the Spectator view of “knowing as an internal activity in minds” are baffled by the thesis “that the object and situation in which inquiry is completed consists of existentially changed constituents and relations from that with which inquiry begins . . . [but] knowledge is not a modification of objects of knowledge”.

Although Dicker and Thayer have made a good start, Dewey's theory of knowledge has not yet been elaborated in a way that brings it into the mainstream of Anglo-American epistemology. Few epistemologists today appreciate what Dewey's thought can contribute to the solution of problems addressed by other theorists of knowledge.

## WRITINGS

- The Quest for Certainty*, in *John Dewey: The Later Works 1925–1953*, vol. 4, 1929 (Carbondale, IL: Southern Illinois University Press, 1984).
- The Influence of Darwin on Philosophy and Other Essays in Contemporary Thought* (New York: Henry Holt, 1910).
- Logic: The Theory of Inquiry*, in *John Dewey: The Later Works 1925–1953*, vol. 12, 1938 (Carbondale, IL: Southern Illinois University Press, 1986).

## BIBLIOGRAPHY

- Dicker, G.: *Dewey's Theory of Knowing* (Philadelphia, PA: Philosophical Monographs, 1976).
- Murphy, A. E.: "Dewey's Epistemology and Metaphysics," in *The Philosophy of John Dewey*, ed. P. A. Schilpp and L. E. Hahn, 3rd edn (Carbondale, IL: Southern Illinois University Press, 1989).
- Rorty, R.: *Philosophy and the Mirror of Nature* (Princeton, NJ: Princeton University Press, 1979).
- Thayer, H. S.: "Dewey and the Theory of Knowledge," *Transactions of the C.S. Peirce Society* 26 (1990), 443–58.

PETER H. HARE

**dialectic (Hegel)** Hegel uses several kinds of "dialectic".

*Dialectical analysis* determines the content and proper application of concepts. It begins with an elementary, general concept from a domain of inquiry, examines its content and range of application, and criticizes its failure to account for salient features of examples in its purported domain. This justifies introducing a more sophisticated concept to account for the domain, on which the analysis is

repeated. The result of dialectical analysis is an integrated network of concepts which specifies the proper domain of each and which preserves the legitimate content of earlier concepts in the final, most comprehensive and adequate concept.

For example, Hegel's *Logic* analyses concepts which purport to characterize the whole of reality. The first concept treated is "being", which is criticized for its descriptive vacuity and for connoting stasis, which ill describes a fundamental trait of reality, viz. change. These defects justify introducing the interim concept of "nothing" and then the concept of "becoming". This concept is then in its turn submitted to analysis.

*Dialectical arguments* offer indirect proof. They justify controversial principles for a domain by criticizing the simplest principle from that domain. Hegel believes that inadequacies in a principle can be generated internally – between the principle and examples from its domain. These inadequacies specify more accurately the proper range of application of the principle *and* they justify the introduction of a more sophisticated principle which purports to account for the original examples and for the pitfalls of the previous principle. The dialectical examination is then repeated. Increasingly sophisticated principles are justified by showing that they are the simplest principles which can account accurately for the relevant phenomena in the domain. Such arguments often argue regressively from an obvious phenomenon to demonstrate either necessary or sufficient conditions for the possibility of that phenomenon.

For example, Hegel's *Phenomenology of Spirit* begins its defence of a non-foundational epistemology by arguing against non-conceptual knowledge (knowledge by acquaintance) (*see* KNOWLEDGE BY ACQUAINTANCE/BY DESCRIPTION). He purports to show that no such view can account for our obvious abilities to distinguish among different objects of knowledge or to specify the relevant spatial or temporal scope of ostensive reference without admitting that concepts are essential even in the most elementary examples of human knowledge. This failure justifies introducing a view of knowledge that admits



elementary concepts for sensory qualities. This view is then submitted to analysis and used in a further indirect proof.

*Dialectical relations* hold between things, concepts or phenomena when two or more of them appear to be independent but are in fact interdependent. Typically, these dependencies would now be expressed as biconditional relations.

*Dialectical developments* occur in history or in society when a historical or social phenomenon either depends upon or generates a distinct and opposed phenomenon, where these phenomena ultimately are encompassed within a larger framework. *Dialectical explanations* explain dialectical relations or dialectical developments by emphasizing their dialectical character.

See also HEGEL.

#### BIBLIOGRAPHY

- Baum, M.: *Die Entstehung der hegelschen Dialektik* (Bonn: Bouvier, 1986).  
 Muller, G. E.: "The Hegel Legend of 'Thesis, Antithesis and Synthesis,'" *Journal for the History of Ideas* 19 (1958), 411–14.  
 Pinkard, T.: *Hegel's Dialectic: The Explanation of Possibility* (Philadelphia, PA: Temple University Press, 1988).  
 Westphal, K. R.: "Hegel's Solution to the Dilemma of the Criterion," *History of Philosophy Quarterly* 5 (1988), 173–88.  
 Wolff, M.: *Der Begriff des Widerspruchs. Eine Studie zur Dialektik Kants und Hegels* (Königstein: Hain, 1981).

KENNETH R. WESTPHAL

**dialectic (Plato)** From *dialegethai*, to converse, dialectic is in general the pursuit of philosophical issues through conversation. Pre-Platonic philosophers like Zeno and some Sophists cultivated ways of refuting opponents in discussion, and Socrates regarded conversation as essential to philosophical activity, partly because it forced people to say what they really believed. Plato uses various methods under the general heading of "asking and answering questions" (*Cratylus*

390c). In earlier works theses and definitions are scrutinized and usually refuted. In the *Meno* and *Phaedo* dialectic employs a method of "hypothesis", in which propositions about Forms are examined and provisionally accepted. In the *Republic*, however, Plato seems to think that this method can ultimately attain the certainty of an "unhypothesized" principle, seemingly concerned with the Good (510–11, 533–4). Subsequent works, like the *Phaedrus* and *Sophist*, are silent on hypothesis, and associate dialectic with "collection and division", used for dividing and classifying kinds or Forms, and so providing definitions.

See also PLATO.

#### BIBLIOGRAPHY

- Robinson, R.: *Plato's Earlier Dialectic*, 2nd edn (Oxford: Oxford University Press, 1953).  
 Stenzel, J.: *Studien zur Entwicklung der platonischen Dialektik* (Leipzig–Berlin, 1931).  
 Vlastos, G.: "The Socratic Elenchus," *Oxford Studies in Ancient Philosophy* 1 (1983), 27–58.

NICHOLAS P. WHITE

**different constructions in terms of "knows"** The question of the nature of the different constructions in terms of "knows" and of their interrelations can be approached either in terms of the general semantical and pragmatic nature of the concept of knowledge, or else through the facts of ordinary usage and of concepts taken from ordinary usage and from our *prima facie* intuitions about the alleged inferential relationships between natural language sentences. For reasons of space, the scope of this article is restricted to the logico-semantical approach. There are nevertheless reasons to believe that this is the only self-explanatory approach and that the true nature of the very concepts and generalizations of the ordinary language approach can only be understood from the semantical vantage point. (Examples of the natural language-driven approaches are found among the papers reprinted in Schwartz,

1977; and in Salmon and Soames, 1988. Cf. also Hintikka and Hintikka, 1989, ch. 9.)

What is the point of having the notion of knowledge in our conceptual repertoire in the first place? The best available answer is that we often want to restrict our attention, for instance our practical preparations, to only some of the contingencies that we might have to heed. We *know that* *S* iff we are entitled to restrict our attention to those scenarios in which *S* is true. What is meant by "entitled to" here is a question that concerns the definition of knowledge. What is discussed in this article is not affected by it. Likewise, the precise nature of the "contingencies" or "scenarios" involved here does not matter to most of what is said in this article. What is crucial is their logical type, i.e. that they are to be described by sentences (propositions) rather than by names or predicates.

This characterization presupposes that a space of possible states of affairs or courses of events is given. They are usually called, and will be called here, by the misleading term "possible worlds". Since no comprehensiveness in space or time is intended, a less misleading term would be "scenario" or perhaps even "situation".

More explicitly, what is involved in *knowing that* is a space of possible worlds on which a two-place relation is denoted for each agent (knower). A world  $W_1$  bears this relation to  $W_0$  with respect to *b* iff  $W_1$  is compatible with whatever it is that *b* knows in  $W_0$ . Such worlds are called epistemic *b*-alternatives to  $W_0$ .

Now *knowing that* can be characterized by the following truth-condition:

- (1) *b* knows that *S* in  $W_0$  iff *S* is true in each epistemic *b*-alternative to  $W_0$ .

Some fine-tuning is needed here in that *S* might be undefined in some possible worlds. This does not affect the theme of this article, however.

The usual formal counterpart " $K_b$ " for "*b* knows that" will be used here.

Besides the *knows that* construction we have to discuss also the following constructions, among others:

- (i) *b* knows whether . . .
- (ii) *b* knows who (what, which, when, where, . . .) . . .
- (iii) *b* knows how . . .
- (iv) *b* knows why . . .
- (v) *b* knows *d* (*knows* with a direct grammatical object).

The possible-worlds analysis opens the door for the analysis of all these different constructions. In some ways, it almost forces the right analysis on us. For instance, what is the logical form of the following sentence?

- (2) Jessica knows who committed the murder.

This can be taken to be of the form

- (3) *j* knows who [say *x*, is such that] *M<sub>x</sub>*.

Here (3) can apparently be analysed as

- (4)  $(\exists x) K_j M_x$

(where "*x*" is taken to range over persons).

This sounds quite plausible. What else can there be to *knowing who* is such-and-such than *knowing of some one* person that he or she is such-and-such? Let us accept (4) provisionally as an analysis of (2).

From the possible-worlds analysis of *knowing that* it is seen what is involved in (4). The knowledge-operator  $K_j$  involves several possible worlds as a member of which the individual *x* is being considered. This makes sense only on the assumption that criteria of identity across the boundaries are objectively given. Imaginary lines connecting the embodiments of the same individual in different worlds are called *world lines*.

The analysis of (2) as (4) reveals much of the logical behaviour of (1), e.g. when it is implied by a simpler sentence of the form

- (5)  $K_j M_c$ .

But can this analysis be generalized? The other *wh*-constructions (ii) differ from *knowing who* only with respect to the range of the bound variable *x*. How these ranges go together with the behaviour of quantification

in natural language is an interesting subject which nevertheless does not affect this article. (These different ranges correspond in fact closely to the different Aristotelian categories; see Hintikka, 1983b.) But can the analysis be extended to more complex cases? What is needed here is an insight into the way quantifiers operate. The crucial thing about them is not the idea of existence, nor the idea of generality, but the idea of a *dependent quantifier*, exemplified by

$$(6) \forall(x) \exists(y) Rxy$$

To understand the logic of quantification is to understand such dependent quantifiers as " $(\exists y)$ " in (6). But if you understand quantifier dependence, you *ipso facto* understand quantifier independence. What is needed is simply a notation for such independence. The independence of the quantifier  $(Q_0x)$  of certain others, say  $(Q_1y)$ ,  $(Q_2z)$ ,  $\dots$ , within the scope of which it occurs, will be indicated by writing it as

$$(7) (Q_0x/Q_1y, Q_2z, \dots)$$

This may need an explanation. (Cf. here Hintikka and Sandu, 1989.) In the conventional logical notation dependence is determined by the scope conventions, e.g. by the brackets that are associated with a quantifier. Unfortunately, the usual scope notation is unnecessarily restrictive, ruling out certain perfectly natural and interpretable possibilities, e.g. the possibility that the scopes of two quantifiers might overlap only partially, so that neither scope is completely within the other. What is needed is a way of temporarily excusing a quantifier from the scope of another, and this is precisely what the slash notation accomplishes.

A full understanding of the new notation is provided by its semantics. Such semantics can be spelled out most easily in its full generality by means of game-theoretical semantics. (See Hintikka, 1983a.) There the independence referred to here in intuitive terms becomes but a special case of the general game-theoretical concept of informational independence. The basic idea is so

intuitive, however, that the reader can be spared the technicalities. Moreover, once you grasp the basic idea, you can see how it can be extended to all other parts of our logical vocabulary.

The crucial features of the analysis of all the different *wh*-constructions (i)–(iv) can now be expressed by saying that these *wh*-words are essentially *informationally independent logical operators* of different logical types and of different categories. Indeed, the analysis of (i)–(iii) can be expressed roughly as follows:

- (i) *b* knows whether  $S_1$  or  $S_2$  iff  
(8)  $K_b (S_1 \vee K_b S_2)$
- (ii) *b* knows who (say *x*) satisfies  $S[x]$  iff  
(9)  $K_b (\exists x/K_b) S[x]$
- (iii) *b* knows how *X* is done iff  
(10)  $K_b (\exists m/K_b) (X \text{ is done by the method } m)$

A moment's reflection (or a minimal acquaintance with game-theoretical semantics) shows that (8)–(10) are logically equivalent to the following:

- (11)  $K_b S_1 \vee K_b S_2$
- (12)  $(\exists x) K_b S[x]$
- (13)  $(\exists m) K_b (X \text{ is done by the method } m)$

These do not use the idea of informational independence. Hence there is nothing intrinsically wrong about such analyses as (4).

What makes the difference is that the independence analysis, unlike the original one exemplified by (4), can be extended to more complex cases. The following are cases in point:

- (14) Albert knows whom everybody admires most.
- (15) Bob knows which of one's parents, one's father or one's mother, everyone loves more.

The most natural readings of (14)–(15) assign to them the following logical forms:

- (16)  $K_a (\forall x) (\exists y/K_a) Axy$
- (17)  $K_b (\forall x) (Fx \vee K_b Mx)$

Somewhat surprisingly, (16)–(17) do not reduce to the usual independence-free notation of epistemic logic. Why not? Why are they not for instance equivalent to the following?

$$(18) (\forall x) (\exists y/K_a) Axy$$

$$(19) (\forall x)(K_a Fx \vee K_a Mx)$$

The reasons are subtle, and cannot be given here in detail. Very roughly speaking, a closer analysis of the semantics of epistemic logic shows that quantifiers binding from the outside into the scope of “ $K_a$ ” so to speak range over individuals known to  $a$ , while there is no such restriction on other quantifiers. But if so, (18)–(19) obviously cannot be the respective logical forms of (14)–(15) whereas (16)–(17) can.

Of course, (14) has an entirely different reading on which its logical form is

$$(20) (\exists y) K_a(\forall x)Axy$$

Since there are no subtle problems about this kind of reading, it is simply disregarded here.

The nature of (16)–(17) is illuminated further by their equivalence with the following:

$$(21) (\exists f) K_a(\forall x) (Axf(x))$$

$$(22) (\exists f) K_a(\forall x) ((f(x) = 1 \ \& \ Fx) \vee (f(x) = 0 \ \& \ Mx))$$

This treatment can be extended to all other *wh*-questions. What makes the difference between the different cases is the logical type of the entities involved. Indeed, the *why*-construction (iv) can be analysed along the same lines as soon as we realize that an *explanans* is of the logical type of proposition. Roughly speaking, the analysis of

$$(23) \text{ It is known why } F$$

will be

$$(24) K(\exists S/K)(S \ \& \ N(S \rightarrow F)).$$

The details, including the nature of the necessity operator “ $N$ ”, have not yet been analysed in the literature.

The independence notation also enables us to see the difference of knowing who someone, say  $b$ , is, independently of how he or she is referred to, and knowing him or her as  $b$ . These can be expressed respectively as:

$$(25) K(\exists x/K) ((b/K) = x) \text{ and}$$

$$(26) K(\exists x/K) (b = x)$$

There is still another sense of *knowing who* in which it amounts to knowing who all the people are who satisfy a certain condition, say  $S[x]$ . This sense is also parallel with the others. It can be expressed by

$$(27) K(\exists X/K) (\forall x)(S[x] \leftrightarrow Xx)$$

One special use of the *wh*-constructions may be worth mentioning. *Knowing how* is used both (a) for knowing how something is done and (b) for knowing that *and* being able to apply to that knowledge in practice (“know-how”). The latter sense is sometimes thought to be *sui generis*. In reality, the corresponding sense of applied knowledge occurs in all the *wh*-constructions (ii) as well, as a survey of ordinary usage easily shows.

What remains is to treat the direct-object construction. Such a treatment has been available for a long time, but epistemologists have failed – or refused – to avail themselves of it, even though it possesses a better claim to psycholinguistic reality than any other logico-epistemological theory. One source of the failure is a failure to analyse the criteria of cross-identification mentioned above. Philosophers have wishfully thought that reality might do their job for them by means of causal links between objects and their names, of supplies of ready-made temporarily persisting objects, etc.

In reality, world lines can be drawn in different ways, and are so drawn in our actual conceptual system. (See *here* Hintikka and Hintikka, 1989, essay 8; Hintikka, 1975, ch. 3.) One set of world lines is the one on which the semantics of *wh*-constructions with *knows* is based. To the extent that you understand the meaning of these constructions, you *ipso facto* master these world lines, which I shall call (like the corresponding method of

identification) *public*. However, another set of world lines is drawn on the basis of some particular knower's, say *b*'s, first-hand cognitive relations to persons, objects, places, times, etc. These relations create a framework which can be used for cross-identification. In the case of visual knowledge, this framework is one's visual space. More generally, the totality of one's direct cognitive relations to other entities creates a kind of story or drama, not to say a soap opera. Persons, objects, places, etc., which play the same role in this personal soap opera can be considered identical, even if their identity is not known by the person or question and although they are therefore different public (publicly identified) persons in his or her epistemic alternatives. I shall call this method of identification (and the world lines based on it) *perspectival*. In the case of visual knowledge, it amounts to identifying persons and objects occupying the same slot in the knower's (seer's) visual space even if the perceiver does not see who or what they are.

The direct-object construction is just like the *wh*-constructions except that it relies on perspectival world lines instead of public ones. If the quantifiers relying on the former are (Ex), (Ay), etc., we have the following parallelism:

(28) *b* knows who *d* is equals

(29)  $K_b (\exists x/K_b) (d = x)$

whereas

(30) *b* knows *d* equals

(31)  $K_b (Ex/K_b) (d = x)$

As in the case of (25) we also have a slightly different reading of (30):

(32)  $K_b (Ex/K_b) ((d/K_b) = x)$

This diagnosis of the direct-object construction admits of more corroborating evidence than can be rehearsed here. Suffice it to suggest that it has even historical implications. For instance, the ancient Greek predilection for the direct-object construction is clearly connected with their preference of situation-

centred conceptualizations. Closer to home, the contrast between the two pairs of quantifiers and the two kinds of world lines is intimately related to Russell's distinction between knowledge by description (and its objects) and knowledge by acquaintance (and objects of acquaintance). (Cf. Russell in Salmon and Soames, 1988: *see* KNOWLEDGE BY ACQUAINTANCE/BY DESCRIPTION.)

The logical analyses outlined above help to put the ordinary language-driven approaches to a deeper perspective. For reasons of space, only a few such applications can be mentioned here. For instance, the difference between concepts informationally dependent and independent of *K* provides for epistemic contexts a good rational reconstruction of the *de dicto* vs. *de re* distinction, which therefore is not an irreducible and inexplicable one. Likewise, the distinction between attributive vs. referential uses of definite descriptions and other referring expressions can easily be analysed. Furthermore, the idea of a rigid designator turns out to be a mirage. First of all, it is relative to a method of drawing world lines. Russellian "logically proper names" like *this* and *that* are supposed to be "rigid designators" for perspectival reference in analogy to proper names, which are claimed to be rigid designators for public reference. Of course, in reality even proper names are not rigid designators in epistemic contexts, for one can very well fail to know who the reference of a proper name is.

Much more interesting than the shortcomings of the ordinary language driven approach are their deeper reasons. (For these reasons, *see also* Hintikka, 1991.) Some of the most crucial ingredients of a satisfactory semantical analysis of *wh*-constructions with *knows* are not expressed in most natural languages with any standard syntactical device. This is especially true of the phenomenon of informational independence, which occurs in so many different grammatical and logical categories that it is so to speak beyond the ingenuity of the linguistic community to express it by any uniform syntactical device.

But if a distinction is not expressed by any uniform syntactical construction or other syntactical device, an approach which starts



from the syntax of natural language and relies heavily on syntactical generalizations will inevitably overlook phenomena like informational independence and therefore *inter alia* the right analysis of *wh*-constructions with *knows*.

Once this is realized, further peculiarities of the logical forms of more complex *wh*-constructions with *knows* can be mastered. All told, there are interesting general theoretical reasons why the foot of the letter – meaning the foot of the syntax of natural language – is not the right starting place for an analysis of the different constructions with *knows*.

The analysis outlined here is an improvement on such earlier treatments as Hintikka (1975, ch. 1 and 1976). Most of the details of earlier discussions remain unaffected by the improvements, however.

The treatment of *knows who* sketched here should be compared with Boër and Lycan (1986). For the linguistic theory of *wh*-constructions, cf. e.g. Hirschbühler (1979).

## BIBLIOGRAPHY

- Boër, S. E. and Lycan, W. G.: *Knowing Who* (Cambridge, MA: MIT Press, 1986).  
 Hintikka, J.: *The Intentions of Intentionally* (Dordrecht: Reidel, 1975), esp. chs 1 and 3.  
 Hintikka, J.: *The Semantics of Questions and the Questions of Semantics* (Helsinki: Societas Philosophica Fennica, 1976).  
 Hintikka, J.: *The Game of Language* (Dordrecht: Reidel, 1983a).  
 Hintikka, J.: "Semantical Games, the Alleged Ambiguity of 'is', and Aristotelian categories," *Synthese* 54 (1983b), 443–68.  
 Hintikka, J.: "Paradigms for Language Theory," *Acta Philosophica Fennica* (1991).  
 Hintikka, J. and Hintikka, M. B.: *The Logic of Epistemology and the Epistemology of Logic* (Dordrecht: Kluwer Academic, 1989), esp. essays 8 and 9.  
 Hintikka, J. and Sandu, G.: "Informational Independence as a Semantical Phenomenon," *Logic, Methodology and Philosophy of Science VIII*, ed. J.E. Fenstad, I. T. Frolov and R. Hilpinen (Amsterdam: North-Holland, 1989), 571–89.  
 Hirschbühler, P.: *Syntax and Semantics of Wh-Constructions* (Bloomington, IN: Indiana University Linguistics Club, 1979).  
 Lewis, D.: *Philosophical Papers*, vol. 1 (New York: Oxford University Press, 1983).

- Schwartz, S. P., ed.: *Naming, Necessity, and Natural Kinds* (Ithaca, NY: Cornell University Press, 1977).  
 Salmon, N. and Soames, S., eds: *Propositions and Attitudes* (Oxford: Oxford University Press, 1988).

JAAKKO HINTIKKA

**direct realism** A view about what the objects of perception are. Direct realism is a type of realism (*see* REALISM), since it is assumed that these objects exist independently of any mind that might perceive them; and so it thereby rules out all forms of idealism and phenomenalism, which hold that there are no such independently existing objects. Its being a "direct" realism rules out those views defended under the rubric of "critical realism", or "representative realism" (*see* REPRESENTATIVE REALISM), in which there is some non-physical intermediary – usually called a "sense-datum" or a "sense impression" – that must first be perceived or experienced in order to perceive the object that exists independently of this perception. According to critical realists, such an intermediary need not be perceived "first" in a temporal sense, but it is a necessary ingredient which suggests to the perceiver an external reality, or which offers the occasion on which to infer the existence of such a reality. Direct realism, on the other hand, denies the need for any recourse to mental go-betweens in order to explain our perception of the physical world.

Often the distinction between direct realism and other theories of perception is explained more fully in terms of what is "immediately" perceived, rather than "mediately" perceived. The terms are Berkeley's (*see* BERKELEY), who claims (1713, p. 46) that one might be said to hear a coach rattling down the street, but this is mediate perception as opposed to what is "in truth and strictness" the immediate perception of a sound. Since the senses "make no inferences", the perceiver is then said to infer the existence of the coach, or to have it suggested to him by means of hearing the sound. Thus, for Berkeley, the distinction between mediate and immediate perception is explained in terms of whether or not either inference or suggestion is present in the perception itself.

Berkeley went on to claim that the objects of immediate perception – sounds, colours, tastes, smells, sizes and shapes – were all “ideas in the mind”. Yet he held that there was no further reality to be inferred from them; so that the objects of mediate perception – what we would call “physical objects” – are reduced to being simply collections of ideas. Thus Berkeley uses the immediate-mediate distinction to defend idealism (see IDEALISM). A direct realist, however, can also make use of Berkeley’s distinction to define his own position. D. M. Armstrong (see ARMSTRONG) does this by claiming that the objects of immediate perception are all occurrences of sensible qualities, such as colours, shapes and sounds; and these are all physical existents, and not ideas or any sort of mental intermediary at all (Armstrong, 1961, pp. xii, 23). Physical objects, all mediately perceived, are the bearers of these properties immediately perceived.

Berkeley’s and Armstrong’s way of drawing the distinction between mediate and immediate perception – by reference to inference or the lack of it – houses major difficulties. We are asked to believe that some psychological element of inference or suggestion enters into our mediate perception of physical objects such as coaches and camels. But this is implausible. First, there are cases in which it is plausible to assert that someone perceived a physical object – a tree, say – even when that person was unaware of perceiving it. (We can infer from his behaviour in carefully walking around it that he did see it, even though he does not remember seeing it.) Armstrong would have to say that in such cases inference was present, because seeing a tree would be a case of mediate perception; although here it would have to be an unconscious inference. But this seems baseless; there is no empirical evidence that any sort of inference was made at all.

Second, it seems that whether a person infers the existence of something from what he perceives is more a question of talent and training than it is a question of what the nature of the objects inferred really is. For instance, if we have three different colour samples, a trained artist might not have to infer

their difference; instead, he might see their difference immediately. Someone with less colour sense, however, might see patches A and B as being the same in colour, and patches B and C as being the same, but also see that A is darker than C. On this basis he might then infer that A is darker than B, and B darker than C; and so inference might be present in determining difference in colour, but colour was supposed to be an object of immediate perception. On the other hand, a park ranger might not have to infer that the animal he sees is a Florida panther; he sees it to be such straightaway. Someone unfamiliar with the Everglades, however, might have to infer this from the creature’s markings. Hence, inference need not be present in cases of perceiving physical objects; yet perception of physical objects was supposed to be mediate perception.

A more straightforward way to distinguish between different objects of perception was advanced by Aristotle in *De Anima* (p. 567) (see ARISTOTLE), where he spoke of objects directly or essentially perceived as opposed to those objects incidentally perceived. The former comprise perceptual properties, either those discerned by only one sense (the “proper sensibles”), such as colour, sound, taste, smell and tactile qualities, or else those discerned by more than one sense, such as size, shape, and motion (the “common sensibles”). The objects incidentally perceived are the concrete individuals which possess the perceptual properties; that is, particular physical objects.

According to Aristotle’s direct realism, we perceive physical objects incidentally; that is, only by means of the direct or essential perception of certain properties that belong to such objects. In other words, by perceiving the real properties of things, and only in this way, can we thereby be said to perceive the things themselves. These perceptual properties, though not existing independently of the objects that have them, are yet held to exist independently of the perceiving subject; and the perception of them is direct in that no mental messengers have to be perceived or sensed in order to perceive these real properties.

Aristotle’s way of defining his position seems superior to the psychological account

offered by Armstrong, since it is unencumbered with the extra baggage of inference or suggestion. Yet a common interpretation of the Aristotelian view leads to grave difficulties. This interpretation identifies the property of the perceived object with a property of the perceiving sense organ. It is based on Aristotle's saying that in perception the soul takes in the form of the object perceived without its matter (*ibid.*, p. 580). On this interpretation, it is easy to think of direct realism as being committed to the view that "colour as seen", or "sound as heard" were independently existing properties of physical objects. But such a view has been rightly disparaged by its critics and labelled as "naïve realism"; for this is a view holding that the way things look or seem is exactly the way things are, even in the absence of perceivers to whom they appear that way.

The chief difficulty of naïve realism is well presented by an argument of Bertrand Russell's (Russell, 1962, p. 9) (*see* RUSSELL). Russell claims that an ordinary table appears to be of different colours from different points of view and under different lighting conditions. Since each of the colours appearing has just as much right to be considered real, we should avoid favouritism and deny that the table has any one particular colour. Russell then went on to say the same sort of thing about its texture, shape and hardness. All of these qualities are what we might call "appearance determined" qualities; that is, they are not real independent of how they appear to perceivers; so the real table, for Russell, was something apart from the directly perceived colours, sounds, smells and tactual qualities – all of which Russell termed "sense-data". It is from these sense-data that Russell believed that we inferred the existence of physical objects.

Russell's argument, however, only works against the "naïve" version of direct realism. It should first be noted that the argument does not show that the table has no real colour, shape, or texture, but only that we might not know which of the apparent properties are the real properties of the table. So the most that Russell can prove with his argument is that we must remain sceptical about the

real properties of the table; but this might be enough to show that we have no right to talk about its real properties at all. If we did have some way of determining which were the real properties, however, then Russell's argument loses its sting. A step towards making this determination this can be taken by questioning Russell's initial supposition that some perceiver-dependent properties might turn out to be real properties. To agree with this assumption is to fall into the error of naïve realism. Instead, the clear-headed direct realist would be on safer ground in denying that the directly apprehended real properties are "colours as seen", "sounds as heard", or "textures as felt"; for this is to confuse real properties of things with the appearances they present to perceivers.

The direct realist should instead begin by insisting that real properties are not perceiver-dependent. This would mean that if colour is to be a real property, it must be specified in terms that do not require essential reference to the visual experience of perceivers. One way to do this would be to identify the colour of a surface with the character of the light waves emitted or reflected from that surface (Armstrong, 1968, p. 283). This would be an *empirical* identification; that is, the predicate "is coloured" and the predicate "reflects or emits light of a certain wave length" would refer to the one and the same property.

To say, then, that fire engines are red even at night would be to say that their surfaces, under normal conditions of illumination, would reflect light at the red end of the colour spectrum. This is still compatible with saying that they are not red in the dark, in that they are not now reflecting any such light. This gets around Russell's problem about choosing the "real colour" of an object. Another way to make this point is to say that the "standing colour" of fire engines remains red no matter what the conditions of illumination; whereas, their "transient colour" changes according to changes in such lighting conditions (Armstrong, 1968, pp. 284–5).

Similar reductions could be made with regard to the other sensible properties that seemed to be perceiver-dependent: sound

could be reduced to sound waves, tastes and smells to the particular shapes of the molecules that lie on the tongue or enter the nose, and tactual qualities such as roughness and smoothness to structural properties of the objects felt. All of these properties would be taken to be distinct from the perceptual experiences that these properties typically give rise to when they cause changes in the perceiver's sense organs. When critics complain that such a reduction would "leave out the greenness of greens and the yellowness of yellows" (Campbell, 1976, p. 67), the direct realist can answer that it is by identifying different colours with distinct light waves that we can best explain how it is that perceivers in the same environment, with similar physical constitutions, can cite similar colour *experiences* of green, or of yellow.

If such a general reductive programme could be made plausible, it would show that Locke's "secondary qualities" – colour, sound, taste and smell – were really "primary qualities" after all, in that they could be specified apart from their typical effects on perceivers (Locke, 1690, p. 135) (*see* PRIMARY AND SECONDARY QUALITIES). A direct realist could then claim that one directly perceives what is real only when there is no difference between the property proximately impinging on the sense organ and that property of the object which gives rise to the sense organ's being affected. For colour, this would mean that the light waves reflected from the surface of the object must match those entering the eyes; and, for sound, it means that the sound waves emitted from the object must match those entering the ear. A difference in the property at the object from that at the sense organ would result in illusion, not veridical perception. Perhaps this is simply a modern version of Aristotle's idea that in genuine perception the soul (now the sense organ) takes in the form of the perceived object.

If it is protested that illusion might also result from an abnormal condition of the perceiver, this can also be accepted. If one's colour experience deviated too far from normal, even when the physical properties at the object and the sense organ were the same, then

misperception or illusion would result. But such illusion could only be noted against a backdrop of veridical perception of real properties. Thus, the chance of illusion due to subjective factors need not lead to Democritus' view of colours, sounds, tastes, and smells as existing merely "by convention". The direct realist could insist that there must be a real basis in veridical perception for any such agreement to take place at all; and veridical perception is best explained in terms of the direct perception of the properties of physical objects. It is explained, in other words, when our perceptual experience is caused in the appropriate way.

This reply on the part of the direct realist does not, of course, serve to refute the global sceptic, who claims that, since our perceptual experience could be just as it is without there being any real properties at all, we have no knowledge of any such properties. But no view of perception alone is sufficient to refute such global scepticism (Pitcher, 1971, p. 219). For such a refutation we must go beyond a theory that claims how best to explain our perception of physical objects, and defend a theory that best explains how we obtain knowledge of the world.

*See also* ARISTOTLE; ARGUMENT FROM ILLUSION; EXPERIENCE; PERCEPTUAL KNOWLEDGE; PROBLEM OF THE EXTERNAL WORLD; REPRESENTATIVE REALISM; SARTRE; SENSATION/COGNITION.

#### BIBLIOGRAPHY

- Aristotle: *De Anima*, trans. J. A. Smith in *The Basic Works of Aristotle*, ed. R. McKeon (New York: Random House, 1941).
- Armstrong, D. M.: *Perception and the Physical World* (London: Routledge and Kegan Paul, 1961).
- Armstrong, D. M.: *A Materialist Theory of the Mind* (London: Routledge and Kegan Paul, 1968).
- Armstrong, D. M.: *The Nature of Mind and Other Essays* (Ithaca, NY: Cornell University Press, 1981).
- Berkeley, G.: *Three Dialogues between Hylas and Philonous* (1713), in *Berkeley: Philosophical Works*, ed. M. R. Ayers (London: Dent, 1975).
- Campbell, K.: *Metaphysics: An Introduction* (Encino: Dickinson, 1976).

Carrier, L. S.: *Experience and the Objects of Perception* (Washington, DC: University Press of America, 1981).  
 Locke, J.: *An Essay Concerning Human Understanding* (1690); ed. P. H. Niddich (Oxford: Oxford University Press, 1975).  
 Pitcher, G.: *A Theory of Perception* (Princeton, NJ: Princeton University Press, 1971).  
 Russell, B.: *The Problems of Philosophy* (1912); 28th impr. (Oxford: Oxford University Press, 1962).

L. S. CARRIER

**disposition** A lump of sugar is cubical, has mass, and is soluble. It is common to distinguish the third of these properties from the first two. Solubility is a *dispositional* as distinguished from a *categorical* trait. In ascribing solubility to sugar, we ascribe to it a tendency: It would dissolve were it placed in a (suitable) liquid.

Ryle (1949) (*see* RYLE) argues that certain states of mind are best regarded dispositionally. Believing, for instance, is not a matter of consciously entertaining thoughts, but of being disposed to say and do various things (including entertaining thoughts), depending on the circumstances. Ryle, however, seems to deny that dispositions are genuine features of the items possessing them: "Dispositional statements are neither reports of observed or observable states of affairs nor yet reports of unobserved or unobservable states of affairs" (p. 125).

"Realists" about dispositions come in two flavours. Some suppose dispositions to have a "basis" in non-dispositional features of objects (Averill, 1990; Prior, 1985). Others (e.g. Goodman, 1955; Mellor, 1974) forgo the dispositional-categorical distinction, and take physical properties to be dispositions, some of which have a basis in other dispositions.

*See also* BEHAVIOURISM.

#### BIBLIOGRAPHY

Averill, E. W.: "Are Physical Properties Dispositions?," *Philosophy of Science* 57 (1990), 118–32.

Goodman, N.: *Fact, Fiction, and Forecast* (London: 1955); 2nd edn (Indianapolis, IN: Bobbs-Merrill, 1965).  
 Mellor, D. H.: "In Defense of Dispositions," *Philosophical Review* 53 (1974), 157–81.  
 Prior, E.: *Dispositions* (Atlantic Highlands, NJ: Humanities Press, 1985).  
 Ryle, G.: *The Concept of Mind* (London: Hutchinson, 1949).

JOHN HEIL

**dogmatism** Like "fanaticism", dogmatism is ordinarily a term of abuse, and a term one doesn't apply to oneself. (How often do you hear someone describe himself as a dogmatist?) The term has a variety of analogically related uses. In one use, to say of someone that she is a dogmatist is to say that she holds her views more strongly than is appropriate, more strongly than the evidence warrants, for example; alternatively, it is to say that she holds her views uncritically, without paying sufficient heed to objections and alternatives, or to the limitations of human reason (*see* KANT). "Dogmatism" is therefore an indexical term; whether you properly apply it to a given doctrine or belief depends upon where you yourself stand.

ALVIN PLANTINGA

**doxastic voluntarism** The thesis that our doxastic attitudes – belief, disbelief, and suspension of judgment – are at least sometimes under our voluntary control. This thesis conflicts with the thesis of doxastic involuntarism, which states that doxastic attitudes are never under our voluntary control. Although doxastic voluntarism is controversial, the view has enjoyed an impressive list of advocates throughout the history of philosophy, including Augustine, Aquinas, Clifford, Descartes, and Locke. More recently, Chisholm (1966), Ginet (2001), and Steup (2000, 2008) have championed the view.

The motivation for doxastic voluntarism derives, to a large extent, from the appeal of internalism and epistemic deontologism. According to the latter view, the justification of our beliefs should be understood in terms of



what we *ought* to believe, what we have a *right* to believe, or what we are *permitted* to believe. Internalism is, roughly, the view that whether one is justified in believing *p* is determined by one's reasons, which should be recognizable on reflection. Deontologism motivates doxastic voluntarism because, if "ought" implies "can", then the idea that we can be held responsible for our beliefs seems to entail that we are in control of the doxastic attitudes that we form. Internalism arguably motivates doxastic voluntarism indirectly by motivating deontologism. As long as one's doxastic attitudes fit one's internally accessible reasons, one has fulfilled one's doxastic duty (Plantinga, 1993; Alston, 1989; Goldman, 2001). Both rationales for doxastic voluntarism are controversial. Feldman (2001) has rejected the claim that deontologism rests on the plausibility of doxastic voluntarism (see also Feldman and Conee, 1985); Fumerton (1988, 1995) and Kornblith (1982) have rejected the claim that internalism rests on the plausibility of deontologism.

If there is indeed a tight connection between internalism and deontologism, and if it is true that the plausibility of deontologism rests on the plausibility of doxastic voluntarism, then whether doxastic voluntarism is true is an important question in epistemology. But, even if these connections fail to hold, whether or not our doxastic attitudes are under our control is an interesting question in its own right, bearing on issues in action theory, the philosophy of psychology, and the philosophy of religion.

In the philosophy of religion, there are at least two major problems that raise the issue of doxastic voluntarism. Consider Pascal's Wager. Pascal argued that, although we will never have sufficient evidence to justify belief in God's existence, it is nevertheless prudentially rational for us to believe that God exists. His reasoning goes as follows:

- (i) If we believe that God exists, and God actually exists, we have a lot to gain.
- (ii) If we believe that God exists and God does not exist, we have nothing much to lose.
- (iii) If we do not believe that God exists, and

God does not exist, we have nothing much to lose.

- (iv) However, if we do not believe that God exists, and God does exist, we are in serious trouble.

Thus, just considering the practical payoff, we should believe that God exists. By doing so, we shall win big or else we shall not lose much. If we fail to believe, we shall suffer an enormous loss or else we shall not lose much.

Pascal's argument has been challenged on many grounds. One objection to it states that his recommendation to believe in God is pointless because doxastic involuntarism is true: Since we do not have voluntary control over our beliefs, we cannot just *decide* to believe in God. This objection can be rebutted in three ways: (i) Even if doxastic voluntarism is false, we can still engage in certain religious practices that would eventually bring about the right religious beliefs in us. We can, some philosophers would argue, indirectly control our beliefs. (ii) Even if doxastic involuntarism is true, and Pascal's advice is therefore pointless, it might still be true that believing in God's existence is in our best interest. (iii) Doxastic involuntarism is false. We do have voluntary control over our beliefs. It is, therefore, within our power to follow Pascal's advice.

Next, consider the issue of salvation. We might think that our salvation depends on whether we believe or fail to believe that God exists. This view arguably hinges on the truth of doxastic voluntarism. If our beliefs are not under our voluntary control, then it is difficult to reconcile divine punishment or reward with the concept of a just and loving God. It seems unfair to punish someone for something outside his or her control. If

- (i) "ought" implies "can" as well as "can refrain",

and owing to doxastic involuntarism

- (ii) we cannot believe what in fact we do not believe and cannot refrain from believing what in fact we do believe,

then atheists should not be punished for disbelieving in God, and theists should not be rewarded for believing in God. So the problem of salvation is this: If it is indeed true that we lack voluntary control over our beliefs, how could God be fair in making salvation depend on belief in his existence? Among the numerous responses to this problem, one would be to deny that that we lack voluntary control over our beliefs.

As noted above, doxastic voluntarism is a matter of considerable dispute. Williams (1973) has argued that it is *conceptually impossible* for beliefs to be under our control. Williams's reasoning in support of this claim can be summed up as follows:

- (1) Belief is, *by definition*, aimed at truth. [Premise]
- (2) If beliefs were under our voluntary control, beliefs could be adopted without regard for truth and on the basis of any reason at all. [Premise]
- (3) Adopting beliefs without regard for truth and on the basis of any reason at all is impossible. [From (1)]

Therefore:

- (4) Beliefs cannot be under our voluntary control. [From (2) and (3)]

In response to this argument, voluntarists would reject premise (2) because doxastic voluntarism does not imply that one can believe for any reason at all. Just because one is unable to do *x* for just *any* reason, it does not follow that doing *x* is not under one's control. For example, typically one cannot cut off one's own leg for just any reason. This does not show that cutting off one's own leg is never under one's control, for we can imagine circumstances in which a person *can* cut off one's leg because it would be rational for one to do so.

Premise (1) is problematic as well. Suppose we interpret (1) as follows:

- (1a) Believing *p* =<sub>df</sub> believing *p* is true.

On this interpretation, the second premise should read as follows:

- (2a) If beliefs were under our voluntary control, then one could adopt a belief *p* without believing *p* is true.

However, doxastic voluntarism is a thesis about control. The thesis says that believing *p*, or believing *p* is true, is sometimes under the believer's control. The thesis does not imply that one could believe *p* without believing *p* is true. Doxastic voluntarists, therefore, have no reason to accept (2a).

According to an alternative interpretation, the first premise amounts to the following:

- (1b) Believing *p* =<sub>df</sub> believing, on the basis of excellent reasons, that *p* is true.

On this second interpretation, premise (2) turns into:

- (2b) If beliefs were under our voluntary control, then one could adopt a belief *p* without, on the basis of excellent reasons, believing that *p* is true.

(1b) is not a plausible definition of a belief. Sometimes, people believe *p* when they do not have excellent reasons for believing *p*. And (2b) is problematic because doxastic voluntarism has no implications about the basis upon which one forms one's beliefs.

There might be further interpretations of the first premise. It seems doubtful, however, that there is an interpretation that meets the following two desiderata: (i) that it supply us with a plausible definition of belief; (ii) that it result in a second premise asserting something that doxastic voluntarism actually implies.

Doxastic voluntarism has also been rejected on the grounds that, while the view is not conceptually impossible, it is in fact empirically false. Alston (1989) contrasts the forming of a doxastic attitude with actions that we regard to be under our voluntary control. He claims that forming a doxastic attitude is not at all like these actions. For example, while we can just raise our hand if we want to, we cannot just believe that the number of stars is even or that  $1 + 1 = 85$ . Audi (2001) adds that forming a belief is similar to the

formation of a damp ring on a coaster under a glass of water (see also Feldman, 2001; and Plantinga, 1993).

In response, it could be argued that doxastic voluntarism is consistent with the empirical facts about belief formation if doxastic control is construed in analogy to compatibilism as a view of free action. Steup (2008) appeals to the kind of reasons-responsiveness compatibilism advocated in Fischer and Ravizza (1998). Roughly, this is the view that one's actions are free as long as they are responsive to one's reasons. Steup applies this approach to our doxastic attitudes: to the extent that they result from reason-responsive cognitive processes, they are under our direct, voluntary control. Thus, even if Alston is right in thinking that we cannot form doxastic attitudes in opposition to what we take our evidence to support, that does not show that our doxastic attitudes are not under our control. In fact, according to Steup, our doxastic attitudes' being determined by our appreciation of the evidence is all the more reason to regard them as being under our voluntary control.

Ryan (2003) argues for a similar conclusion, using different compatibilist ideas about free will. According to Ryan, doxastic attitudes that are not the result of brainwashing and the like are formed intentionally. And, as long as a belief is formed intentionally, it is, Ryan argues, under the believer's control, even if it is determined by the believer's evidence.

#### BIBLIOGRAPHY

- Alston, W.: *Epistemic Justification: Essays in the Theory of Knowledge* (Ithaca, NY: Cornell University Press, 1989).
- Audi, R.: "Doxastic Voluntarism and the Ethics of Belief," in *Knowledge, Truth, and Duty*, ed. M. Steup (Oxford: Oxford University Press, 2001), 93–111.
- Chisholm, R.: *Theory of Knowledge* (Englewood Cliffs, NJ: Prentice-Hall, 1966).
- Conee, E. and Feldman, F.: "Evidentialism," *Philosophical Studies* 48 (1985), 15–35.
- Feldman, R.: "Voluntary Belief and Epistemic Evaluation," in *Knowledge, Truth, and Duty*, ed. M. Steup (Oxford: Oxford University Press, 2001), 77–92.
- Fischer, J. M. and Ravizza, M.: *Responsibility and Control* (Cambridge: Cambridge University Press, 1998).
- Fumerton, R.: "The Internalism/Externalism Controversy," *Philosophical Perspectives* 2 (1988), 443–59.
- Fumerton, R.: *Metaepistemology and Skepticism* (Lanham, MD: Rowman and Littlefield, 1995).
- Ginet, C.: "Deciding to Believe," in *Knowledge, Truth, and Duty*, ed. M. Steup (Oxford: Oxford University Press, 2001), 63–76.
- Goldman, A.: "Internalism Exposed," in *Knowledge, Truth, and Duty*, ed. M. Steup (Oxford: Oxford University Press, 2001), 115–33.
- Kornblith, H.: "The Psychological Turn," *Australasian Journal of Philosophy* 74 (1982), 238–53.
- Plantinga, A.: *Warrant: The Current Debate* (Oxford: Oxford University Press, 1993).
- Pojman, L.: *Religious Belief and the Will* (London and New York: Routledge and Kegan Paul, 1986).
- Ryan, S.: "Doxastic Compatibilism and the Ethics of Belief," *Philosophical Studies* 114 (2003), 47–79.
- Steup, M.: "Doxastic Voluntarism and Epistemic Deontology," *Acta Analytica* 15 (2000), 25–56.
- Steup, M.: *Knowledge, Truth, and Duty* (Oxford: Oxford University Press, 2001).
- Steup, M.: "Doxastic Freedom," *Synthese* 161 (2008), 375–92.
- Williams, B.: *Problems of the Self* (New York: Cambridge University Press, 1973).

SHARON RYAN

**Dutch book argument** see BAYESIAN EPISTEMOLOGY in Part I; PROBABILITY.

# E

---

**empiricism** An epistemological movement according to which: (1) nothing around us can be known to be real unless its existence is revealed in or inferable from information we gain directly in sense experience, or in introspection, or later recall; and (2) genuine, intelligible differences in our claims about this world must express these knowable differences in experience, so that the truth of rival hypotheses ("the world rests on an elephant" vs. "it rests on a tortoise") must make a potential difference to experience or their terms (e.g. "tortoise", "elephant") must be differentially definable in terms of experience.

The empiricist movement found its greatest proponents in the seventeenth and eighteenth centuries (*see* LOCKE; BERKELEY; HUME). However, since then it has continued to have many important and influential advocates and sympathizers (*see* MILL; JAMES; RUSSELL; LEWIS; CARNAP; AYER; QUINE).

As a result of their constraints on knowledge and meaning, empiricists tend to be sceptical of necessary truths that are independent of mind and of language, and of putative eternal abstract entities (e.g. forms or universals of justice, triangularity, etc.) that these truths allegedly concern. (*See* A PRIORI KNOWLEDGE *in Part I*; MATHEMATICAL KNOWLEDGE.) They are particularly sceptical of putative faculties of intellectual, non-sensuous intuition through which such things are allegedly known. Empiricism, like any philosophical movement, is often challenged to show how its claims about the structure of knowledge and meaning can themselves be intelligible and known within the constraints it accepts (*see* PROBLEM OF THE CRITERION). For example, the truth of empiricism itself doesn't seem to be something that can be known *a priori*; but, if it is known empirically, how can empiricism avoid vicious circularity? And are the

claims of empiricism themselves empirically intelligible?

Empiricists furthermore need to tell us (1) what is revealed directly in sense experience and how, and (2) what can be inferred from it and how. The general makeup and workings of the world (e.g. there are crows, all crows are black, crows look black in standard observation conditions, there are electrons, etc.) are thought to be inferable from more particular facts. However, for traditional empiricists, only what we are immediately presented with in consciousness, or are immediately aware of, is non-inferentially knowable. In current sense experience, this consists of the occurrence of particular sensible qualities (that colour, that shape, that flavour, that odour, etc.). From past experience, it consists of those we immediately recall, perhaps less vividly. We might describe an experience as one of "that rotten-egg smell" or of "seeming to smell a rotten egg". However, it is typically thought, what is not revealed directly is whether the odour is actually that of a rotten egg or ever has been, or whether it even inheres in or is produced by anything that exists independently of our experience. Any of these claims go beyond the non-inferentially knowable sensory character of our experience.

As a result of distinguishing so sharply between the inferentially and the non-inferentially knowable, empiricists are always challenged to show how the former can be derived from the latter without making general assumptions about the world and how it affects our experience. If these assumptions are somehow warranted non-empirically, empiricism seems compromised. If they aren't, scepticism (*see* SCEPTICISM) must be embraced. Sometimes scepticism (concerning God, universals, even matter) seems welcome; at other times it isn't so welcome.

Hume forcefully posed the problem (*see* HUME). Traditional empiricism notes that we know the nature of the odour directly but can only infer its probable cause – the egg. Yet any inference must work by simple induction from experienced correlations of that odour and rotten eggs, and this is only possible if we have been able to acquire information about the presence of rotten eggs directly in experience – which traditional empiricism claims to be impossible (*see* PROBLEM OF THE EXTERNAL WORLD).

Some empiricists commonsensically expand the non-inferentially knowable to include what we automatically take ourselves to perceive about our environment on particular occasions (e.g. that we smell a rotten egg). These ordinary judgements made in sense experience are legitimate without being inferable from something more basic in experience. In particular cases, additional evidence may perhaps show them to be defective and illegitimate, but occasional mistakes aren't reason for denying that our senses can directly inform us of what exists outside us. Indeed, as empiricists themselves recognize, our immediate recollections of past experience may be mistaken, but, on pain of massive scepticism, that can't be a reason to deny that memory can be a non-inferential source of information about past experience. Further, although science may challenge our common-sense claims about the nature of the objects we see and their properties, we may none the less be assured that we see objects outside us with properties responsible for the way we experience them. Locke, for instance, argued that we can have non-inferential knowledge and that there is at least something there affecting our senses. However, wherever exactly we place the inferential/non-inferential distinction, Hume's problems reappear. For example, how do we know that the black thing we see is an egg, that the black egg we see is the same egg as the rotten one we smell, and the same egg as the white one we saw yesterday, that exposure to warm air caused the egg to rot, that that was so because of such and such physical or chemical constitution of the egg, etc.?

Other empiricists expand the forms of legitimate non-deductive inference instead. After

all, as Hume argued, the empiricist credentials of simple induction are suspect. It seems to rest on conceivably false assumptions for which, on pain of circularity, we can't have empirical evidence, e.g. that nature tends to be uniform, that our experience isn't idiosyncratic (*see* PROBLEMS OF INDUCTION). If induction isn't privileged, perhaps, as Russell (1912) suggested, particular hypotheses concerning persisting objects and processes distinct from but responsible for our experience are warranted because they simplify and explain the regularities and irregularities in our experience (*see* INFERENCE TO THE BEST EXPLANATION). Indeed, science may provide us over time with explanatory hypotheses that are better than, and radically different from, those of common sense.

Of course, empiricists might also expand both the non-inferentially knowable and the forms of non-deductive inference. In any case, the reliability of particular inductions and simplifications in common sense and science, Russell thought, depends on the truth of very general postulates about the regularities that hold in the world, in experience, and in the relation between the two. These matters of fact couldn't in turn be established by experience without circularity. Whatever truth there was in empiricism as an account of how we reliably gain information about the world couldn't itself be established by experience, although it couldn't be refuted by experience, either. Some empiricists might claim that the unrefuted reliability of such sources of information is sufficient for them to provide us with knowledge or at least warranted belief (*see* RELIABILISM). However, Russell, like most traditional epistemologists, demanded more. This left him in no better position than Hume. Sometimes he suggested that these general postulates of science and common sense might be somehow non-empirically known, sometimes instead that they might be warranted by fitting the data and rendering as coherent as possible our instinctive beliefs and particular scientific inductions and simplifications. Limitations to empiricism needed acknowledgement, however reluctantly.

Many empiricists, however, change the nature of the inference by reducing the



content of our claims about reality to claims about appearance. Berkeley (*see* BERKELEY) suggested that claims about cherries I don't experience were just claims about what experiences other persons, possibly divine, do and don't have or imagine. Mill (*see* MILL), along with most empiricists since, suggested instead that they were really just complex claims about what sense experiences one would have (e.g. experiences of red or sweetness) were one to have others (e.g. experiences of opening one's eyes or putting something to mouth) (*see* PHENOMENALISM). They were merely claims about the *possibilities* of experience. They weren't claims about a material world distinct from these possibilities, yet responsible for them. Thus, they could be warranted by straightforward induction (including analogy) from experienced and remembered patterns of experience. Induction clearly remains a problem for this form of empiricism. However, the standard objection to the phenomenalist construal of empiricism, due to Chisholm (*see* CHISHOLM), is that the permanent possibility of material objects looking different in different material conditions shows that material object claims can't be equivalent in meaning to claims solely about actual and possible experience. This charge has led many to abandon empiricist epistemology altogether.

For other empiricists such as Quine (*see* QUINE), the fact that particular hypotheses explain and predict experience only in conjunction with other particular hypotheses about the circumstances of observation, and with general hypotheses about how the world and observers interact, shows only that empiricists can't reduce hypotheses one by one to possible experience. Hypotheses can always be saved in the face of recalcitrant experience by revising other hypotheses. Apart from observation sentences, individual claims in isolation from others don't get tested by experience and don't bear empirical significance. Only sets of hypotheses or theories do. This position is known as holism (*see* HOLISM).

Indeed, empiricism may survive its traditional foundations. For Quine the content of non-inferentially warranted observation

reports is behaviouristically determined by ranges of sensory stimulation prompting the unreserved assent of speakers, rather than by sensuous qualities presenting themselves to our immediate awareness. There may be further hope in holism. Since the empirical significance and test of our theories ultimately lies in the consequences their truth has for experience, warrant for our theories seems ultimately a matter of the warrant which experience so far provides for predicting these consequences. Natural science appeals to natural selection to explain why induction over certain patterns of stimulation is reliable and thus why these patterns of observation and induction are warranted. Of course, natural science is itself the product of induction and simplifying hypothesizing, and warranted as a result (*see* NATURALIZED EPISTEMOLOGY).

Empiricist strictures are clearly preserved in the genesis of knowledge, but less clearly in its warrant. Does science, by allowing us to give an explanation of the reliability of our forms of observation and induction, show why particular observations and inductions are warranted, and thus confirm, alter, and increase the warrant of particular observations, hypotheses, and inductions? Or does it provide all their warrant, even initially? In either case, the truth in empiricism has become a contingent matter not known *a priori* but revealed by empirical science, and its norms are perhaps relevant only to us and the sentient creatures from whom we have evolved. If so, we don't need to embrace the claims of poetry over empirical science to challenge empiricism's credentials. Particularly objectionable perhaps, even on scientific grounds, is the apparent vision of our cognitive and linguistic lives as consisting in learned verbal responses to stimuli, the successful prediction and control of which is the primary function of acquiring these dispositions, however much the patterns they manifest reflect inherited and uninherited environmental adaptations.

Some empiricists, inspired by Berkeley, have thought talk of material objects as things over and above actual and possible experience quite unintelligible, not simply

dubiously warranted. Talk of purple triangles can make sense without experience of purple triangles because we can think of what appropriately compares and contrasts with purple squares and non-purple triangles in our experience. Perhaps (possibly *contra* Berkeley), we can even abstract from such comparisons and contrasts to form a way of thinking of purple triangles in general. However, according to the kind of empiricism under consideration here, experience cannot generate a conception of entirely unexperienced (unperceived, unimagined, unconceived) items, still less one of objects distinct from the possible experience of anyone and everyone but (partly) responsible for these possibilities. (Others think we can show the unintelligibility of talk of truth as correspondence between thought and mind independent fact.)

Inspired by Moore, Russell (1912) influentially claimed that this argument confuses our experiences or conceptions (our mental acts) with their objects or contents. The former are necessarily mental acts, but the latter aren't. Thus we have a basis in experience for an idea of objects that aren't simply possibilities of experience. Other, more radical, empiricists may object that the act/object distinction itself isn't simply given to consciousness but requires explication in empiricist terms. Even so, Russell (1948) argued that *non-experience* must be definable with whatever definition of experience empiricists use to state their own view, logical concepts like "not", "some", "=", and logical ingenuity. The empiricist demand that meaningful talk be definable in experiential terms could then be compatible with intelligible talk of matter as something quite unlike immediate experience and more than just possibilities of experience.

As for logical concepts, empiricists sometimes ground our understanding of them in our experience of mental states, *e.g.* "not" in our frustration. Critics often object that this absurdly turns judgements like "This isn't blue" into judgements about our minds and confuses terms for logical operations we perform on names, predicates and propositions with names or predicates standing for experienced mental states. Other empiricists argue that logical notions are definable in logically

and empirically unproblematic fashion, *e.g.* as truth-tables define logical operators as connectives yielding truth/falsehood from truths/falsehoods of empirically unproblematic meaning.

Similar questions arise with "categories" like substance, causality, person, self. Empiricists often reduce the categories to collections of qualities, events or experiences, ordered only in space and time but otherwise without necessary connection or point of collective reference; or distinguish categorially-related qualities and events from otherwise unrelated ones only in terms of how we are disposed to think, feel, infer or act with respect to them. For instance, Hume notoriously gave two definitions of cause: an event followed by another where events like the former are regularly followed by the latter, and also an event followed by another where the appearance of the former leads one to *expect* the latter. The first leaves out necessary connection between cause and effect; the second leaves in only the feeling of necessity or expectation we may have when we make causal judgements.

Rationalists (*see* RATIONALISM) scorn empiricists for giving woefully inadequate analyses of crucial notions or for confusing psychology with philosophy. They conclude that categories cannot be explicated in experiential terms and their application to experience must be underwritten non-empirically by appeal to general principles not warranted by experience.

Empiricists instead often see Hume as insightfully recognizing the real, intelligible, albeit downgraded, content of the categories. Or, as insightfully recognizing the real, albeit cognitively downgraded, status of categorial judgements as mere dispositions of inference and action. In the latter case, the descriptions we are disposed to give of events, qualities, and their spatio-temporal order are true/false, warranted/unwarranted, but the dispositions themselves aren't. Recently, other empiricists have rejected the reductionism that underwrites downgrading. We learn to ascribe most concepts ("red" as much as "cherry" or "cause") on the basis of distinctive experiential conditions and patterns (*e.g.* looking red) which

are neither necessary nor sufficient for their truth. These core experiences then constitute good but defeasible reasons for these ascriptions, Pollock argues, for example, as a matter of logical necessity or meaning, not empirically confirmed fact.

Empiricists often argue that what makes propositions necessarily true isn't a realm of necessary facts but just the way we think or speak. Many claim that these propositions are reducible upon definition of their terms to tautologies whose denials are self-contradictory (see ANALYTICITY; LOGICAL POSITIVISM). Their form rather than what they say about the world makes them true. They are thus knowable *a priori* rather than through sense experience. However, the necessity of logical truth and the adequacy of definition still need explanation. Broader appeals can be made to our inability to conceive or combine ideas in our minds (see LOCKE; HUME) or to our resolutions or conventions never to assert/deny certain statements or combinations of statements (see HOBBS; AYER; LEWIS; CARNAP). Critics frequently argue that accounting for all and only necessary truths requires appealing to what can't be said without contradiction or to what logically follows from our resolutions/conventions of speech – precisely what needs explaining (Pap, 1958).

Moreover, consider how well we know basic general logical and mathematical principles, sometimes even after considering just a few examples. There is little obvious connection to the quantity and quality of our introspected attempts at conceiving combinations of ideas, or to our introspective knowledge of our psychological, practical, or linguistic proclivities. Nor, for that matter, to the warrant the variety and quantity of perceived instances of such principles might provide through inductive or explanatory inference. Yet suppose we allow that through consideration of particular examples, actual and hypothetical, we arrive by abstraction at a knowledge transcending them in content and the inferential warrant they yield. That seems tantamount to embracing rationalist metaphors of intellectually seeing and grasping the truth – metaphors for which empiricists typically condemn *a priori* intuition and

its objects as hopelessly occult. Some empiricists and philosophers sympathetic to empiricism (e.g. Locke, Russell) seem willing to embrace the metaphors up to a point. Others sometimes argue instead that knowledge of necessities is knowledge of the rules or norms governing thinking and speaking, not knowledge of our social or psychological customs or of the contents of our minds. They need to address more clearly what these distinctions involve (see INTUITION AND DEDUCTION).

Quine, on the other hand, perhaps more radically and more consistently empiricist, denies that there is any principled distinction to be drawn between those sentences that are true by virtue of the rules of language and those that are true by virtue of the facts, or between those we would never reject, no matter what experience may bring, and those we would under certain circumstances. In so doing, he abandons the necessary/contingent, *a priori*/a posteriori distinctions (see NECESSARY/CONTINGENT; A PRIORI/A POSTERIORI) for a broader holism (see HOLISM) according to which the theories or sets of beliefs that get tested by experience include not only physical theories and beliefs about the world, but also the logical principles that allow them to have implications for experience. Any member of the set, including the logical principles, might, at least in principle, be rejected in the face of recalcitrant experience. Critics (e.g. Bonjour, 1998) object that the rationality of abandoning elements of an empirically testable theory because of logical inconsistencies within the theory or between the theory and experience requires logical principles that yield these inconsistencies and that aren't themselves just elements in the empirically testable theory.

From Kant to Wittgenstein (see KANT; WITTGENSTEIN; PRIVATE LANGUAGE ARGUMENT) and Sellars (see SELLARS), empiricists have also been challenged to show how even the most basic knowledge or understanding is possible without allegedly derivative knowledge. Sellars, for example, argues that, for "that's red" or "that's the same in colour" to be correct and express knowledge, more than a reliable disposition to say such things before red objects is needed; governance by

rules and thus general knowledge concerning when such sentences are correctly uttered and thus legitimate is required. That in turn requires more particular and general knowledge, all conforming to norms or standards of thinking whose authority we recognize as interpersonally binding. Empiricism then is deeply incoherent, whatever the ultimate explanation of this authority – whether as principles necessary for any possible empirical consciousness, principles necessary for any effective agency, or just reflections of our agreement about particular cases. Such arguments are the bane of empiricists, partly for their seductiveness, partly for their obscurity, partly for their far-reaching importance.

See also EPICURUS; LOGICAL POSITIVISM; RATIONALISM.

#### BIBLIOGRAPHY

- Ayer, A. J.: *Language, Truth, and Logic*, 2nd edn (London: Gollancz, 1946).  
 Bonjour, L.: *In Defence of Pure Reason* (Cambridge: Cambridge University Press, 1998).  
 Lewis, C. I.: *An Analysis of Knowledge and Valuation* (La Salle, IL: Open Court, 1946).  
 Pap, A.: *Semantics and Necessary Truth* (New Haven, CT: Yale University Press, 1958).  
 Pollock, J.: *Knowledge and Justification* (Princeton, NJ: Princeton University Press, 1974).  
 Quine, W. V.: *Ontological Relativity* (New York: Columbia University Press), chs 3, 5.  
 Russell, B.: *The Problems of Philosophy* (1912) (Oxford: Oxford University Press, 1967).  
 Russell, B.: *Human Knowledge: Its Scope and Limits* (London: Allen and Unwin, 1948).  
 Sellars, W.: *Science, Perception, and Reality* (London: Routledge and Kegan Paul, 1963).  
 Woolhouse, R. S.: *The Empiricists* (Oxford: Oxford University Press, 1988).

BRUCE HUNTER

**Epicurus (c.341–271 BC)** Greek philosopher, founder of a school of philosophy at Athens. Epicurus is arguably the first thinker in antiquity to develop a full-fledged empiricist (see EMPIRICISM) epistemology. Although not the first to claim that knowledge is derived from the senses or that abstract conceptions

arise through the repeated imprint of sensory experience on the memory, Epicurus is the first to have combined these claims with the assertion that sensation alone, as opposed to reason, offers an indubitable foundation for knowledge and a sure defence against scepticism. It is this reliance on sensation to meet sceptical challenges about justification that sets him apart from his predecessors and gives his theory a self-consciously empiricist cast. At the same time, however, fundamental differences in perspective and method separate Epicurus from many later empiricists. He does not take as his point of departure, for instance, the familiar empiricist doctrine that agents have complete and incorrigible access to their own sense-contents; nor does he begin from the perspective of the subject's own awareness and then attempt to justify further inferences to external objects. Rather, he has in mind a wider naturalistic programme rooted in his atomist analysis of perception. He provides detailed naturalistic explanations of the causal effects of various external objects on percipients, with the consequence that the representational and informational character of the resulting sensory states is often wider than what is subjectively accessible to agents themselves.

Epicurus' most distinctive and notorious doctrine is that all sensations are true. Much of his effort on its behalf is spent attacking the following two alternatives: (1) all sensations are false, or (2) some sensations are false. He insists that the scepticism endorsed in (1) is self-refuting since it is impossible to commit oneself consistently to such a view either in theory or in practice. If (2) were true, he argues, a criterion would need to be established for distinguishing true sensations from false; but there can be no such criterion other than sensation itself. Against the objection that different senses may provide conflicting reports, not all of which can therefore be true, Epicurus argues that no sense ever actually contradicts the report of another, since the objects of the various senses are all different. Of course, when faced with incommensurable reports from different senses, it is not clear why we should not suspend judgment rather than conclude that they are all true. To meet

such sceptical objections, Epicurus appeals to his general causal theory of sensation. What guarantees the truth of all sensations, he claims, are the causal processes that bring them about. The senses are reliable in their own spheres, since they merely transmit their information mechanically and passively, adding nothing to what they are recording. As in Locke, there is a strong emphasis both on the passivity of the perceiver and on the mechanisms of representation, though for the Epicurean the process of representation is underwritten not by mental impressions but by the physical images (*eidola*) continually being emitted by external objects. In isolation, such *eidola* are not infallible guides to the nature of external objects *per se*; but when properly assessed, he claims, they lead to true judgments about external reality.

Epicurus' account is vulnerable, no doubt, to many of the objections raised against imagist theories of sensation and thought. But it was instrumental in establishing an overall theoretical framework that was to have enduring influence from Gassendi and Hobbes to J. S. Mill.

See also SEXTUS EMPIRICUS.

#### WRITINGS

Many of the relevant texts are set out and lucidly discussed in A. A. Long and D. N. Sedley, *The Hellenistic Philosophers* (Cambridge: Cambridge University Press, 1987), which includes a detailed bibliography.

PHILLIP MITSIS

**epistemic deontologism** A way of thinking about the concept of epistemic justification in terms of epistemic duty. There are deontological and non-deontological approaches to characterizing epistemic justification. We can distinguish between them by considering the following question:

*The Epistemic Deontologism Question*

When a person considers a proposition, is it ever true that she *epistemically ought* to believe the proposition, disbelieve the proposition, or

suspend judgment on the truth-value of the proposition?

Some philosophers think that the answer to this question is "yes". These philosophers take a deontological approach towards epistemic justification. Philosophers who think the answer is "no" take a non-deontological approach to epistemic justification.

There are many different versions of epistemic deontologism, all of which have in common the claim that doxastic attitudes can be properly evaluated using deontological concepts. For example, some deontologists understand "S is justified in believing *p*" as "S *ought* to believe *p*". Others would analyze it using locutions such as the following: "It is permissible for S to believe *p*", "S is not blameworthy in believing *p*", or "S is not irresponsible in believing *p*".

In modern philosophy, René Descartes and John Locke are important advocates of epistemic deontologism. In his fourth Meditation, Descartes claimed:

But if I abstain from giving my judgment on any thing when I do not perceive it with sufficient clearness and distinctness, it is plain that I act rightly and am not deceived. But if I determine to deny or affirm, I no longer make use as I should of my free will, and if I affirm what is not true, it is evident that I deceive myself; even though I judge according to truth, this comes about only by chance, and I do not escape the blame of misusing my freedom; for the light of nature teaches us that the knowledge of the understanding should always precede the determination of the will.

(1955, p. 176)

Clearly, Descartes thinks that we can believe *rightly* or *wrongly*, that some of our beliefs can be subject to *blame* and *praise*, and that we *ought* to abstain from belief when we do not perceive its truth with sufficient clarity and distinctness. In similar fashion, John Locke writes:

He that believes, without having any reason for believing, may be in love with his own fancies; but neither seeks truth as he ought, nor pays the obedience due his maker, who would have him use those discerning faculties he has given



him, to keep him out of mistake and error. He that does not do this to the best of his power, however, he sometimes lights on truth, is in the right but by chance; and I know not whether the luckiness of the accident will excuse the irregularity of his proceeding. This at least is certain, that he must be accountable for whatever mistakes he runs into: whereas he that makes use of the light and faculties God has given him, and seeks sincerely to discover truth, by those helps and abilities he has, may have this satisfaction in doing his duty as a rational creature, that though he should miss truth, he will not miss the reward for it. For he governs his assent right, and places it as he should, who in any case or matter whatsoever, believes or disbelieves, according as reason directs him.

(1959, pp. 413–14)

Another historical source of epistemic deontology is W. K. Clifford, who claimed boldly:

It is wrong always, everywhere, and for anyone, to believe anything upon insufficient evidence.

(1879, p. 183)

Twentieth-century deontologists include A. J. Ayer, Roderick Chisholm, Carl Ginet, and Richard Feldman. According to Ayer (1955), a necessary condition of knowledge is the “right to be sure”. Similarly, Chisholm proposed:

To know that *h* is true will be not only to have true opinion with respect to *h*, but to have a certain right or duty with respect to *h*.

(1966, p. 11)

Ginet holds that one

is justified in being confident that *p* iff it is not the case that one ought not to be confident that *p*; one could not be justly reproached for being confident that *p*.

(1975, p. 28)

Finally, Feldman contends:

Whenever one’s evidence on balance supports a proposition, then one has an *epistemic right* to believe that proposition. Believing it is *permitted*. Believing anything for which one lacks good reasons is *epistemically prohibited*.

(2001, pp. 88–9)

Although the term “deontology” is used in both ethics and epistemology, the distinction between deontological and non-deontological views in epistemology is quite different from the distinction between deontological and non-deontological views in ethics. Ethical systems that are based on Kant’s Categorical Imperative or the Golden Rule are examples of deontological theories. They qualify as deontological because they assert that whether an action is right or wrong is determined, not by its consequences, but rather by its being or not being in agreement with the relevant rules of morality. It is characteristic of such theories that, according to them, an action can be morally right even if it has bad consequences. This is precisely what non-deontological theories, such as act-utilitarianism, deny. Thus, in ethics, the debate over whether deontology is the correct approach concerns the moral relevance of an action’s consequences. The debate over deontology in ethics is therefore significantly different from the debate in epistemology, which concerns the legitimacy of using evaluative terms such as “ought”, “should”, and “permissible” when judging the epistemic status of a belief. This debate has no strict analog in ethics. Advocates of both deontological and non-deontological views in epistemology could agree that actions can be properly evaluated in deontological terms.

William Alston, although he is one of the main critics of epistemic deontology, interestingly concedes:

The terms “justified”, “justification”, and their cognates are most naturally understood in what we may term a “deontological” way, as having to do with obligation, permission, requirement, blame, and the like. (Alston, 1989, p. 115)

Alvin Plantinga, like Alston an opponent of epistemic deontology, makes a similar concession:

Indeed the whole notion of epistemic justification has its origin and home in this deontological territory of duty and permission, and it is only by way of analogical extension that the term “epistemic justification” is applied in other

ways. Originally and at bottom, epistemic justification is deontological justification: deontological justification with respect to the regulation of belief. (Plantinga, 1993, p. 14)

But, if these critics of epistemic deontology concede that much, on what grounds do they oppose epistemic deontology?

There are two main objections to epistemic deontology. The first of these, due to Alston (1989, essays 4 and 5), is based on the claim that our doxastic attitudes are not under our direct, voluntary control. This objection can be stated as follows:

#### The Involuntarism Argument

- (1) If it is correct to evaluate doxastic attitudes deontologically, then it is possible to exert voluntary control over our doxastic attitudes.

It is not possible to exert voluntary control over doxastic attitudes. Therefore:

- (2) It is not correct to evaluate doxastic attitudes deontologically.

Like Alston, Plantinga (1993) and Williams (1975) consider this argument an effective refutation of epistemic deontology. The first premise seems to follow from the widely endorsed ought-implies-can principle. This principle holds that one can be obligated to do A only if it is within one's power to do A. Opponents of epistemic deontology motivate the second premise citing multiple examples of what seems obvious: it is not within our power to choose our doxastic attitudes. The doxastic attitudes we take on seem to be automatic and spontaneous responses beyond our voluntary control.

The Involuntarism Argument is controversial. Feldman (2001), Ryan (2003), and Wolterstorff (1997) have challenged premise (1). Feldman and Wolterstorff argue that epistemic "oughts" function differently from ethical "oughts", which is why we can have epistemic obligations that we cannot fulfill while we cannot have ethical obligations of that kind. Ryan argues that the "ought"-implies-"can" principle is false for both ethical

and epistemic "oughts". Thus, she argues, even if doxastic attitudes are not under our voluntary control, it can still make sense to view epistemic justification deontologically.

Ginet (2001), Steup (2000, 2007), and Ryan (2003) have argued against premise (2). Ginet claims that in situations when our evidence is not conclusive we can decide which doxastic attitude to adopt. Both Steup and Ryan argue that our doxastic attitudes are as much under our control as actions that we would ordinarily regard as free.

The second major criticism of epistemic deontology is that deontological justification is not truth-conducive. According to this objection, while epistemic justification is supposed to make our beliefs likely to be true, a belief can be deontologically justified without being formed in a way that makes that belief likely to be true. Alston (1989, essay 4) proposes the following example to demonstrate this possibility:

I may have done what could reasonably be expected of me in the management and cultivation of my doxastic life and still hold a belief on outrageously inadequate grounds. There are several possible sources of such a discrepancy. First, there is what we might call "cultural isolation". If I have grown up in an isolated community in which everyone unhesitatingly accepts the traditions of the tribe as authoritative, then if I have never encountered anything that seems to cast doubt on the traditions and have never thought to question them, I can hardly be blamed for taking them as authoritative. There is nothing I could reasonably be expected to do that would render me more exposed to counterevidence. (We can suppose that the traditions all have to do with events distant in time and/or space, matters on which I could not be expected to gather evidence on my own.) I am [deontologically justified] in believing these things. And yet the fact that it is the tradition of the tribe [to believe] that p may be a very poor reason for believing that p.

(1989, p. 95)

The gist of Alston's argument is as follows: The members of the isolated community form beliefs on the basis of their traditions. Because of their isolated condition, they cannot be blamed for their beliefs. Consequently,

deontology implies that their beliefs are justified. However, we intuitively judge that their beliefs would actually be unjustified if the reasons supplied by their traditions are poor.

Many philosophers reject epistemic deontology on the basis of precisely this kind of argument (Pryor, 2001, pp. 114–15; and BonJour, 2003, pp. 175–7). However, there are several ways in which deontologists can respond to the truth-conduciveness objection.

First, deontologists might insist that meeting one's epistemic duty involves more than mere blamelessness. It might, for example, require meeting the evidentialist's standard of believing *p* if and only if believing *p* fits one's evidence. So, if the isolated subjects believe in a way that doesn't fit their evidence, their beliefs would not be deontologically justified even if their beliefs are held blamelessly (Conee and Feldman, 2001).

Second, deontologists might argue that one is epistemically blameless in believing *p* if and only if believing *p* fits one's evidence. If blamelessness is thus construed, Alston's culturally isolated subjects are not blameless in believing in the way they do (Steup, 1988).

Third, deontologists could point out that the details of Alston's objection are not sufficiently specified. Alston never makes clear exactly what his subjects believe and what their evidence amounts to. Suppose a more complete description of the example tells us that Alston's isolated subjects have excellent evidence for believing something that is either false or (in an objective, externalist sense of probability) unlikely to be true. Deontologists will then argue that Alston's intuition – the beliefs in question are actually unjustified – rests on an externalist conception of justification that they reject. Alternatively, suppose we are to understand that Alston's isolated subjects have epistemically defective reasons for believing something that is widely accepted in their cultural group. In that case, deontologists will point out that Alston uncharitably saddles deontology with a relativist notion of justification, to be contrasted with an objectivist notion, according to which one ought to believe what one's evidence actually supports rather than what one happens to think one's evidence supports. Deontologists have no reason to

reject such objective standards of epistemic obligation. Hence, given this alternative construal of Alston's example, they would not concede the premise that, according to deontology, the beliefs in question are justified.

To conclude, epistemic deontology is a highly controversial view. While it is true that there are serious objections to it, it also seems that deontologists have the resources to rebut these objections.

#### BIBLIOGRAPHY

- Ayer, A. J.: *The Problem of Knowledge* (New York: St Martin's Press, 1955).
- Alston, W.: *Epistemic Justification. Essays in the Theory of Knowledge* (Ithaca, NY: Cornell University Press, 1989).
- BonJour, L. and Sosa, E.: *Epistemic Justification* (Malden, MA: Blackwell Publishing, 2003).
- Chisholm, R.: *Theory of Knowledge* (Englewood Cliffs, NJ: Prentice-Hall, 1966).
- Clifford, W. K.: "The Ethics of Belief," in *Lectures and Essays*, vol. 2 (London: Macmillan, 1879), 163–205.
- Conee, E. and Feldman, R.: "Internalism Defended," in *Epistemology: Internalism and Externalism*, ed. Hilary Kornblith (Oxford: Blackwell, 2001), 231–60.
- Descartes, R.: *Meditations on First Philosophy, The Philosophical Works of Descartes*, ed. E. S. Haldane and G. R. T. Ross, vol. 1 (New York: Dover, 1955).
- Feldman, R.: "Voluntary Belief and Epistemic Evaluation," in *Knowledge, Truth, and Duty*, ed. M. Steup (Oxford: Oxford University Press, 2001), 77–92.
- Ginet, C.: *Knowledge, Perception, and Memory* (Dordrecht: Reidel, 1975).
- Ginet, C.: "Deciding to Believe," in *Knowledge, Truth, and Duty*, ed. M. Steup (Oxford: Oxford University Press, 2001), 63–76.
- Locke, J.: *An Essay Concerning Human Understanding*, ed. A. C. Fraser (New York: Dover, 1959).
- Plantinga, A.: *Warrant: The Current Debate* (Oxford: Oxford University Press, 1993).
- Pryor, J.: "Highlights of Recent Epistemology," *British Journal for the Philosophy of Science* (2001), 95–124.
- Ryan, S.: "Doxastic Compatibilism and the Ethics of Belief," *Philosophical Studies* 114 (2003), 47–79.
- Steup, M.: "The Deontic Conception of Epistemic Justification," *Philosophical Studies* 53 (1988), 65–84.

- Steup, M.: "Doxastic Voluntarism and Epistemic Deontology," *Acta Analytica* 15 (2000), 25–56.
- Steup, M.: "Doxastic Freedom," *Synthese* 2007.
- Williams, B.: *Problems of the Self* (New York: Cambridge University Press, 1973).
- Wolterstorff, N.: "Obligations of Belief: Two Concepts," in *The Philosophy of Roderick Chisholm*, ed. Lewis E. Hahn (La Salle, IL: Open Court, 1997), 217–38.

SHARON RYAN

**epistemic luck** Fortuitousness in possessing a true belief, a condition traditionally thought to be incompatible with knowledge. Generically, the notion of epistemic luck is used to describe any of a wide variety of ways in which it is accidental, coincidental or fortuitous that a person has a true belief that *p*.

#### THE INCOMPATIBILITY THESIS

A venerable tradition, dating back to Plato, has taken epistemic luck *simpliciter* to be incompatible with knowledge. Epistemologists have long held that knowledge requires more than mere true belief, because a person can fortuitously arrive at a true belief by guessing, and lucky guesses fall short of knowledge. Although the role of the justification condition in the traditional justified-true-belief [JTB] analysis of knowledge is to rule out lucky guesses as instances of knowledge, it's possible, given any fallibilistic account of justification, to have a justified belief that is only luckily true, as Edmund Gettier (1963) demonstrates with the following example: Smith justifiably believes that Jones owns a Ford [*J*] based on good, but misleading, evidence. On the basis of her justified-but-false belief that *J*, Smith justifiably infers that *J or B* (where *B* = Brown is in Barcelona), despite having no idea of Brown's whereabouts. Coincidentally, Brown is in Barcelona. So, it's just a matter of luck that Smith's justified belief that *J or B* is true. Since Smith's *luck-infused* justified-true-belief that *J or B* clearly falls short of knowledge, fallibilistic justification is incapable of eliminating all forms of knowledge-destroying luck.

#### THE INESCAPABILITY OF EPISTEMIC LUCK

An all-pervasive phenomenon, epistemic luck infects every epistemology in one form or other. Its inescapability can be demonstrated as follows: To convert true belief to knowledge, every viable epistemology requires satisfying either some internalistic justification condition or some externalistic condition (that may or may not be a justification condition). But neither an internalistic nor an externalistic condition can completely succeed in eliminating epistemic luck. As the new evil demon problem demonstrates, internalistic justification is not conceptually connected to truth in any robust way, for demon-world victims have internalistically justified beliefs almost all of which are false. Given the absence of a robust truth connection, it's always in some sense a matter of luck when a *merely* internalistically justified belief turns out to be true. Truth-connected externalist approaches (e.g. reliabilist, truth-tracking, and safety-based accounts) avoid this kind of epistemic luck. However, they are subject to another kind of ineliminable epistemic luck. Consider, e.g., the externalist condition of being a reliably produced belief. While a reliably produced belief's being true is not a matter of luck, one's cognitive faculties being reliable is a matter of luck, for in a demon world these same cognitive faculties are highly unreliable. Or consider the externalist condition of being a *safe* belief (to be explained below). While a safe belief's being true isn't lucky, having safe beliefs is, for in a demon world none of one's beliefs are safe. Analogous considerations can be applied to any externalist condition of knowledge. Since every epistemology incorporates either an internalistic justification condition or an externalistic condition, no epistemology can rid us of luck's intractable presence.

#### THE SKEPTICAL CHALLENGE

Epistemic luck, then, is unavoidable. If all forms of epistemic luck are incompatible with

knowledge, as the incompatibility thesis maintains, skepticism is correct. The task facing the anti-skeptical epistemologist is to reconcile the rather strong intuition that epistemic luck is not compatible with knowledge with the equally evident assumption that it must be.

#### REJECTING THE INCOMPATIBILITY THESIS

Peter Unger (1968), the first epistemologist to note that not all forms of epistemic luck are incompatible with knowledge, identifies the following three types of benign epistemic luck: (1) *Propositional luck*: It can be entirely accidental that  $p$ , and  $S$  can still know that  $p$ . For example, a person who witnesses an automobile accident can certainly know that the accident occurred. (2) *Existential luck*: For  $S$  to know that  $p$ ,  $S$  must exist, and it might be extraordinarily lucky that  $S$  exists. If  $S$  is the lone survivor of a fiery plane crash,  $S$  is lucky to be alive, but  $S$ 's existential luck does not preclude her from knowing that she survived the crash. (3) *Facultative luck*: To know that  $p$ ,  $S$  must possess the cognitive skills requisite for knowledge. Suppose  $S$  is shot in the head, but the bullet narrowly misses all vital regions of the brain required for conceptual thought and knowledge.  $S$  is overwhelmingly lucky that she still possesses the cognitive capacities needed for knowledge, but since she does possess them she is still capable of knowing many things, including that she was shot in the head. Unger has successfully identified three types of harmless epistemic luck, but not all forms of epistemic luck are benign. What's needed is an account of knowledge-undermining luck.

#### KNOWLEDGE-DESTROYING EPISTEMIC LUCK

Mylan Engel Jr (1992) distinguishes two kinds of epistemic luck – evidential luck and veritic luck – and argues that only the latter is incompatible with knowledge. Engel characterizes these two types of luck as follows:

EL: A person  $S$  is *evidentially lucky* in believing that  $p$  in circumstances  $C$  iff it is just a matter of luck that  $S$  has the evidence  $e$  for  $p$  that she does; but, given  $e$ , it is not a matter of luck that her belief that  $p$  is true in  $C$ .

VL: A person  $S$  is *veritically lucky* in believing that  $p$  in circumstances  $C$  iff, given  $S$ 's evidence for  $p$ , it is just a matter of luck that  $S$ 's belief that  $p$  is true in  $C$ .

To see that evidential luck is compatible with knowledge, suppose that a bank robber's mask slips momentarily during a holdup, and the startled teller sees clearly that the robber is the bank president. In such a situation, the teller would clearly be lucky to have the evidence she does, but she would nevertheless know that the bank president is the villain.

Engel argues that all genuine Gettier cases involve veritic luck. In Gettier's original example presented above, Smith's belief that  $J$  or  $B$  is veritically lucky: Given Smith's misleading evidence of Jones's Ford-ownership status and her total lack of evidence concerning Brown's whereabouts, it's just a matter of luck that Smith's belief that  $J$  or  $B$  is true. Veritic luck with respect to  $p$  is incompatible with knowing that  $p$ , because it undercuts the connection between  $S$ 's evidence for  $p$  and  $p$  in a way that makes it entirely coincidental that  $p$  is true.

#### NEO-MOOREAN EPISTEMOLOGY

Duncan Pritchard (2003) replaces Engel's evidence-based characterization of knowledge-destroying luck with the following modal analysis:

MVL: For all agents  $S$  and propositions  $p$ , the truth of  $S$ 's belief that  $p$  is veritically lucky iff  $S$ 's belief that  $p$  is true in the actual world  $\alpha$  but false in nearly all nearby possible worlds in which  $S$  forms the belief in the same manner as in  $\alpha$ . (2003: 111)



MVL differs from Engel's characterization in the following way: it concerns the connection between the *method of belief formation* and proposition believed, rather than the connection between *S's evidence* and the proposition for which it is evidence. Pritchard argues that a safety-based neo-Moorean account, according to which knowledge is safe true belief, is capable of eliminating veritic luck. Pritchard's neo-Moorean account has a consequence that many will find unacceptable: It entails *S* knows that her lottery ticket will lose (provided she believes it will and it does), because her belief that her ticket loses is safe. Pritchard tries to circumvent this objection by noting that *S*'s ticket wins in "a small cluster" of nearby worlds. However, since its winning in a small cluster of nearby worlds is compatible with its losing in nearly all nearby worlds, Pritchard would need to replace his weak safety constraint with strong safety to prevent *S* from knowing that her ticket will lose.

#### SENSITIVITY

The post-Gettier literature is rife with attempts at supplementing or amending the traditional JTB analysis with a satisfactory anti-luck constraint on knowledge. The first wave of proposals includes adding a no-essential-false-assumptions condition to JTB, replacing JTB with a causal theory of knowing, and supplementing JTB with a defeasibility condition. These and similar proposals fall prey to ever more complicated Gettier-style examples. The general consensus is that none of these proposals succeeds. First-wave luck-eliminating proposals invoke counterfactual or subjunctive constraints on knowing, principal among them: *sensitivity* and *safety*. *S*'s belief that *p* is *sensitive* to *p*'s truth-value iff *S* would not believe that *p* if *p* were false (i.e. iff *S* does not believe *p* in the closest  $\sim p$ -worlds). To be sure, sensitive belief does preclude veritic luck, but it does so at a steep price. First, the sensitivity condition results in closure failure. (See SKEPTICISM AND CLOSURE in Part I; ZEBRAS AND CLEVERLY DISGUISED MULES.) Second, as Jonathan Vogel (1999) shows with

*Hole-In-One*, knowledge does not require sensitivity. The fourth hole at Augusta National Golf Course where The Masters is played is a tricky 240-yard par 3, euphemistically called "Flowering Crabapple". In 2007, not one player shot a hole-in-one on this diabolical hole, and there were only eleven birdies throughout four rounds of play. Right now, I know that not all seventy-two players in this year's Masters will shoot a hole-in-one on Flowering Crabapple in the first round of play, but my belief to this effect is not sensitive. Were every golfer to shoot a hole-in-one on Flowering Crabapple in Round One of the Masters in defiance of the astronomical odds against it, I would still believe that they weren't going to do so. So sensitivity is not necessary for knowledge.

#### SAFETY

Considerations such as these have led a number of epistemologists (Sosa, 1999 and 2000; Williamson, 2000a and 2000b; Pritchard, 2005) to replace the sensitivity condition with some sort of safety condition. Safety comes in different strengths: *S*'s true belief that *p* is *strongly safe* iff were *S* to believe that *p*, *p* would be true, i.e. in all the closest *Bp* worlds, *p* is true. *S*'s true belief that *p* is *weakly safe* iff *S* would not easily be mistaken with respect to *p*, i.e. *p* is true in the overwhelming majority of close *Bp* worlds.

Peter Murphy (2005) uses Kripke's famous Nozick-bashing counter-example to show that strong safety results in closure failure. Suppose the following is true of Barn County: The landscape is peppered with barn façades, there are a few real barns in the county, some of the real barns are red and some are blue, but all the façades are red. Driving through Barn County, Smith is unaware that most of the barn-looking structures are façades. She looks at a blue barn and comes to believe that she is looking at a blue barn. Her belief is safe. In all nearby worlds where she believes she is looking at a *blue barn*, she is looking at a blue barn. However, her belief that she is looking at a *barn* is not safe. There are many nearby worlds where she believes

she's looking at a barn, but is really just looking at a façade. So strong safety entails that Smith knows she's looking at a *blue barn*, but doesn't know she's looking at a *barn*.

Weak safety is open to a different worry. If knowledge only requires weakly safety-justified true belief, then a person who justifiably believes her lottery ticket will lose *knows* that her ticket will lose (unless, of course, it happens to win). Most epistemologists (though not all) insist that people don't know their lottery tickets will lose, prior to hearing the announced results. Anyone convinced that people don't know their tickets will lose, before learning of the results, will think weak safety too weak an anti-luck constraint for knowledge.

Avram Hiller and Ram Neta (2007) argue that no safety condition can eliminate all cases of veritic luck as follows: Start with a justified-but-false belief like Smith's belief that Jones owns a Ford. Next, have Smith justifiably infer a disjunction of the form *J* or *B*, where, unbeknownst to Smith, *B* is true in all nearby worlds. Let *B* = Brown will not win a Grammy. Suppose, unbeknownst to Smith, that Brown is totally devoid of musical talent and there is no remotely close world where Brown wins a Grammy. Then Smith's true belief that *J* or *B* will be safe, but veritically lucky none the less.

#### INELIMINABLE REFLECTIVE LUCK

As noted, Pritchard maintains that the threat posed by veritic luck can be eliminated by adopting a safety-based neo-Moorean account of knowledge. He thinks, however, that there is another even more worrisome kind of luck – reflective epistemic luck – that cannot be so eliminated. Such luck arises when *from the agent's reflective position* it's just a matter of luck that her belief is true. More precisely:

MRL: For all *S* and *p*, the truth of *S*'s belief that *p* is *reflectively lucky S's* belief that *p* is true in  $\alpha$  but, in nearly all nearby possible worlds consistent with what *S* is able to know by

reflection alone, were *S* to believe *p*, *p* would be false (2003: 122).

Pritchard thinks that reflective luck is not with ordinary knowledge (only veritic luck is), but he insists that reflective luck is incompatible with a much-desired internalistic kind of robust reflective knowledge. He contends that skeptical challenges force us to confront the human epistemic predicament. We cannot know, on the basis of reflection alone, that the skeptic's radical hypotheses are false. We cannot know, by reflection alone, that we are not BIVs. We all believe that the skeptic's hypotheses are false, but we lack adequate reflectively accessible internalistic grounds for thinking so. Skeptical challenges thus force us to recognize the reflectively lucky nature of our anti-skeptical beliefs, which, in turn, explains the enduring epistemic *Angst* that skeptical hypotheses engender.

See also CONTEXTUALISM in Part I; EPISTEMIC VIRTUE; GETTIER PROBLEM; NOZICK; SENSITIVITY AND SAFETY.

#### BIBLIOGRAPHY

- Engel Jr., M.: "Is Epistemic Luck Compatible with Knowledge?," *Southern Journal of Philosophy* 30 (1992), 59–75.
- Gettier, E.: "Is Justified True Belief Knowledge?," *Analysis* 23 (1963), 121–3.
- Heller, M.: "The Proper Role for Contextualism in an Anti-luck Epistemology," *Philosophical Perspectives*, 13, *Epistemology* (1999), 115–29.
- Hiller, A. and Neta, R.: "Safety and Epistemic Luck," *Synthese* 158 (2007), 303–13.
- Murphy, P.: "Closure Failure for Safety," *Philosophia* 33 (2005), 331–4.
- Nozick, R.: "Knowledge and Skepticism," in *Philosophical Explanations* (Cambridge, MA: Harvard University Press, 1981).
- Pritchard, D.: *Epistemic Luck* (Oxford: Oxford University Press, 2005).
- Pritchard, D.: "Virtue Epistemology and Epistemic Luck," *Metaphilosophy* 34 (2003), 106–30.
- Sosa, E.: "Skepticism and Contextualism," *Philosophical Issues*, 10, *Skepticism* (2000), 1–18.
- Sosa, E.: "How to Defeat Opposition to Moore," *Philosophical Perspectives*, 13, *Epistemology* (1999), 141–54.

- Unger, P.: "An Analysis of Factual Knowledge," *Journal of Philosophy* 65 (1968), 157–70.
- Vogel, J.: "The New Relevant Alternatives Theory," *Philosophical Perspectives*, 13, *Epistemology* (1999), 155–80.
- Williamson, T.: "Scepticism and Evidence," *Philosophy and Phenomenal Research* 60 (2000a), 613–28.
- Williamson, T.: *Knowledge and Its Limits* (Oxford: Oxford University Press, 2000b).

MYLAN ENGEL

**epistemic supervenience** The thesis that, necessarily, whenever something has an epistemic property, E, it has a collection of non-epistemic properties, possession of which entails that it has E. Epistemic properties include *justification*, *rationality*, *reasonableness*, *warrant*, and *knowledge*. Non-epistemic properties include *truth*, *falsity*, *reliably produced*, and *psychologically certain*. Epistemologists tend to agree that an acceptable epistemological theory must respect the supervenience of epistemic on non-epistemic properties.

The supervenience thesis can be understood *ontologically*, in which case it pertains to *properties* (or *truths* or *facts*), or *linguistically*, in which case it applies to *ascriptions*. Those inclined to anti-realism about epistemic properties will favor the latter (Klagge, 1988). Epistemologists, generally disinclined to anti-realism about epistemic properties, focus on the former.

Supervenience is a relation between classes of properties (or truths, or facts). The A-properties supervene on the B-properties (the "subvenient" or "base" properties) just in case no two things can differ in their A-properties without also differing in some of their B-properties. In short:

- (S) There cannot be an A-difference without a B-difference.

Supervenience theses vary with respect to what they quantify over. *Individual* supervenience theses quantify over individuals, *regional* over regions of space-time, and *global* over entire possible worlds. Alternative versions of the thesis also result from different

interpretations of the modal force of "cannot" (nomological, metaphysical, analytic, or logical). A further issue concerns the possible distribution patterns of A- and B-properties. *Weak* supervenience theses restrict patterns only within respective possible worlds (intra-world restriction), whereas *strong* ones restrict them across possible worlds (inter-world restriction). In epistemology, supervenience theses are typically, if not always, individual, strong, and stated in terms of metaphysical necessity.

The supervenience of the A-properties on the B-properties *does not* guarantee that the A-properties "depend on" or otherwise obtain "because of" or "in virtue of" the B-properties. Further argument is needed to establish these more robust metaphysical claims (Grimes, 1988; Kim, 2002, pp. xviii–xxii). For example, the singleton set whose only member is the property of being self-identical supervenes on the set of color properties, but things are not self-identical because of their color properties.

The thesis of epistemic supervenience asserts the following:

- (ES) Necessarily, whenever something has an epistemic property, E, it has a collection of non-epistemic properties, N, possession of which entails that it has E (Van Cleve, 1985, pp. 97–8; Sosa, 1991, pp. 156, 183).

While ES covers all epistemic properties of all possible doxastic attitudes and cognizers, for simplicity we focus on the justification of beliefs. The relevant modality is metaphysical. N might include complex relational properties. Technically, ES is neutral on whether there are any epistemic properties.

Statements of the thesis vary, most conspicuously in that sometimes, instead of "non-epistemic" properties, E is said to supervene on "non-normative" or "non-evaluative" or "natural" or "descriptive" properties (compare Goldman, 1979, p. 340; Chisholm, 1989, pp. 42 f; and Sosa, 1991, p. 153). Sometimes an author will use two or more of these interchangeably (e.g. Steup, 1996, pp. 30–43; Lehrer, 1997, pp. 64–5), but the differences among them are potentially

significant. Consider divine command epistemology, according to which it is a necessary truth that a belief is justified if and only if God commands the subject to hold it. Having been commanded by God is neither an epistemic nor an evaluative property, so divine command epistemology entails ES, as well as the version of ES with “non-evaluative” substituted for “non-epistemic”. But it is not a *natural* property in any recognizable sense of “natural”, so divine command epistemology does not entail the version of ES with “natural” substituted for “non-epistemic”.

To deny ES is to claim that at least some epistemic properties are *autonomous* (Sosa, 1991, pp. 153–4) or *independent* of the non-epistemic (Lehrer, 1997, p. 64). One could consistently deny the possibility of a satisfactory *analysis* of any epistemic concept in completely non-epistemic terms, yet at the same time endorse ES. Ernest Sosa (1991, pp. 153–4) calls this combination of views *pessimism*.

Proponents of ES need not deny that epistemic properties also supervene on epistemic properties. Indeed, it would be foolish to deny this further claim since every set of properties trivially supervenes on itself.

A standard argument in favor of ES proceeds from the observation that the evaluative in general supervenes on the non-evaluative (Alston, 1976, p. 170; Kim, 1988, p. 310; Sosa, 1991, pp. 152, 179, 192):

1. All evaluative properties supervene on non-evaluative properties.
2. All epistemic properties are evaluative properties.
3. Therefore, all epistemic properties supervene on non-evaluative properties.

The conclusion of this argument is stronger than the official statement of ES; it entails ES, but ES does not in turn entail it. Unlike the conclusion of the present argument, ES would not be falsified if it turned out that epistemic properties supervene on *non-epistemically* evaluative properties (e.g. moral properties) that do not in turn supervene on non-evaluative ones.

James Van Cleve (1985, pp. 98–9) offers the following argument in favor of epistemic

supervenience (substituting “non-epistemic” for “natural”).

1. Either the epistemic supervenes on the non-epistemic, or there could be a justified belief with no non-epistemic properties, or there could be a justified belief with some non-epistemic properties but none that entails its epistemic properties.
2. There could not be a justified belief with no non-epistemic properties.
3. There could not be a justified belief with some non-epistemic properties but none that entails its epistemic properties.
4. Therefore, the epistemic supervenes on the non-epistemic.

Premise 2 is true because any belief will have some non-epistemic property, e.g. temporal properties. Denying premise 3 would be “absurd” because that would imply “that there could be another belief just like [a paradigmatically justified belief] in all [non-epistemic] respects – directed at the same proposition, caused by similar causes, accompanied by similar experiences, related in the same ways to other beliefs of its subject, and so on – yet not justified”. Yet “a difference in epistemic status must surely be traceable to some *further* [non-epistemic] difference”. It is unclear how persuasive this defense of premise 3 would be to someone who didn’t already share the intuition that ES is true.

Keith Lehrer is the chief opponent of epistemic supervenience and has offered several arguments against the view. At times Lehrer suggests (1997, p. 70; 1999, p. 1071) that ES fails because we cannot give an account of epistemic properties “without the use of epistemic terms”. But this confuses pessimism with the autonomy of epistemic properties (Sosa, 2003, pp. 27f). However, Lehrer’s arguments discussed hereafter do not trade on this confusion.

One of Lehrer’s arguments proceeds from the supposed failure of alethic supervenience (Lehrer, 1997, pp. 68 f, 73–5), along with an element of Lehrer’s positive theory of justification, to wit, that (undefeated) justification requires that the subject have true beliefs about what she is trustworthy in

accepting. Justification won't supervene on natural properties unless truth does, then, because a *necessary* condition for justification is that the subject have *true* beliefs about her trustworthiness. However, this fails to acknowledge the possibility that the set of truths about what subjects are trustworthy in accepting could supervene on the subjects' natural properties even if truths *in general* do not. But Lehrer *does* deny that truths about trustworthiness supervene on natural or even non-epistemic properties: "There is nothing about me, short of my being trustworthy in what I accept, that necessitates that I am trustworthy in what I accept" (1997, p. 72). Since trustworthiness itself is said to be an epistemic property, this would falsify ES. In response, Van Cleve (1999, pp. 1054 f) remarks that if Lehrer is correct, then it is possible for a trustworthy person to "cease being trustworthy" even though nothing about her or her environment changed in non-epistemic respects. According to Van Cleve, however, it is extremely counter-intuitive that trustworthiness could "float free in that way . . . fluctuating though nothing else fluctuates with it . . .".

Another of Lehrer's arguments (1999, p. 1070; 2003, p. 318) begins with the claim that metaphysical necessity and possibility "track consistency": "If the claim that *x* exemplifies *F* and not *G* is consistent, then it is possible that *x* exemplifies *F* and not *G*, and it is not necessarily true that if *x* exemplifies *F*, then *x* exemplifies *G*" (2003, p. 318). Let "*F*" predicate any non-epistemic property, or combination thereof, that the supervenience theorist proposes, and "*G*" the property *justifiedly believes that he has hands*. The claim that *S* is *F* and not *G* is consistent. So it is metaphysically possible that *S* is *F* and not *G*. Therefore, ES is false. For the argument to succeed, we need a suitable characterization of *consistency*. If consistency just means logical consistency, then consistency does not imply metaphysical possibility: it is logically consistent to say that water is a chemical element, or that I am petting an orca but not a mammal, but neither of these things is metaphysically possible. If consistency means

metaphysical possibility, then the argument begs the question.

Lehrer (2003, p. 320) also offers this argument:

1. If epistemic properties supervene on natural properties, then epistemic properties are epiphenomenal.
2. Epistemic properties are not epiphenomenal.
3. Therefore, epistemic properties do not supervene on natural properties.

To say epistemic properties are epiphenomenal is to say that they are causally inefficacious, mere "causal dangles". A similar argument, which Lehrer also embraces, applies to moral properties. Premise 1 is the weak link. From the fact that a property supervenes on natural properties, we cannot conclude that it is epiphenomenal. Paradigmatically causally efficacious properties, such as *having a force of 150 newtons*, supervene on natural properties, since they *are* natural properties.

Earl Conee and Richard Feldman advocate "mentalism", the view that justification "strongly supervenes" on the subject's total mental condition, which entails that "if any two possible individuals are exactly alike mentally, then they are alike justificationaly, e.g., the same beliefs are justified for them to the same extent" (2001, p. 56). They argue for mentalism on the grounds that it best explains our intuitions about a wide range of particular cases. They also maintain that mentalism is the only or best way to defend a broadly internalist perspective on justification, given that standard arguments based on a deontological conception of justification have been undermined.

Mentalism pertains solely to justification. It does not state, for instance, that knowledge or warrant supervenes on the mental. ES could be true even if mentalism were false, and vice versa. Epistemic properties might supervene on non-epistemic properties, e.g. contingent reliability relations, despite failing to supervene on the mental, in which case ES would be true and mentalism false. Mentalism would be true and ES false if some of justification's



subvenient mental properties were epistemic properties that did not in turn supervene on non-epistemic properties. And, even if knowledge failed to supervene on non-epistemic properties, justification might still supervene on non-epistemic mental properties, in which case mentalism would be true and ES false.

## BIBLIOGRAPHY

- Alston, W. P.: "Two Types of Foundationalism," *Journal of Philosophy* 73 (1976), 165–85.
- Chisholm, R.: *Theory of Knowledge*, 3rd edn (Englewood Cliffs, NJ: Prentice Hall, 1989).
- Conee, E. and Feldman, R.: "Internalism Defended," *American Philosophical Quarterly* 38 (2001), 1–18. Reprinted in Conee and Feldman, 2004.
- Conee, E. and Feldman, R.: *Evidentialism: Essays in Epistemology* (Oxford: Oxford University Press, 2004).
- Goldman, A. I.: "What Is Justified Belief?," in *Justification and Knowledge*, ed. G. S. Pappas (Dordrecht: Reidel, 1979), 1–23. Reprinted in Sosa and Kim, 2000.
- Grimes, T. R.: "The Myth of Supervenience," *Pacific Philosophical Quarterly* 69 (1988), 152–60. Reprinted in Kim, 2000.
- Kim, J.: "What Is 'Naturalized Epistemology'?", *Philosophical Perspectives*, 2, *Epistemology* (1988). Reprinted in Sosa and Kim, 2000.
- Kim, J., ed.: *Supervenience* (Burlington, VT: Ashgate, 2000).
- Klagge, J. C.: "Supervenience: Ontological and Ascriptive," *Australasian Journal of Philosophy* 66 (1988), 461–70. Reprinted in Kim, 2000.
- Lehrer, K.: *Self Trust: A Study of Reason, Knowledge, and Autonomy* (Oxford: Clarendon Press, 1997).
- Lehrer, K.: "Reply to Van Cleve," *Philosophy and Phenomenological Research* 59 (1999), 1068–71.
- Lehrer, K.: "Reply to Sosa," in *The Epistemology of Keith Lehrer*, ed. E. J. Olsson (Dordrecht: Kluwer, 2003), 317–20.
- Sosa, E.: *Knowledge in Perspective: Selected Essays in Epistemology* (Cambridge: Cambridge University Press, 1991).
- Sosa, E.: "Epistemology: Does It Depend on Independence?," in *The Epistemology of Keith Lehrer*, ed. E. J. Olsson (Dordrecht: Kluwer, 2003), 23–30.
- Sosa, E. and Kim, J. (eds): *Epistemology: An Anthology* (Malden, MA: Blackwell, 2000).
- Steup, M.: *An Introduction to Contemporary Epistemology* (Upper Saddle River, NJ: Prentice Hall, 1996).

Van Cleve, J.: "Foundationalism, Epistemic Principles, and the Cartesian Circle," *The Philosophical Review* 88 (1979), 55–91. Reprinted in Sosa and Kim, 2000.

Van Cleve, J.: "Epistemic Supervenience and the Circle of Belief," *The Monist* 68 (1985).

Van Cleve, J.: "Epistemic Supervenience Revisited," *Philosophy and Phenomenological Research* 59 (1999), 1049–55.

JOHN TURRI

**epistemic virtue** A virtue is an excellence of character manifesting itself through stable habits and dispositions. Following Aristotle's distinction between two kinds of human excellence – virtues of character, *ethike areté*, and virtues of thought, *dianoetike areté* – we may distinguish between, on the one hand, moral and practical virtues and, on the other hand, epistemic virtues. Narrowly defined, an *epistemic* virtue is an excellence of character instrumental to the acquisition of true belief and knowledge. According to a broader conception, epistemic virtues also include character traits instrumental to the achievement of further intellectual goals like wisdom and understanding. On the narrow conception, epistemic virtues are primarily identified with reliable cognitive faculties such as perception, memory, and sound reasoning. On the broader conception, they also include traits such as intellectual curiosity, honesty, and open-mindedness. Advocates of the narrow conception would insist that the latter traits are genuinely epistemic virtues only if they promote the acquisition of true belief and knowledge. However, other authors have questioned the distinction between broadly intellectual and narrowly epistemic virtues (Bloor, 1984; Laudan, 1984).

Broadly speaking, to say that an agent manifests an epistemic virtue is to evaluate positively the way that the agent has conducted an inquiry or acquired/maintained a belief. Such evaluation becomes relevant to epistemology once we impose an *aretic* *virtue-theoretic* condition on knowledge. According to such a condition, an agent's belief is an instance of knowledge only if the belief's acquisition is due to the agent's stable

intellectual character traits. What motivates imposing an aretaic condition is the recognition that knowledge is incompatible with *epistemic luck*: a belief's being true due to coincidence. If a belief is true as a result of epistemic luck, it is not an epistemic achievement and thus does not qualify as an instance of knowledge. Epistemologists are thus confronted with an important desideratum: to articulate a condition of knowledge that prevents a belief from being true due to luck. According to advocates of virtue epistemology, the sought-after condition is an aretaic one: a condition to be articulated by appeal to the agent's intellectual character traits.

One way of addressing the issue of epistemic luck is to make what is known as *safety* a necessary condition of knowledge: An agent's true belief that *p* is an instance of knowledge that *p* only if the agent's belief could not easily have been false, that is, if the agent's belief will not be false in a broad range of cases relevantly similar to the case we are evaluating. The question at hand is this: Must the safety condition be viewed as an aretaic condition, as a condition involving essential appeal to epistemic virtues? According to Duncan Pritchard, the safety condition is non-aretaic since, once it is imposed, it leaves no substantive role to play for the virtues, thus making any aretaic condition redundant (Pritchard, 2005). In response to such concerns, Ernest Sosa has argued that a belief's being safe is essentially due to the agent's stable intellectual character traits. So, according to Sosa, safety, if properly understood, must be viewed as an aretaic condition (Sosa, 2007).

Pritchard argues that a safety-based anti-luck account makes an aretaic condition redundant, but he allows the intellectual virtues to serve important roles *outside* analysis of knowledge – for example, roles connected with how we acquire. There may be two separate projects to engage, where the one that involves the virtues focuses not on analyzing knowledge, but on concerns such as epistemic rationality. The intellectual vices and virtues remain worthy of study in this project even if they have nothing directly to contribute to a successful theory of knowledge.

Other virtue theorists are drawn to a “two projects” approach because it expands epistemology by encouraging study of the intellectual virtues in relation to a broader range of epistemic ends such as wisdom, understanding, and knowing how (Kvanvig, 1992; Riggs, 2007a). Still, if one maintains that knowledge is an achievement, and that a safety-based account fails properly to capture this point, one will argue that veritic luck is not the only anti-luck concern which epistemologists face. In this case one might hold an aretaic condition to be required for knowledge, even while acknowledging the need for an independent safety condition to address veritic luck (Axtell, 2007).

In recent years, a new body of literature on epistemic value epistemological axiology has emerged. The themes discussed therein are best-approached in terms of *ensembles*: complex theories made up of a variety of elements, combining claims about the nature and the value of epistemic virtue. According to the *Reliabilist-Instrumentalist Ensemble*, an agent can acquire knowledge – by exercising reliable faculties – even while lacking reflectively good grounds for the proposition she knows. But such instances of knowledge fall short of being instances of *reflective* knowledge. Hence, according to this approach, we must distinguish between faculty virtues and reflective virtues, the latter grounding reflective knowledge, the former non-reflective knowledge (Sosa, 2007).

Alvin Goldman, a well-known advocate of the reliabilist approach, views true belief as the cardinal value. The epistemic virtues are related to this value teleologically: they are processes, traits, or actions that tend to produce or promote true belief (Goldman, 2002). According to other advocates of the R-I Ensemble such as Greco and Sosa, an epistemic virtue is a competence to acquire true beliefs in a particular field (Sosa, 2007; Greco, 2002). Embedded in this approach is the claim that truth is intrinsically valuable, for without this claim the contrast between truth as a *telos* and epistemic virtues as the instrumental means for securing it could not be drawn. At the heart of R-I Ensemble, then, we find a conception of truth as an intrinsic end and the

epistemic virtues as the instrumental means to achieve this end, resulting in a reliabilist account of the virtues and an instrumentalist account of the virtues.

Critics of the R–I Ensemble question the very claim that truth is the core epistemic goal, charging its proponents with holding an austere externalism akin to a “machine-model” of knowledge. Flowing from this concern, critics have articulated several objections. First, Linda Zagzebski has argued that the R–I Ensemble is unable to account for the intuition that knowledge is valuable and indeed more than mere true belief (Zagzebski, 2003). Second, it has been suggested that the aim of belief formation should explain both the authority of epistemic norms and also their motivational efficacy. If the aim is narrowly identified with truth alone, this desideratum cannot be met (Owens, 2006). According to the third objection, proponents of the R–I Ensemble misconceive the problem of non-veritic luck and therefore function of anti-luck and/or *aretai* on knowing. As a result, R–I treatments of luck do not reflect our everyday practices of knowledge attribution, which are sensitive to the way *non-veritic* kinds of epistemic luck (constitutive, evidential, and circumstantial) undercut our responsibility and control (Riggs, 2007b; Axtell 2007). Finally, critics of the approach of austere externalism objected to its problematic commitment to epistemic value monism (Jones, 1997; Grimm, 2007; Kelly, 2003). Such a commitment is manifest, for example, in Goldman’s externalist brand of virtue epistemology, at the heart of which lies what he calls *veritism*: the view that truth is the primary epistemic value (Goldman and Olsson, 2007). Epistemic value monism, then, is the thesis that, among the various epistemic values, truth is the primary one, and that all other epistemic values derive their status as epistemic values from the value of truth. Advocates of epistemic value pluralism reject this claim, insisting that, in addition to truth, there are further, independent epistemic values such as justification, understanding, and wisdom.

At the other end of the spectrum of available views, Zagzebski advocates a neo-

Aristotelian approach to the virtues. Combining a motivationist account of the nature of the virtues with a constitutivist of the value of the virtues, her view, sharply opposed to the externalist approach of the R–I Ensemble, could be classified as a *Motivationist–Constitutivist Ensemble*. According to the M–C Ensemble, since a good character’s goodness is intrinsic, intellectual virtues are valuable not by virtue of being suitably related to some *telos* but by virtue of being valuable in themselves. Indeed, intellectual virtue is partly constitutive of the intellectual *telos*. The *telos* turn is conceived in holistic and normative terms: for example, in terms of intellectual flourishing. What is important in any such account is the personal worth of the virtues, which is a function of motivational and other factors internal to agency (Dancy, 2000). For Zagzebski, epistemic value arises from what we care about; it is caring that gives rise to the demand for epistemic conscientiousness (Zagzebski, 2003).

In Zagzebski’s account, then, motivation plays a double role: the epistemic virtues arise from suitable motivations, and such motivations render the virtues valuable. Such a motivationist account of the nature of the virtues is non-consequentialist: a trait’s status as an intellectual virtue is not a function of its external cognitive outputs or consequences, but rather of motives and characteristic emotions that attend to paradigmatic instances of human knowing. Knowledge, then, demands reflective engagement, so that by implication beliefs formed in an insufficiently reflective fashion – for example, those of animals, young children, or unenlightened chicken-sexers – do not qualify as instances of knowledge.

Objections to the M–C Ensemble focus on the question of whether a motivation-based account of the virtues can provide the needed degree of unity to the sources of epistemic virtue to avoid appearing *ad hoc* (Greco, 2002). Zagzebski argues that a strong unification of the intellectual virtues is indeed available on her view, but requires accepting *assimilation*: subsuming “the intellectual virtues under the general category of the moral virtues, or *aretai ethikai*, roughly as

Aristotle understands the latter” (1996). In sum, critics of Zagzebski’s approach object that the M–C Ensemble is non-naturalistic (Goldman, 2007), internalist despite avowals to the contrary (Pritchard, 2005), extravagantly assimilationist (Driver, 2000), and *ad hoc* as it presents a fragmented conception of the sources of virtue (Olson, 2006).

Zagzebski’s “radically motivation-based” account of the nature of epistemic virtue illustrates just how sharp the contrasts can be between the R–I and the M–C Ensembles. However, these two accounts of epistemic virtue by no means exhaust the range of available views. Many virtue theorists view epistemic evaluation as context-sensitive and intertwined with moral evaluation and practical reasoning. Along a range lying between the R–I and M–C Ensembles are an array of views qualifying as versions of either *virtue reliabilism* or *virtue responsibilism*. What motivates these views is the intention to mediate the R–I/M–C conflict in ways that allow the virtues to be integrated into a “credit-related family” of concepts (Riggs, 2007a). Zagzebski herself points out that, despite crucial differences, many reliabilists share with her a *credit theory* of knowledge. Sosa and Greco are prime examples of this compromise approach since each, while rejecting internalism, also distances himself from Goldman’s form of reliabilism on the grounds that knowledge is an achievement for which the agent deserves credit.

Why do credit theories appeal to parties who appear to be residing at opposite sides of a seemingly unbridgeable divide? Perhaps the appeal arises from a shared thesis of *epistemic compatibilism*: the view that the intuitions behind internalism and externalism needn’t be conceived as driving competing accounts of the same univocal concept, namely justification. The compatibilist yearning to retain insights from each approach within a single account of the intellectual side of life is clearly evident in both virtue reliabilist and virtue responsibilist thought. Despite their sympathies with the claim that the virtues are intrinsically valuable, responsibilists are wedded to neither neo-

Aristotelianism, nor assimilationism, nor internalism. Indeed, compatibilism is formally a kind of externalism in which the epistemic roles of the reflective virtues, responsibility and credit are shown *not* to depend upon the truth of access internalism. Examples of responsibilist treatments of epistemic virtue include care-based feminist and social epistemologies (Code, 1987; Dalmiya, 2002), responsibilist epistemologies for philosophy of religion (Roberts and Wood, 2007), and *inquiry-based* accounts of epistemic virtue, exploiting synergies between virtue theory and classical pragmatism (Axtell, 2007; Hookway, 2000; Olson, 2006).

Arguably, alternative proposals for side-stepping the clash between the R–I and M–C Ensembles – that involve complex relations between epistemic luck, virtue, and value – have not yet been adequately explored in the literature. But the fact that credit theories are often endorsed on both sides of the divide suggests that our understanding of the virtues will be advanced if the relationship between credit theories and epistemic compatibilism is explored further and alternative ensembles are articulated. Whether the R–I or the M–C Ensembles or instead some other ensemble will prevail is a question further developments in epistemology will answer.

#### BIBLIOGRAPHY

- Axtell, G.: “Two for the Show: Anti-luck and Virtue Epistemologies in Consonance,” *Synthese* (2007).  
 Axtell G., ed.: *Knowledge, Belief, and Character* (Lanham, MD: Rowman & Littlefield, 2000).  
 Baehr, J.: “Character in Epistemology,” *Philosophical Studies* (2006), 479–514.  
 Bloor, D.: “The Sociology of Reasons,” in *Scientific Rationality: The Sociological Turn*, ed. J. R. Brown (Dordrecht: Reidel, 1984).  
 Brady, M. and Pritchard, D., eds.: *Moral and Epistemic Virtues* (Chichester: Wiley-Blackwell, 2003).  
 Code, L.: *Epistemic Responsibility* (Hanover, NH: University Press of New England, 1987).  
 Dancy, J.: “Supervenience, Virtues, and Consequences,” in Axtell, G. (ed.) 2000, 73–86.  
 Dalmiya, V.: “Why Should a Knower Care?,” *Hypatia* 17: 1 (2002), 34–52.



Driver, J.: "Moral and Epistemic Virtue," in Axtell, G. (ed.) 2000, 123–34.

Goldman, A.: "The Unity of the Epistemic Virtues," in his *Pathways to Knowledge* (New York: Oxford University Press, 2002).

Goldman, A. and Olsson, E.: "Reliabilism and the Value of Knowledge," in *Epistemic Value*, ed. D. Pritchard (Oxford: Oxford University Press, 2007).

Greco, J.: "Virtue and Luck, Epistemic and Otherwise," in Brady, M. and Pritchard, D. (eds), 2003, 258–70.

Greco, J.: "Two Kinds of Intellectual Virtue," *Philosophy and Phenomenological Research* 60: 1 (2002), 179–84.

Grimm, S.: "Epistemic Normativity," in *The Value of Knowledge*, ed. A. Haddock, A. Millar and D. Pritchard (Oxford: Oxford University Press, 2007).

Hookway, C.: "Regulating Inquiry," in Axtell, G. (ed.) 2000, 149–60.

Jones, W.: "Why Do We Value Knowledge?," *American Philosophical Quarterly* (1997), 423–39.

Kelly, T.: "Epistemic Rationality as Instrumental Rationality: A Critique," *Philosophy and Phenomenological Research* 3 (2003), 612–40.

Kvanvig, J.: *The Intellectual Virtues and the Life of the Mind* (Lanham, MD: Rowman and Littlefield, 1992).

Montmarquet, J. A.: "Epistemic Virtue," *Mind* 96 (1987), 482–97.

Laudan, L.: "The Pseudo-science of Science?," in *Scientific Rationality: The Sociological Turn*, ed. J. R. Brown (Dordrecht: Reidel, 1984).

Olson, P.: "Flourishing Rationality: A Pragmatic Approach to Virtue and Reason," unpublished dissertation, 2006.

Owens, D.: "Does Belief Have an Aim?," *Philosophical Studies* 1 (2006), 105–29.

Pritchard, D.: *Epistemic Luck* (Oxford: Oxford University Press, 2005).

Quine, W. V.: "Reply to White," in *The Philosophy of W. V. Quine*, ed. L. Hahn and P. A. Schilpp (La Salle, IL: Open Court, 1986).

Riggs, W.: "The Value Turn in Epistemology," in *New Waves in Epistemology*, ed. V. Hendricks and D. H. Pritchard (Burlington, VT: Ashgate, 2007a).

Riggs, W.: "Why Epistemologists Are So Down on Their Luck," in *Synthese* (2007b).

Roberts, R. C. and Wood, W. Jay: *Intellectual Virtues: An Essay in Regulative Epistemology* (Oxford: Oxford University Press, 2007).

Sosa, E.: *A Virtue Epistemology: Apt Belief and Reflective Knowledge* (Oxford: Oxford University Press, 2007).

Zagzebski, L.: "Epistemic Value and the Primacy of What We Care About," *Philosophical Studies* 33 (2003), 353–77.

Zagzebski, L.: *Virtues of the Mind* (Cambridge: Cambridge University Press, 1996).

GUY AXTELL

**epoche** The epoche or the transcendental reduction, is for Husserl (see HUSSERL) the basic method of phenomenology (see PHENOMENOLOGY). Instead of focusing on the normal objects of our acts, be they physical objects, actions, persons or general features that many objects can have in common (see ESSENCE), one reflects on the structures of one's own consciousness, and studies the *noemata*, the features that make one's consciousness be consciousness of those objects. Although we are not normally aware of these features, they are a *sine qua non* for the appearance of a world. Husserl therefore calls them transcendental, and the reflection that leads to them, where the ordinary objects are bracketed, the transcendental reduction. Husserl got the idea of the transcendental reduction in 1906. It marks the transition from the early phenomenology of the *Logical Investigations* (1900–1) to the "idealist" phenomenology of the *Ideas* and later works, in Husserl's very special sense of "idealism".

DAGFINN FOLLESDAL

**essence (Husserl)** For Husserl (see HUSSERL) an object's essence (*eidos*) is not something peculiar to that object, but a feature that the object can share with other objects, e.g. the triangularity of a triangle or the greenness of a tree. Mathematics is the most highly developed study of essences. By performing an eidetic reduction, one passes from focusing on the individual physical object to focusing on one of its essences. By further adding a transcendental reduction (see EPOCHE), one reflects on the *noemata* (structures) of the former acts and thereby arrives at phenomenology, which is the study of the *noemata* of acts directed towards essences.

DAGFINN FOLLESDAL



**essence (Plato)** Definitions stating “what X is” are central to Plato’s philosophy (see PLATO). He often maintains that one must know “what” something is before one can know other facts about it (e.g. *Meno* 71, 100), and sometimes that to know “what” something is, one must be able to give a definition of it. He usually regards a definition as describing an *eidos*, “Form”, or *ousia*, “essence” or “being”. He regards these as objective entities, not constituted by human usage or judgements. He suggests sometimes that a definition gives an analysis of something into constituents, and sometimes that it specifies something by features distinguishing it from all other things. Plato seemingly denies that any sensible particular is an *ousia*, and so that there is any saying “what it is”, i.e. that it is or has an essence. On one interpretation, however, Plato believes in “abstract particulars”, which do have essential connections to Forms.

## BIBLIOGRAPHY

- Allen, R. E.: *Plato’s Euthyphro and the Earlier Theory of Forms* (London: Routledge and Kegan Paul, 1970).  
 Gosling, J. C. B.: *Plato* (London: Routledge, 1973), ch. 11.  
 Penner, T.: *The Ascent from Nominalism* (Dordrecht: Reidel, 1987).  
 Robinson, R.: *Plato’s Earlier Dialectic*, 2nd edn (Oxford: Clarendon Press, 1953).  
 White, N. P.: *Plato on Knowledge and Reality* (Indianapolis, IN: Hackett, 1976).

NICHOLAS P. WHITE

**essentialism** see NECESSARY/CONTINGENT.

**ethics and epistemology** What are the issues on which some knowledge of ethics would be of help, or even essential, for an epistemologist? The following are a few of the main ones.

First, what is the nature of the epistemic “ought”? This question includes such issues as the sense in which we can have epistemic duties; whether, if we cannot (suitably) control our beliefs, there can be beliefs that we ought to have or have reason to have; also

more generally the similarities and differences between epistemic and moral “oughts”, for instance the question whether some or all of our epistemic duties are moral duties. This last question can be framed in terms of character: To what extent is a good epistemic character (part of) a good moral character? This question is not solely the province of virtue epistemologists.

Second, what are the epistemic virtues and what role should investigation of them play in a complete epistemology? Clearly, virtue ethicists and virtue epistemologists have a lot to learn from each other. Epistemologists, however, do not have to be virtue theorists in order to find philosophical value in an examination of the virtues.

Third, what is a reason for believing? An enormous amount of work has been done on reasons for action by ethicists recently, much of which can be adapted for use by epistemologists. And vice-versa: Can Bayesian conceptions of epistemic reasons be adapted for use in ethics? Is a moral reason a consideration that raises the probability that the action is right?

Fourth, how should we best understand the vexed notion of epistemic justification? In ethics there have been protracted debates about whether our duties are grounded in how things are, in how the available evidence suggests that they are, or in how we suppose them to be. Struggles such as these are conducted around the distinction between objective and subjective. An analogous distinction can be found in the theory of justification.

Fifth, what is the relation between principles and particular cases? This includes the question whether our theory should be (in Chisholm’s sense) particularist or generalist, an issue that has much troubled ethicists, and also the question whether we should expect or need to find principles at all, which the ethical particularist denies.

Sixth, what is the relation between evaluation of the belief and evaluation of the believer? In ethics, much has been made of the distinction between act and agent. Is it possible to hold that the act is wrong but the agent blameless? Can analogous things be said in epistemology? Can there be beliefs

that are wrong but which the agent is blameless in forming, or is the very idea of a wrong belief (once contrasted with the wrongness of the believing) incoherent?

Seventh, we are told that belief aims at the truth. This claim, which is hard to understand and even harder to defend, is analogous to the claim that action aims at the good, a claim that is equally puzzling.

Finally, and much more obviously, virtue epistemology and feminist epistemology are consciously intended to draw on their ethical cousins and sisters.

In general, to the extent that epistemology thinks of itself as concerned to understand epistemic *practice*, it is likely to stray towards and to be capable of learning from results in ethics – always assuming that ethics is capable of producing results in the first place.

I have been suggesting that epistemologists can learn from ethicists; but of course the reverse is true, too.

#### BIBLIOGRAPHY

- Alston, W. P.: *Epistemic Justification: Essays in the Theory of Knowledge* (Ithaca, NY: Cornell University Press, 1989).
- Chisholm, R. M.: *Perceiving: A Philosophical Study* (Ithaca, NY: Cornell University Press, 1957), part I: "The Ethics of Belief."
- Clifford, W. K.: "The Ethics of Belief," in his *The Ethics of Belief and Other Essays* (London: Watts, 1947), 70–96.
- Steup, M.: *Knowledge, Truth, and Duty* (Oxford: Oxford University Press, 2001).

JONATHAN DANCY

**evidence** The notion of evidence figures prominently in several epistemological issues. A good way to raise the central philosophical questions about evidence is in the context of a discussion of a theory of epistemic justification known as "evidentialism". Evidentialism, suggested by Chisholm (1977) and defended explicitly in Feldman and Conee (1985), holds that a belief is epistemically justified for a person if and only if the person's evidence supports that belief. Working out the details of this view requires resolving several

questions about the concept of evidence, including the following: (1) What sorts of things can be evidence? (2) Under what conditions does a body of evidence support a particular proposition or belief? (3) What is it for someone to have something as evidence? Of course, these questions retain their interest whatever the merits of evidentialism.

#### WHAT IS EVIDENCE?

The concept of evidence appealed to in evidentialism, and in epistemology generally, differs from the related concept of evidence used in the law. In the law, or at least in informal discussions of the law, evidence includes physical objects and events. Weapons and footprints, for example, are ordinarily said to be evidence. In philosophical discussions, evidence is generally taken to be either internal states such as beliefs, or the believed propositions themselves. Thus, the belief (or proposition) that a weapon of certain type was used might be evidence for one person's guilt.

A crucial question about the nature of evidence is whether evidence is limited to other beliefs (or believed propositions) or whether it includes other mental states such as perceptual experiences. Various reasons have been advanced for thinking that only beliefs can be evidence, one being that the evidence for a belief confers justification on the belief, but only something that is itself justified can confer justification on anything else, and only beliefs (or other doxastic states) can be justified. (See Bonjour, 1985, ch. 4.) Sosa (1974, 1980) argues that non-doxastic states, such as experiences, can also count as evidence. On this view, some beliefs are basic, in the sense that they are justified by experience rather than by other beliefs. Sosa argues that the experiences which justify basic beliefs need not be justified themselves. Van Cleve (1985), adopting a point made by Sosa (1980), contends that only states that are themselves justified could "transmit" justification, but non-justified states might "generate" justification. Both Sosa and Van Cleve claim that since justification supervenes (*see* EPISTEMIC SUPERVENIENCE) on non-epistemic properties,

there must be some non-epistemic states that are sufficient for, and thus generate, justification. (See FOUNDATIONALISM; CRITERIA AND KNOWLEDGE; INFINITE REGRESS ARGUMENT.)

#### EVIDENTIAL SUPPORT

Holding that experiences count as evidence adds complexity to an already difficult set of questions about the evidential support relation. The new questions are about exactly what makes it the case that an experiential state counts as evidence for one belief (or proposition) rather than another. It is easy to be fooled by superficial linguistic facts that seem to link certain experiences to certain beliefs. It may seem clear that the fact that something looks blue to S justifies S in believing that the thing is blue (absent any counterevidence). More generally, if a thing looks F to S, then S is justified in believing that it is F provided S does not have any evidence against its being F. (See Van Cleve, 1985, pp. 96–7; Moser, 1985, ch. 5.) This may seem right, but the formulation masks complexities. To say that something “appears blue” to a person is to say, roughly, that it induces a certain sort of internal state in the person. To say that it is blue is to say that it has certain physical properties of some sort. It appears, pending further analysis of colour words, that these propositions are only contingently related and that our inclination to think it obvious that one justifies the other results from the accident that the word “blue” appears in the sentences used to express both propositions. This temptation would be eliminated if we described the internal appearance state in some other terminology. (Why think that the fact that an object appears in manner 1,256 justifies the belief that the object is blue?)

Furthermore, if one says that the experience of seeing a blue object normally justifies one in believing that one sees something blue, then it is hard to see how one can avoid saying that the experience of seeing a 23-sided object normally justifies one in believing that one sees something 23-sided. But this conclusion is implausible; not all experiences typically justify the corresponding proposition

about the experienced object. (See Sosa, 1988, p. 171.)

To this, one might reply that to those of us who aren't equipped to “pick up” on 23-sidedness, 23-sided things don't appear 23-sided, whereas blue things typically do appear blue to us. So, the cases are disanalogous. This reply raises questions about the nature of appearances. Imagine a person who was designed to sense 23-sidedness. It seems possible that the visual image that such a person has when looking at a 23-sided object would be the same as the one a normal person would have when looking at that object. But the reply holds that the 23-sided object appears differently to these two individuals. While there is a difference in their abilities to extract information from a visual array, it is difficult to understand what makes their appearances different.

In addition to the questions about how experiential states provide evidential support, there are many traditional epistemological issues which can be framed as questions about the nature or extension of the evidential support relation. Traditional debates about our knowledge of the external world or of other minds and questions about knowledge based on induction are largely questions about the adequacy of our evidence for external world propositions, propositions about other minds, and of inductive evidence generally.

It is extraordinarily difficult to state in a general way the conditions under which a body of evidence provides evidential support for a belief. The mere existence of a logical or probabilistic connection between the evidence and the belief is not sufficient for evidential support. If it were adequate, then all the distant and unseen necessary or probabilistic consequences of one's justified beliefs would themselves be justified. Since that is clearly unacceptable, one might say instead that if evidence *e* provides epistemic support for proposition *p* for person S, then *e* must entail or make probable *p* and S must “grasp” the connection between *e* and *p*. This reply seems to over-intellectualize the situation, since people seem not to grasp such matters routinely, and it invites a troublesome

regress if requiring this “grasp” of the evidential connection amounts to requiring the justified belief that *e* supports *p*. There is no generally accepted view about what is necessary or sufficient for epistemic support.

#### HAVING EVIDENCE

A further question about evidence concerns exactly what it is to *have* something as evidence. Stored somewhere in one's memory are an enormous number of facts. Many of these facts may bear on some proposition, *p*, that one believes. While considering *p*, one may think of only some of these stored facts. If prompted in one way, one might recall some of these facts, and if prompted in other ways, one might recall other facts. Some of them may be accessible only with complex and detailed prompting. But which of these facts are part of the evidence one has and are relevant to the assessment of the epistemic merit of the current belief? A highly restrictive view would limit the evidence to what one actually has currently in mind. A highly liberal view would include as part of one's evidence everything stored in one's mind. This renders justified some beliefs that seem, from an intuitive viewpoint, quite unreasonable. There is no clearly acceptable way to carve out a theory positioned between these two extremes (see Feldman, 1988).

#### RIVALS TO EVIDENTIALISM

A different set of questions about evidence concerns the connection between evidence and epistemic justification. Evidentialism holds that questions about epistemic justification turn entirely upon matters pertaining to evidence. Rival views hold that other sorts of matters play a central role in determining which beliefs are justified. For example, Kornblith (1983) argues that a belief is epistemically justified only if the believer has gone about gathering evidence for it in an epistemically responsible manner (see EPISTEMIC VIRTUE). Goldman (1986) defends reliabilism (see RELIABILISM) which, like some other causal theories of justification, implies

that having supporting evidence is neither necessary nor sufficient for justification, since on standard understandings of reliabilism, a belief can be caused in a reliable way even though the believer does not have anything that could plausibly be regarded as good evidence for it. The debate on these matters is surely not settled, but it is instructive to notice that defenders of evidentialism's rivals, such as Goldman (1986), often go to some lengths to adjust their theories so that they share the straightforward implications of evidentialism. They do not defend the implications of the simple versions of their theories.

#### BIBLIOGRAPHY

- BonJour, L.: *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1985).
- Chisholm, R. M.: *Theory of Knowledge*, 2nd edn (Englewood Cliffs, NJ: Prentice-Hall, 1977).
- Feldman, R.: “Having Evidence,” in *Essays Presented to Edmund Gettier*, ed. D. Austin (Dordrecht: Kluwer, 1988), 83–104.
- Feldman, R. and Conee, E.: “Evidentialism,” *Philosophical Studies* 48 (1985), 15–34.
- Feldman, R. and Conee, E.: *Evidentialism: Essays in Epistemology* (Oxford: Clarendon Press, 2004).
- Goldman, A.: *Epistemology and Cognition* (Cambridge, MA: Harvard University Press, 1986).
- Kornblith, H.: “Justified Belief and Epistemically Responsible Action,” *Philosophical Review* 92 (1983), 33–48.
- Moser, P.: *Empirical Justification* (Dordrecht: Reidel, 1985).
- Sosa, E.: “How Do You Know?,” *American Philosophical Quarterly* 11 (1974), 113–22.
- Sosa, E.: “The Foundations of Foundations,” *Nous* 14 (1980), 547–64.
- Sosa, E.: “Beyond Scepticism, to the Best of Our Knowledge,” *Mind* 97 (1988), 153–88.
- Van Cleve, J.: “Epistemic Supervenience and the Rule of Belief,” *The Monist* 68 (1985), 90–104.

RICHARD FELDMAN

**evolutionary argument against naturalism** An argument proposed by Alvin Plantinga (henceforth EAAN), which purports to show that metaphysical naturalism is self-defeating and hence cannot be rationally

accepted. In addition, Plantinga argues that theism does not face self-defeat in the same way that naturalism does. In what follows, I shall describe EAAN and outline some of the main objections to it.

To begin with, let "N" stand for metaphysical naturalism, the claim that there is no God and nothing like God; let "E" stand for the view that human cognitive faculties have evolved by way of the mechanisms that are studied by contemporary evolutionary theory; and let "R" stand for the claim that the beliefs produced by those cognitive faculties are for the most part true.

EAAN has three stages, each of which involves defending a certain premise:

- (1)  $P(R/N\&E)$  is either low or inscrutable (meaning that we cannot determine whether it is low or high). Call this the *Probability Thesis*.
- (2) Anyone who accepts N and E and the Probability Thesis has a defeater for R. This is the *Defeater Thesis*.
- (3) Anyone who has a defeater for R has an undefeated defeater for each of his beliefs.

From these premises, it follows that anyone who accepts N and E and the Probability Thesis has an undefeated defeater for each of his beliefs, including his belief in metaphysical naturalism. But one who is a naturalist must accept E (it is, says Plantinga, the only option for the naturalist when it comes to explaining the diversity of life). Hence, naturalism is self-defeating. Let us see how these three premises are defended.

Plantinga defends the Probability Thesis by inviting us to consider the case of a hypothetical population of creatures on a planet a lot like earth, formed by blind, undirected evolution, and to assume that naturalism is true. What is  $P(R/N\&E)$  specified, not to us, but to them? Plantinga notes that, when we consider this hypothetical population, there are four possibilities:

- P1: There is no causal connection between belief and behavior.
- P2: Beliefs are the effects of behavior but are not among the causes of behavior.

P3: Beliefs do causally affect behavior, but not by virtue of their content.

P4: Beliefs do causally affect behavior in virtue of their content.

Plantinga then says that, since these four possibilities are jointly exhaustive and mutually exclusive, the probability we want to assess, namely  $P(R/N\&E)$ , is given by the following weighted average:

$$\begin{aligned} P(R/N\&E) = & P(R/N\&E\&P1) P(P1/N\&E) \\ & + P(R/N\&E\&P2) P(P2/N\&E) \\ & + P(R/N\&E\&P3) P(P3/N\&E) \\ & + P(R/N\&E\&P4) P(P4/N\&E). \end{aligned}$$

The Probability Thesis is then justified by estimating this weighted average.  $P(R/N\&E\&Pi)$  is estimated as low for  $i = 1, 2, 3$ , because in these cases beliefs will be invisible to natural selection and so there will be no selection pressure towards their being mostly true. It seems, initially, as though  $P(R/N\&E\&P4)$  is going to be very high, but Plantinga contests this estimate by presenting examples of beliefs which are false but which, when combined with strange desires, lead to felicitous action. In the latter case, Plantinga concludes that the probability will be at best moderately high, not very much more than a half.

It now remains to estimate the probabilities of the form  $P(Pi/N\&E)$ , for  $i = 1, 2, 3, 4$ . Here, Plantinga thinks that, because of the enormous difficulties that naturalists (almost all of whom are at present materialists) face in avoiding P3,  $P(P3/N\&E)$  is very high. Now, P1, P2, P3, and P4 are mutually exclusive and jointly exhaustive, and their respective probabilities sum to 1. Thus, each of P1, P2, and P4 must be estimated as having low probability on N&E. Plantinga claims that a reasonable estimate of the probabilities leads to an estimate of  $P(R/N\&E)$  as being somewhat less than a half.

Plantinga grants, however, that estimating probabilities in this sort of context is a dubious business. So he concedes that it would be proper to take the relevant probabilities to be inscrutable to us, leading to the conclusion that  $P(R/N\&E)$  is inscrutable to us. In this way, Plantinga arrives at his



conclusion that  $P(R/N\&E)$  is either low or inscrutable.

In his self-profile in this volume, Plantinga has given a new argument for the Probability Thesis, which does not consider different possibilities for the relation between belief and action, and which supports the stronger conclusion that  $P(R/N\&E)$  is low (rather than the conclusion that it is low or inscrutable).

The Defeater Thesis is defended by appealing to hypothetical cases that, it is claimed, are clearly analogous to the case of the naturalist in EAAN. Since, in these cases, the subject has a defeater for R, the same is true of the naturalist who accepts the Probability Thesis. Two hypothetical cases that have tended to predominate in discussions of EAAN are *The Case of the Cartesian Demon* and *The Case of the Drug XX*. The former is described below, and a version of the latter is described in Plantinga's self-profile in this volume.

#### *The Case of the Cartesian Demon*

Suppose a man comes to believe that he is the creation of a demon that, as imagined by Descartes, is immensely powerful and immensely knowledgeable. Suppose that he also comes to believe that this demon is not particularly concerned with making his creations cognitively reliable, and on at least some occasions has been quite pleased to make them unreliable, and moreover has made them unreliable in such a way that they continue to think of themselves as paragons of reliability, being unable to detect the cognitive disaster that has befallen them. Thinking about this, the man comes to the conclusion that  $P(R/D)$  is low or inscrutable, where R is specified to himself, and D is the proposition that the man has been created by the demon. Then the man has a defeater for R.

Plantinga defends the third premise by arguing that, if the naturalist has a defeater for R, this generates a defeater for the rest of his beliefs as well. The reason is that all of the naturalist's beliefs are products of his cognitive faculties, which constitute their source. Once the reliability of that source comes into question, so do the beliefs generated by the source. Moreover, the defeater for R that the naturalist acquires cannot itself be defeated,

since everything that could be a defeater-defeater is itself subject to defeat. To support this, Plantinga says that to rely on one's cognitive faculties to form a defeater-defeater of the defeater one has for R would be like trusting a man to tell you he is not a liar when you have already been given excellent reasons to doubt his honesty.

Let us now consider some objections to EAAN. Most of the controversy regarding the argument has focused on the Defeater Thesis. There has been one main worry that critics have had about this claim. The objections to it that we shall describe are manifestations of this worry, which can be expressed as follows: what exactly is the connection between the naturalist's acceptance of the Probability Thesis on the one hand, and her acquisition of a defeater for R on the other? One of the most natural expressions of this worry is the Perspiration Objection.

#### *The Perspiration Objection*

The probability that the function of perspiration is to cool the body given (just) N&E is also low. But surely it would be absurd to claim that this gives the naturalist a defeater for this belief. Thus, it is also absurd to claim that the naturalist has a defeater for R in virtue of accepting the Probability Thesis.

There is no defeater in the perspiration case because the naturalist has other evidence for his beliefs about the function of perspiration, beyond just N&E. So could not the naturalist appeal to other evidence for his beliefs about R? This thought leads naturally to the Total Evidence Objection for EAAN.

#### *The Total Evidence Objection*

The naturalist has many other beliefs besides N&E. The probability of R relative to N&E conjoined with these other beliefs is quite high. Thus, the naturalist need not have a defeater for R in virtue of accepting the Probability Thesis.

Many philosophers (including Plantinga) hold that, in addition to propositional evidence, beliefs can also be warranted in virtue of *non-propositional* evidence. This leads to yet another objection, due to Michael Bergmann,

which we can call the Non-propositional Evidence Objection.

*The Non-propositional Evidence Objection*

Even if R has low probability on all the available propositional evidence, the naturalist could still have non-propositional evidence for R which makes it rational to continue to hold on to R. Hence, the naturalist need not have a defeater for R merely in virtue of accepting the Probability Thesis.

These objections comprise just a small sample of the arguments against EAAN that have appeared in the published literature on the argument. Many of these, along with Plantinga's responses to them, are articulated and discussed in Beilby (2002).

See also DEFEASIBILITY; and the self-profile of ALVIN PLANTINGA in Part II.

#### BIBLIOGRAPHY

- Beilby, James, ed.: *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument against Naturalism* (Ithaca, NY, and London: Cornell University Press, 2002).  
 Bergmann, Michael: "Commonsense Naturalism," in Beilby 2002, 61–90.

OMAR MIRZA

**evolutionary epistemology** This is an approach to the theory of knowledge that sees an important connection between the growth of knowledge and biological evolution.

An evolutionary epistemologist claims that the development of human knowledge proceeds through some natural selection process, the best example of which is Darwin's theory of biological natural selection. The three major components of the model of natural selection are variation, selection and retention. According to Darwin's theory of natural selection, variations are not pre-designed to perform certain functions. Rather, those variations that perform useful functions are selected, while those that do not are not selected; such selection is responsible for the appearance that variations intentionally occur. In the

modern theory of evolution, genetic mutations provide the blind variations (blind in the sense that variations are not influenced by the effects they would have – the likelihood of a mutation is not correlated with the benefits or liabilities that mutation would confer on the organism), the environment provides the filter of selection, and reproduction provides the retention. Fit is achieved because those organisms with features that make them less adapted for survival do not survive in competition with other organisms in the environment that have features which are better adapted. Evolutionary epistemology applies this blind variation and selective retention model to the growth of scientific knowledge and to human thought processes in general.

The parallel between biological evolution and conceptual (or "epistemic") evolution can be seen as either literal or analogical. The literal version of evolutionary epistemology sees biological evolution as the main cause of the growth of knowledge. On this view, called the "evolution of cognitive mechanisms program" (EEM) by Bradie (1986) and the "Darwinian approach to epistemology" by Ruse (1986), the growth of knowledge occurs through blind variation and selective retention because biological natural selection itself is the cause of epistemic variation and selection. The most plausible version of the literal view does not hold that all human beliefs are innate but rather that the mental mechanisms which guide the acquisition of non-innate beliefs are themselves innate and the result of biological natural selection. Ruse (1986, ch. 5) defends a version of literal evolutionary epistemology which he links to sociobiology. (See Bradie's essay in Rescher, 1990, pp. 33–8, for criticism of this view.)

On the analogical version of evolutionary epistemology, called the "evolution of theories program" (EET) by Bradie (1986) and the "Spencerian approach" (after the nineteenth-century philosopher Herbert Spencer) by Ruse (1986), the development of human knowledge is governed by a process analogous to biological natural selection, rather than by an instance of the mechanism itself. This version of evolutionary epistemology, introduced and elaborated by Donald

Campbell (1974a and 1974b) as well as Karl Popper (*see* POPPER), sees the (partial) fit between theories and the world as explained by a mental process of trial and error known as epistemic natural selection.

Both versions of evolutionary epistemology are usually taken to be types of naturalized epistemology (*see* NATURALIZED EPISTEMOLOGY) because both take some empirical facts as a starting point for their epistemological project (*see* Quine, 1969, and *see* QUINE). The literal version of evolutionary epistemology begins by accepting evolutionary theory and a materialist approach to the mind and, from these, constructs an account of knowledge and its development. In contrast, the analogical version does not require the truth of biological evolution; it simply draws on biological evolution as a source for the model of natural selection. For this version of evolutionary epistemology to be true, the model of natural selection need only apply to the growth of knowledge, not to the origin and development of species. Crudely put, evolutionary epistemology of the analogical sort could still be true even if creationism is the correct theory of the origin of species.

Although they do not begin by assuming evolutionary theory, most analogical evolutionary epistemologists are naturalized epistemologists as well; their empirical assumptions (at least implicitly) come from psychology and cognitive science, not evolutionary theory. Sometimes, however, evolutionary epistemology is characterized in a seemingly non-naturalistic fashion. Campbell (1974b, p. 142) says that "if one is expanding knowledge beyond what one knows, one has no choice but to explore without the benefit of wisdom" (i.e. blindly). This, Campbell admits, makes evolutionary epistemology close to being a tautology (and so not naturalistic). Evolutionary epistemology does assert the analytic claim that when expanding one's knowledge beyond what one knows, one must proceed to something that is not already known, but, more interestingly, it also makes the synthetic claim that when expanding one's knowledge beyond what one knows, one must proceed by blind variation and selective retention. This

claim is synthetic because it can be empirically falsified. The central claim of evolutionary epistemology is *synthetic*, not *analytic*. If the central claim were analytic, then all non-evolutionary epistemologies would be logically contradictory, which they are not. Campbell is right that evolutionary epistemology does have the analytic feature he mentions, but he is wrong to think that this is a *distinguishing* feature, since any plausible epistemology has the same analytic feature (*see* Skagestad, 1978, p. 613).

Two of the deeper issues that arise in the literature involve questions about realism (i.e. what sort of metaphysical commitments does an evolutionary epistemologist have to make?) (*see* REALISM), and progress (i.e. according to evolutionary epistemology, does knowledge develop towards a goal?). (On realism, *see* Campbell, 1974a, pp. 447–50; Bradie, 1986, pp. 444–51; Skagestad, 1978, pp. 617–19; Ruse's essay in Rescher, 1990, pp. 101–10; and Stein's essay in Rescher, 1990, pp. 119–29. On progress, *see* Bradie, 1986, pp. 426–7; Ruse, 1986, and Stein, 1990.) With respect to realism, many evolutionary epistemologists endorse what is called hypothetical realism, a view that combines a version of epistemological scepticism (*see* SCEPTICISM) and tentative acceptance of metaphysical realism (*see* REALISM). With respect to progress, the problem is that biological evolution is not goal-directed, but the growth of human knowledge seems to be. Campbell (1974a) worries about the potential disanalogy here but is willing to bite the bullet and admit that epistemic evolution progresses towards a goal (truth) while biological evolution does not. Some have argued that evolutionary epistemologists must give up the "truth-tropic" sense of progress because a natural selection model is in essence non-teleological; instead, following Kuhn (1970), an operational sense of progress can be embraced along with an evolutionary epistemology.

Among the most frequent and serious criticisms levelled against evolutionary epistemology is that the analogical version of the view is false because epistemic variation is not blind (*see*, for example, Skagestad, 1978, 613–16; and Ruse, 1986, ch. 2). Stein and

Lipton (1990) have argued, however, that this objection fails because, while epistemic variation is non random, its constraints come from heuristics which, for the most part, are the result of epistemic blind variation and selective retention. Further, Stein and Lipton argue that these heuristics are analogous to biological pre-adaptations, evolutionary precursors (such as a half-wing, a precursor to a wing) which have some function other than the function of their descendent structures; the guidedness of epistemic variation is, on this view, not the source of disanalogy, but the source of a more articulated account of the analogy.

Many evolutionary epistemologists try to combine the literal and the analogical versions (see Bradie, 1986, pp. 403–11; Stein and Lipton, 1990, pp. 42–6), saying that those beliefs and cognitive mechanisms which are innate result from natural selection of the biological sort and those which are not innate result from natural selection of the epistemic sort. This is reasonable as long as the two parts of this hybrid view are kept distinct. An analogical version of evolutionary epistemology with *biological* variation as its only source of blindness would be a null theory; this would be the case if all our beliefs are innate or if our non-innate beliefs are not the result of blind variation. An appeal to the blindness of biological variation is thus not a legitimate way to produce a hybrid version of evolutionary epistemology since doing so trivializes the theory. For similar reasons, such an appeal will not save an analogical version of evolutionary epistemology from arguments to the effect that epistemic variation is not blind (see Stein and Lipton, 1990, pp. 42–5).

Although it is a relatively new approach to theory of knowledge, evolutionary epistemology has attracted much attention, primarily because it represents a serious attempt to flesh out a naturalized epistemology by drawing on several disciplines. If science is relevant to understanding the nature and development of knowledge, then evolutionary theory is among the disciplines worth a look. Insofar as evolutionary epistemology looks there, it is an interesting and potentially fruitful epistemological programme.

See also GENETIC EPISTEMOLOGY; NATURALIZED EPISTEMOLOGY.

## BIBLIOGRAPHY

- Bradie, M.: "Assessing Evolutionary Epistemology," *Biology and Philosophy* 1 (1986), 401–59.
- Callebaut, W. and Pinxter, R., eds: *Evolutionary Epistemology: A Multiparadigm Program* (Dordrecht: Reidel, 1987); includes an extensive bibliography.
- Campbell, D. T.: "Evolutionary epistemology," in *The Philosophy of Karl Popper*, book 1, ed. P. A. Schilpp (LaSalle, IL: Open Court, 1974a), 413–63; reprinted in *Evolution, Theory of Rationality and the Sociology of Knowledge*, ed. G. Radnitsky and W. W. Bartley III (LaSalle, IL: Open Court, 1987), 47–89.
- Campbell, D. T.: "Unjustified Variation and Selection in Scientific Discovery," in *Studies in Philosophy of Biology*, ed. F. J. Ayala and T. Dobzhansky (Berkeley, CA: University of California Press, 1974b), 139–61.
- Kuhn, T.: *The Structure of Scientific Revolutions*, 2nd edn (Chicago, IL: University of Chicago Press, 1970).
- Quine, W. V.: "Epistemology Naturalized," in his *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969), 69–90.
- Rescher, N., ed.: *Evolution, Cognition and Realism* (Lanham, MD: University Press of America, 1990).
- Ruse, M.: *Taking Darwin Seriously* (Oxford: Blackwell, 1986).
- Skagestad, P.: "Taking Evolution Seriously: Critical Comments on D.T. Campbell's Evolutionary Epistemology," *The Monist* 61 (1978), 611–21.
- Stein, E. and Lipton, P.: "Where Guesses Come From: Evolutionary Epistemology and the Anomaly of Guided Variation," *Biology and Philosophy* 4 (1990), 33–56.

EDWARD STEIN

**existence** Philosophers often debate the existence of different kinds of things: nominalists question the reality of abstract objects like classes, numbers and universals; some positivists doubt the existence of theoretical entities like neutrons or genes; and there are debates over whether there are sense-data, events, and so on. Some philosophers may be happy to talk about abstract objects and theoretical entities while denying that they

really exist. This requires a “metaphysical” concept of “real existence”: we debate whether numbers, neutrons and sense-data are really existing things. But it is difficult to see what this concept involves and the rules to be employed in settling such debates are very unclear.

Questions of existence seem always to involve general kinds of things: do numbers, sense data or neutrons exist? Some philosophers conclude that existence is not a property of individual things, “exists” is not an ordinary predicate. If I refer to something, and then predicate existence of it, my utterance seems to be tautological: the object must exist for me to be able to refer to it, so predicating existence of it adds nothing. And to say of something that it did not exist would be contradictory.

According to Rudolf Carnap (see CARNAP), philosophical questions of existence always concern whether to adopt a general linguistic or conceptual framework. Questions of which framework to employ do not concern whether the entities posited by the framework “really exist”; they are rather settled by its pragmatic usefulness. Philosophical debates over existence misconstrue “pragmatic” questions of choice of framework as substantive questions of fact. Once a framework is adopted, there are substantive “internal” questions: are there any prime numbers between 10 and 20? “External” questions about choice of framework have a different status.

More recent philosophers, notably Quine (see QUINE), have questioned the distinction between linguistic framework and internal questions arising within it. But Quine agrees that we have no “metaphysical” concept of existence against which different purported entities can be measured. If quantification over numbers (or over sense-data) forms part of the general theoretical framework which best explains our experience, the claim that there are such things (that they exist) is true. Scruples about admitting the existence of too many different kinds of objects depend not on a metaphysical concept of existence but rather on a desire for a simple and economical theoretical framework.

See also ONTOLOGICAL COMMITMENT.

## BIBLIOGRAPHY

- Carnap, R.: *Meaning and Necessity*, 2nd edn (Chicago, IL: Chicago University Press, 1956), 205–21.  
 Moore, G. E.: “Is Existence a Predicate?,” *Proceedings of the Aristotelian Society*, supp. vol. 15 (1936), 175–88.  
 Quine, W. V.: *From a Logical Point of View* (Cambridge, MA: Harvard University Press, 1953), ch. 1.  
 Quine, W. V.: *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969), ch. 4.

CHRISTOPHER HOOKWAY

**experience, theories of** It is not possible to define experience in an illuminating way. Readers, however, know what experiences are through acquaintance with some of their own, e.g. a visual experience of a green after-image, a feeling of physical nausea or a tactile experience of an abrasive surface (which might be caused by an actual surface – rough or smooth – or which might be part of a dream, or the product of a vivid sensory imagination).

The essential feature of every experience is that it *feels* a certain way – that there is something that it is like to have it. We may refer to this feature of an experience as its *character*.

Another core feature of the sorts of experiences with which this article is concerned is that they have representational *content*. (Unless otherwise indicated, the term “experience” will be reserved for these below.) The most obvious cases of experiences with content are sense experiences of the kind normally involved in perception. We may describe such experiences by mentioning their sensory modalities and their contents, e.g. *a gustatory experience (modality) of chocolate ice cream (content)*, but do so more commonly by means of perceptual verbs combined with noun phrases specifying their contents, as in “Macbeth saw a dagger”. This is, however, ambiguous between the perceptual claim “There was a (material) dagger in the world which Macbeth perceived visually” and “Macbeth had a visual experience of a dagger” (the reading with which we are concerned).



As in the case of other mental states and events with content, it is important to distinguish between the properties which an experience *represents* and the properties which it *possesses*. To talk of the representational properties of an experience is to say something about its content, not to attribute those properties to the experience itself. Like every other experience, a visual experience of a pink square is a mental event, and it is therefore not itself either pink or square, even though it represents those properties. It is, perhaps, fleeting, pleasant or unusual, even though it does not represent those properties. An experience may represent a property which it possesses, and it may even do so in virtue of possessing that property, as in the case of a rapidly changing (complex) experience representing something as changing rapidly, but this is the exception and not the rule.

Which properties can be (directly) represented in sense experience is subject to debate. Traditionalists include only properties whose presence could not be doubted by a subject having appropriate experiences, e.g., colour and shape in the case of visual experience, and (apparent) shape, surface texture, hardness, etc., in the case of tactile experience. This view is natural to anyone who has an egocentric, Cartesian (*see* CARTESIANISM) perspective in epistemology, and who wishes for pure data in experience to serve as logically certain foundations for knowledge (*see* SENSE-DATA). Others who do not think that this wish can be satisfied, and who are more impressed with the role of experience in providing animals with ecologically significant information about the world around them, claim that sense experiences represent properties, characteristics and kinds which are much richer and much more wide-ranging than the traditional sensory qualities. We don't see only colours and shapes, they tell us, but also earth, water, men, women and fire; we don't smell only odours, but also food and filth. There is no space here to examine the factors relevant to a choice between these alternatives. In this article the more liberal view will be assumed except when it is incompatible with a position under discussion.

Given the modality and content of a sense experience, most of us will be aware of its character even though we cannot describe that character directly. This suggests that character and content are not really distinct, and there is a close tie between them. For one thing, the relative complexity of the character of a sense experience places limitations on its possible content; e.g. a tactile experience of something touching one's left ear is just too simple to carry the same amount of content as a typical everyday visual experience. Furthermore, the content of a sense experience of a given character depends on the normal causes of appropriately similar experiences, e.g. the sort of gustatory experience which we have when eating chocolate would not represent chocolate unless it were normally caused by chocolate. Granting a contingent tie between the character of an experience and its possible causal origins, it again follows that its possible content is limited by its character.

Character and content are none the less irreducibly different, for the following reasons. (1) There are experiences which completely lack content, e.g. certain bodily pleasures. (2) Not every aspect of the character of an experience with content is relevant to that content, e.g. the unpleasantness of an aural experience of chalk squeaking on a board may have no representational significance. (3) Experiences in different modalities may overlap in content without a parallel overlap in character, e.g. visual and tactile experiences of circularity feel completely different. (4) The content of an experience with a given character may vary according to the background of the subject, e.g. a certain aural experience may come to have the content "singing bird" only after the subject has learned something about birds.

#### ONTOLOGICAL THEORIES

##### *The act/object analysis*

According to the act/object analysis of experience (which is a special case of the act/object analysis of consciousness) (*see* ACT/

OBJECT ANALYSIS), every experience involves an object of experience even if it has no material object. Two main lines of argument may be offered in support of this view, one phenomenological and the other semantic.

In outline, the phenomenological argument is as follows. Whenever we have an experience, even if nothing beyond the experience answers to it, we seem to be presented with something *through* the experience (which is itself diaphanous). The object of the experience is whatever is so presented to us – be it an individual thing, an event, or a state of affairs.

The semantic argument is that objects of experience are required in order to make sense of certain features of our talk about experience, including, in particular, the following. (1) Simple attributions of experience (e.g. “Rod is experiencing a pink square”) seem to be relational. (2) We appear to refer to objects of experience and to attribute properties to them (e.g. “The after-image which John experienced was green”). (3) We appear to quantify over objects of experience (e.g. “Macbeth saw something which his wife did not see”).

The act/object analysis faces several problems concerning the status of objects of experience. Currently the most common view is that they are sense-data (*see* SENSE-DATA) – private mental entities which actually possess the traditional sensory qualities represented by the experiences of which they are the objects. But the very idea of an essentially private entity is suspect. Moreover, since an experience may apparently represent something as having a determinable property (e.g. redness) without representing it as having any subordinate determinate property (e.g. any specific shade of red), a sense-datum may actually *have* a determinable property without having any determinate property subordinate to it. Even more disturbing is that sense-data may have contradictory properties, since experiences can have contradictory contents. A case in point is the waterfall illusion: if you stare at a waterfall for a minute and then immediately fixate on a nearby rock, you are likely to have an experience of the rock’s moving upwards while it remains in exactly the same

place. The sense-datum theorist must either deny that there are such experiences or admit contradictory objects.

These problems can be avoided by treating objects of experience as *properties*. This, however, fails to do justice to the appearances, for experience seems not to present us with bare properties (however complex), but with properties embodied in individuals. The view that objects of experience are Meinongian objects accommodates this point. It is also attractive in so far as (1) it allows experiences to represent properties other than traditional sensory qualities, and (2) it allows for the identification of objects of experience and objects of perception in the case of experiences which constitute perceptions (*see* ACT/OBJECT ANALYSIS). But most philosophers will feel that the Meinongian’s acceptance of impossible objects is too high a price to pay for these benefits.

A general problem for the act/object analysis is that the question of whether two subjects are experiencing one and the same thing (as opposed to having exactly similar experiences) appears to have an answer only on the assumption that the experiences concerned are perceptions with material objects. But in terms of the act/object analysis the question must have an answer even when this condition is not satisfied. (The answer is always negative on the sense-datum theory; it could be positive on other versions of the act/object analysis, depending on the facts of the case.)

In view of the above problems, the case for the act/object analysis should be reassessed. The phenomenological argument is not, on reflection, convincing, for it is easy enough to grant that any experience appears to present us with an object without accepting that it actually does. The semantic argument is more impressive, but is none the less answerable. The seemingly relational structure of attributions of experience is a challenge dealt with below in connection with the adverbial theory. Apparent reference to and quantification over objects of experience can be handled by analysing them as reference to experiences themselves and quantification

over experiences tacitly typed according to content. (Thus "The after-image which John experienced was green" becomes "John's after-image experience was an experience of green", and "Macbeth saw something which his wife did not see" becomes "Macbeth had a visual experience which his wife did not have".)

### *Pure cognitivism*

Pure cognitivism attempts to avoid the problems facing the act/object analysis by reducing experiences to cognitive events or associated dispositions; e.g. Suzy's experience of a rough surface beneath her hand might be identified with the event of her acquiring the belief that there is a rough surface beneath her hand, or, if she does not acquire this belief, with a disposition to acquire it which has somehow been blocked.

This position has attractions. It does full justice to the cognitive contents of experience, and to the important role of experience as a source of belief acquisition. It would also help clear the way for a naturalistic theory of mind, since there seems to be some prospect of a physicalist/functionalist account of belief and other intentional states. But pure cognitivism is completely undermined by its failure to accommodate the fact that experiences have a felt character which cannot be reduced to their content (see above).

### *The adverbial theory*

The adverbial theory (see ADVERBIAL THEORY) is an attempt to undermine the act/object analysis by suggesting a semantic account of attributions of experience which does not require objects of experience. Unfortunately, the oddities of explicit adverbialisations of such statements have driven off potential supporters of the theory. Furthermore, the theory remains largely undeveloped, and attempted refutations have traded on this. It may, however, be founded on sound basic intuitions, and there is reason to believe that an effective development of the theory (which is merely hinted at below) is possible.

The relevant intuitions are (1) that when we say that someone is experiencing "an A",

or has an experience "of an A", we are using this content-expression to specify the type of thing which the experience is especially apt to fit, (2) that doing this is a matter of saying something about the experience itself (and maybe also about the normal causes of like experiences), and (3) that there is no good reason to suppose that it involves the description of an object which the experience is "of". Thus the effective role of the content-expression in a statement of experience is to modify the verb it complements, not to introduce a special type of object.

Perhaps the most important criticism of the adverbial theory is the "many property problem", according to which the theory does not have the resources to distinguish between e.g.

(1) Frank has an experience of a brown triangle

and

(2) Frank has an experience of brown and an experience of a triangle.

which is entailed by (1) but does not entail it. The act/object analysis can easily accommodate the difference between (1) and (2) by claiming that the truth of (1) requires a single object of experience which is both brown and triangular, while that of (2) allows for the possibility of two objects of experience, one brown and the other triangular. Note, however, that (1) is equivalent to

(1\*) Frank has an experience of something's being both brown and triangular,

and (2) is equivalent to

(2\*) Frank has an experience of something's being brown and an experience of something's being triangular,

and the difference between these can be explained quite simply in terms of logical scope without invoking objects of experience. The adverbialist may use this to answer the many-property problem by arguing that the

phrase "a brown triangle" in (1) does exactly the same work as the clause "something's being both brown and triangular" in (1\*). This is perfectly compatible with the view that it also has the "adverbial" function of modifying the verb "has an experience of", for it specifies the experience more narrowly just by giving a necessary condition for the satisfaction of the experience (the condition being that there is something both brown and triangular before Frank).

### *The state theory*

A final position which should be mentioned is the state theory, according to which a sense experience of an *A* is an occurrent, non-relational state of the kind which the subject would be in when perceiving an *A*. Suitably qualified, this claim is no doubt true, but its significance is subject to debate. Here it is enough to remark that the claim is compatible with both pure cognitivism and the adverbial theory, and that state theorists are probably best advised to adopt adverbialism as a means of developing their intuitions.

See also ACT/OBJECT ANALYSIS; ADVERBIAL THEORY; ARGUMENT FROM ILLUSION; DIRECT REALISM; REPRESENTATION; REPRESENTATIVE REALISM; SENSATION/COGNITION; SENSE-DATA.

### BIBLIOGRAPHY

- Anscombe, G. E. M.: "The Intentionality of Sensation: A Grammatical Feature," in *Analytical Philosophy: Second Series*, ed. R. J. Butler (Oxford: Blackwell, 1968).
- Armstrong, D. M.: *A Materialist Theory of Mind* (London: Routledge and Kegan Paul, 1968).
- Castañeda, H.-N.: "Perception, Belief, and the Structure of Physical Objects and Consciousness," *Synthese* 35 (1977), 285–351.
- Clark, R.: "Sensing, Perceiving, Thinking," in *Essays on the Philosophy of Roderick M. Chisholm*, ed. E. Sosa, *Grazer Philosophische Studien* (1981), 273–95.
- Jackson, F.: "On the Adverbial Theory of Visual Experience," *Metaphilosophy* 6 (1975), 127–35.
- Jackson, F.: *Perception: A Representative Theory* (Cambridge: Cambridge University Press, 1977).
- Leon, M.: "Character, Content, and the Ontology of Experience," *Australasian Journal of Philosophy* 65 (1987), 377–99.
- Peacocke, C.: *Sense and Content: Experience, Thought, and Their Relations* (Oxford: Clarendon Press, 1983).
- Pendlebury, M.: "Sense Experiences and Their Contents: A Defence of the Propositional View," *Inquiry* 33 (1990), 215–30.
- Sellars, W.: "The Adverbial Theory of Objects of Sensation," *Metaphilosophy* 6 (1975), 144–60.
- Swartz, R. J., ed.: *Perceiving, Sensing, and Knowing* (Berkeley, CA: University of California Press, 1965).
- Tye, M.: "The Adverbial Theory: A Defence of Sellars against Jackson," *Metaphilosophy* 6 (1975), 135–43.

MICHAEL PENDLEBURY

**explanation** Since at least the time of Aristotle philosophers have emphasized the importance of explanatory knowledge. In simplest terms, we want to know not only *what* is the case but also *why* it is. This consideration suggests that we define an explanation as an answer to a why-question. Such a definition would, however, be too broad, because some why-questions are requests for consolation (Why did *my* son have to die?) or moral justification (Why should women not be paid the same as men for the same work?). It would also be too narrow because some explanations are responses to how-questions (How does radar work?) or how-possibly-questions (How is it possible for cats always to land on their feet?).

In its most general sense, "to explain" means to make clear, to make plain, or to provide understanding. Definitions of this sort are philosophically unhelpful, for the terms used in the definiens are no less problematic than the term to be defined. Moreover, since a wide variety of things require explanation, and since many different types of explanation exist, a more complex explication is required. Our discussion will be facilitated by introducing a bit of technical terminology. The term "explanandum" is used to refer to that which is to be explained; the term "explanans" refers to that which does the explaining. The

explanans and the explanandum taken together constitute the explanation.

One common type of explanation occurs when deliberate human actions are explained in terms of conscious purposes. "Why did you go to the pharmacy yesterday?" "Because I had a headache and needed to get some aspirin." It is tacitly assumed that aspirin is an appropriate medication for headaches and that going to the pharmacy would be an efficient way of getting some. Such explanations are, of course, *teleological*, referring, as they do, to goals. The explanans is not the realization of a future goal – if the pharmacy happened to be closed for stock-taking the aspirin would not have been obtained there, but that would not invalidate the explanation. Some philosophers would say that the antecedent desire to achieve the end is what does the explaining; others might say that the explaining is done by the nature of the goal and the fact that the action promoted the chances of realizing it (e.g. Taylor, 1964). In any case it should not be automatically assumed that such explanations are causal. Philosophers differ considerably on whether these explanations are to be framed in terms of causes or reasons (see REASONS/CAUSES), and there are many differing analyses of such concepts as intention and agency. Expanding the domain beyond consciousness, Freud (see FREUD) maintained, in addition, that much human behaviour can be explained in terms of *unconscious* wishes. These Freudian explanations should probably be construed as basically causal.

Problems arise when teleological explanations are offered in other contexts. The behaviour of non-human animals is often explained in terms of purpose; e.g. the mouse ran to escape from the cat. In such cases the existence of *conscious* purpose seems dubious. The situation is still more problematic when a super-empirical purpose is invoked – e.g. the explanation of living species in terms of God's purpose, or the vitalistic explanation of biological phenomena in terms of an entelechy or vital principle. In recent years an "anthropic principle" has received attention in cosmology (see Barrow and Tipler, 1986). All such explanations

have been condemned by many philosophers as *anthropomorphic*.

The foregoing objection notwithstanding, philosophers and scientists often maintain that *functional explanations* play an important and legitimate role in various sciences such as evolutionary biology, anthropology and sociology. For example, in the famous case of the peppered moth in Liverpool, the change in colour from the light phase to the dark phase and back again to the light phase provided adaptation to a changing environment and fulfilled the *function* of reducing predation on the species. In the study of primitive societies anthropologists have maintained that various rituals (e.g. a rain dance), which may be inefficacious in bringing about their *manifest* goals (e.g. producing rain), actually fulfil the *latent* function of increasing social cohesion at a period of stress (e.g. during a drought). Philosophers who admit teleological and/or functional explanations in common sense and science often take pains to argue that such explanations can be analysed entirely in terms of efficient causes, thereby escaping the charge of anthropomorphism (see Wright, 1976); again, however, not all philosophers agree.

Mainly to avoid the incursion of unwanted theology, metaphysics, or anthropomorphism into science, many philosophers and scientists – especially during the first half of the twentieth century – held that science provides only descriptions and predictions of natural phenomena, but not explanations. Beginning in the 1930s, however, a series of influential philosophers of science – including Karl Popper (1959) Carl Hempel and Paul Oppenheim (1948), and Hempel (1965) – maintained that empirical science *can* explain natural phenomena without appealing to metaphysics or theology. It appears that this view is now accepted by the vast majority of philosophers of science, though there is sharp disagreement on the nature of scientific explanation.

The foregoing approach, developed by Hempel, Popper and others, became virtually a "received view" in the 1960s and 1970s. According to this view, to give a scientific explanation of any natural phenomenon is to show how this phenomenon can be



subsumed under a law of nature. A particular rupture in a water pipe can be explained by citing the universal law that water expands when it freezes and the fact that the temperature of the water in the pipe dropped below the freezing point. General laws, as well as particular facts, can be explained by subsumption. The law of conservation of linear momentum can be explained by derivation from Newton's second and third laws of motion. Each of these explanations is a deductive argument; the premisses constitute the explanans and the conclusion is the explanandum. The explanans contains one or more statements of universal laws and, in many cases, statements describing initial conditions. This pattern of explanation is known as the *deductive-nomological* (D-N) model. Any such argument shows that the explanandum *had to occur* given the explanans.

Many, though not all, adherents of the received view allow for explanation by subsumption under statistical laws. Hempel (1965) offers as an example the case of a man who recovered quickly from a streptococcus infection as a result of treatment with penicillin. Although not all strep infections clear up quickly under this treatment, the probability of recovery in such cases is high, and this is sufficient for legitimate explanation according to Hempel. This example conforms to the *inductive-statistical* (I-S) model. Such explanations are viewed as arguments, but they are inductive rather than deductive. In these cases the explanans confers high inductive probability on the explanandum. An explanation of a particular fact satisfying either the D-N or the I-S model is an argument to the effect that the fact in question *was to be expected* by virtue of the explanans.

The received view has been subjected to strenuous criticism by adherents of the *causal/mechanical* approach to scientific explanation (see Salmon, 1990). Many objections to the received view were engendered by the absence of causal constraints (due largely to worries about Hume's critique) on the D-N and I-S models. Beginning in the late 1950s, Michael Scriven advanced serious counterexamples to Hempel's models; he was followed in the 1960s by Wesley Salmon and in the

1970s by Peter Railton. In general, according to this view, one explains phenomena by identifying causes (a death is explained as resulting from a massive cerebral haemorrhage) or by exposing underlying mechanisms (the behaviour of a gas is explained in terms of the motions of constituent molecules).

A *unification* approach to explanation has been developed by Michael Friedman and Philip Kitcher (see Kitcher, in Kitcher and Salmon, 1989). The basic idea is that we understand our world more adequately to the extent that we can reduce the number of independent assumptions we must introduce to account for what goes on in it. Accordingly, we understand phenomena to the degree that we can fit them into a general world picture or *Weltanschauung*. In order to serve in *scientific* explanations, the world picture must be scientifically well founded.

In contrast to the foregoing views – which stress such factors as logical relations, laws of nature, and causality – a number of philosophers (e.g. Achinstein, 1983; van Fraassen, 1980, ch. 5) have urged that explanation, and *not* just scientific explanation, can be analysed entirely in pragmatic terms.

During the past half-century much philosophical attention has been focused on explanation in science and in history. Considerable controversy has surrounded the question of whether historical explanation must be scientific, or whether history requires explanations of different types. Many diverse views have been articulated; the foregoing brief survey does not exhaust the variety. (For a more comprehensive account, see Salmon, 1990.)

In everyday life we encounter many types of explanation, which appear not to raise philosophical difficulties, in addition to those already discussed. Prior to take-off a flight attendant explains how to use the safety equipment on the aeroplane. In a museum the guide explains the significance of a famous painting. A mathematics teacher explains a geometrical proof to a bewildered student. A newspaper story explains how a prisoner escaped. Additional examples come easily to mind. The main point is to remember the great variety of contexts in which explanations are sought and given.

Another item of importance to epistemology is the widely held notion that non-demonstrative inference can be characterized as inference to the best explanation (see INFERENCE TO THE BEST EXPLANATION). Given the variety of views on the nature of explanation, this popular slogan can hardly provide a useful philosophical analysis.

See also HISTORICAL KNOWLEDGE; SOCIAL SCIENCES.

#### BIBLIOGRAPHY

- Achinstein, P.: *The Nature of Explanation* (New York: Oxford University Press, 1983).
- Barrow, J. D. and Tipler, F. J.: *The Anthropic Cosmological Principle* (Oxford: Clarendon Press, 1986).
- Hempel, C. G.: *Aspects of Scientific Explanation* (New York: The Free Press, 1965).
- Hempel, C. G. and Oppenheim, P.: "Studies in the Logic of Explanation," *Philosophy of Science* 15 (1948), 135–75; reprinted, with added postscript, in Hempel (1965).
- Kitcher, P. and Salmon, W. C., eds: *Scientific Explanation*, vol. 13, *Minnesota Studies in the Philosophy of Science* (Minneapolis, MN: University of Minnesota Press, 1989).
- Popper, K. R.: *The Logic of Scientific Discovery* (New York: Basic Books, 1959).
- Railton, P.: "Probability, Explanation, and Information," *Synthese* 48 (1981), 233–56.
- Rubin, D.-H.: *Explaining Explanation* (London: Routledge, 1990).
- Salmon, W. C.: *Four Decades of Scientific Explanation* (Minneapolis, MN: University of Minnesota Press, 1990); reprinted from Kitcher and Salmon, 1989.
- Taylor, C.: *The Explanation of Behaviour* (London: Routledge and Kegan Paul, 1964).
- Van Fraassen, B. C.: *The Scientific Image* (Oxford: Clarendon Press, 1980).
- Wright, L.: *Teleological Explanations* (Berkeley, CA: University of California Press, 1976).

WESLEY C. SALMON

**external world** see PROBLEM OF THE EXTERNAL WORLD.

**externalism/internalism** The most generally accepted account of this distinction is

that a theory of justification is *internalist* if and only if it requires that all of the factors needed for a belief to be epistemically justified for a given person be *cognitively accessible* to that person, *internal* to his cognitive perspective; and *externalist*, if it allows that at least some of the justifying factors need not be thus accessible, so that they can be *external* to the believer's cognitive perspective, beyond his ken. However, epistemologists often use the distinction between internalist and externalist theories of epistemic justification without offering any very explicit explication.

The externalism/internalism distinction has been mainly applied (as above) to theories of epistemic justification. It has also been applied in a closely related way to accounts of knowledge and in a rather different way to accounts of belief and thought content. We will consider each of these applications, devoting most of our attention to the first.

The internalist requirement of cognitive accessibility can be interpreted in at least two ways: a strong version of internalism would require that the believer actually be aware of the justifying factors in order to be justified; while a weaker version would require only that he be capable of becoming aware of them by focusing his attention appropriately, but without the need for any change of position, new information, etc. Though the phrase "cognitively accessible" suggests the weak interpretation, the main intuitive motivation for internalism, viz. the idea that epistemic justification requires that the believer actually have in his cognitive possession a reason for thinking that the belief is true (see further below), would require the strong interpretation.

Perhaps the clearest example of an internalist position would be a foundationalist view (see FOUNDATIONALISM) according to which foundational beliefs pertain to immediately experienced states of mind and other beliefs are justified by standing in cognitively accessible logical or inferential relations to such foundational beliefs. Such a view could count as either a strong or a weak version of internalism, depending on whether actual awareness of the justifying elements or only the capacity to become aware of them is required. Similarly, a coherentist view (see

COHERENTISM) could also be internalist, if both the beliefs or other states with which a justificandum belief is required to cohere and the coherence relations themselves are reflectively accessible.

It should be carefully noticed that when internalism is construed in this way, it is neither necessary nor sufficient by itself for internalism that the justifying factors literally be internal mental states of the person in question. Not necessary, because on at least some views, e.g. a direct realist view of perception, something other than a mental state of the believer can be cognitively accessible; not sufficient, because there are views according to which at least some mental states need not be actual (strong version) or even possible (weak version) objects of cognitive awareness. Also, on this way of drawing the distinction, a hybrid view (like the ones discussed below), according to which some of the factors required for justification must be cognitively accessible while others need not and in general will not be, would count as an externalist view. Obviously too, a view that was externalist in relation to a strong version of internalism (by not requiring that the believer actually be aware of all justifying factors) could still be internalist in relation to a weak version (by requiring that he at least be capable of becoming aware of them).

The most prominent recent externalist views have been versions of reliabilism (*see* RELIABILISM), whose main requirement for justification is roughly that the belief be produced in a way or via a process that makes it objectively likely that the belief is true. (For discussion of the variety of specific forms that a reliabilist account can take, *see* Goldman, 1986, pp. 43–53.) What makes such a view externalist is the absence of any requirement that the person for whom the belief is justified have any sort of cognitive access to the relation of reliability in question. Lacking such access, such a person will in general have no reason for thinking that the belief is true or likely to be true, but will, on such an account, none the less be epistemically justified in accepting it. Thus such a view arguably marks a major break from the mod-

ern epistemological tradition, stemming from Descartes (*see* DESCARTES), which identifies epistemic justification with having a reason, perhaps even a conclusive reason, for thinking that the belief is true. An epistemologist working within this tradition is likely to feel that the externalist, rather than offering a competing account of the same concept of epistemic justification with which the traditional epistemologist is concerned, has simply changed the subject.

Two general lines of argument are commonly advanced in favour of justificatory externalism. The first starts from the allegedly commonsensical premiss that knowledge can be unproblematically ascribed to relatively unsophisticated adults, to young children, and even to higher animals. It is then argued that such ascriptions would be untenable on the standard internalist accounts of epistemic justification (assuming that epistemic justification is a necessary condition for knowledge), since the beliefs and inferences involved in such accounts are too complicated and sophisticated to be plausibly ascribed to such subjects. Thus only an externalist view can make sense of such commonsense ascriptions and this, on the presumption that commonsense is correct, constitutes a strong argument in favour of externalism. An internalist may respond by challenging the initial premiss, arguing that such ascriptions of knowledge are exaggerated, while perhaps at the same time claiming that the cognitive situation of at least some of the subjects in question is less restricted than the argument claims. A quite different response would be to reject the assumption that epistemic justification is a necessary condition for knowledge, perhaps by adopting an externalist account of knowledge, rather than justification, like those discussed below.

The second general line of argument for externalism points out that internalist views have conspicuously failed to provide defensible, non-sceptical solutions to the classical problems of epistemology. In striking contrast, however, such problems are in general easily solvable on an externalist view. For example, Goldman (1986, pp. 393–4) offers a one-page solution, in a footnote, of the

problem of induction. Thus if we assume both that the various relevant forms of scepticism are false and that the failure of internalist views so far is unlikely to be remedied in the future, we have good reason to think that some externalist view is true. Obviously the cogency of this argument depends on the plausibility of the two assumptions just noted. An internalist can reply, first, that it is not obvious that internalist epistemology is doomed to failure, that the explanation for the present lack of success may simply be the extreme difficulty of the problems in question. Secondly, it can be argued that most or even all of the appeal of the assumption that the various forms of scepticism are false depends essentially on the intuitive conviction that we do have reasons in our grasp for thinking that the various beliefs questioned by the sceptic are true – a conviction that the proponent of this argument must of course reject.

The main objection to externalism rests on the intuition that the basic requirement for epistemic justification is that the acceptance of the belief in question be rational or responsible in relation to the cognitive goal of truth, which seems to require in turn that the believer actually be aware of a reason for thinking that the belief is true (or at the very least, that such a reason be available to him). Since the satisfaction of an externalist condition is neither necessary nor sufficient for the existence of such a cognitively accessible reason, it is argued, externalism is mistaken as an account of epistemic justification. This general point has been elaborated by appeal to two sorts of putative intuitive counterexamples to externalism. The first of these challenges the necessity of the externalist conditions for epistemic justification by appealing to examples of belief which seem intuitively to be justified, but for which the externalist conditions are not satisfied. The standard examples of this sort are cases where beliefs are produced in some very non-standard way, e.g. by a Cartesian demon, but none the less in such a way that the subjective experience of the believer is indistinguishable from that of someone whose beliefs are produced more normally (see Foley, 1985). Cases of this

general sort can be constructed in which any of the standard externalist conditions, e.g. that the belief be a result of a reliable process, fail to be satisfied. The intuitive claim is that the believer in such a case is none the less epistemically justified, as much so as one whose belief is produced in a more normal way, and hence that externalist accounts of justification must be mistaken.

Perhaps the most interesting reply to this sort of counterexample, on behalf of reliabilism specifically, holds that reliability of a cognitive process is to be assessed in “normal” possible worlds, i.e. in possible worlds that are actually the way our world is commonsensically believed to be, rather than in the world which actually contains the belief being judged. Since the cognitive processes employed in the Cartesian demon case are, we may assume, reliable when assessed in this way, the reliabilist can agree that such beliefs are justified. The obvious further issue is whether or not there is an adequate rationale for this construal of reliabilism, so that the reply is not merely *ad hoc*. (See Goldman, 1986, pp. 107, 113.)

The second, correlative way of elaborating the general objection to justificatory externalism challenges the sufficiency of the various externalist conditions by citing cases where those conditions are satisfied, but where the believers in question seem intuitively not to be justified. Here the most widely discussed examples have to do with possible occult cognitive capacities like clairvoyance. Considering the point in application once again to reliabilism specifically, the claim is that a reliable clairvoyant who has no reason to think that he has such a cognitive power, and perhaps even good reasons to the contrary, is not rational or responsible and hence not epistemically justified in accepting the beliefs that result from his clairvoyance, despite the fact that the reliabilist condition is satisfied.

One sort of response to this latter sort of objection is to “bite the bullet” and insist that such believers are in fact justified, dismissing the seeming intuitions to the contrary as latent internalist prejudice. A more widely adopted response attempts to impose additional

conditions, usually of a roughly internalist sort, which will rule out the offending example while still stopping far short of a full internalism (*see e.g.* Goldman, 1986, pp. 111–12). But while there is little doubt that such modified versions of externalism can indeed handle particular cases well enough to avoid clear intuitive implausibility, the issue is whether there will not always be equally problematic cases that they cannot handle, and also whether there is any clear motivation for the additional requirements other than the general internalist view of justification that externalists are committed to reject.

A view in this same general vein, one that might be described as a hybrid of internalism and externalism (*see* Swain, 1981; Alston, 1989, ch. 9), holds that epistemic justification requires that there be a justificatory factor that is cognitively accessible to the believer in question (though it need not be actually grasped), thus ruling out, *e.g.*, a pure reliabilism. At the same time, however, though it must be objectively true that beliefs for which such a factor is available are likely to be true, this further fact need not be in any way grasped or cognitively accessible to the believer. In effect, of the two premisses needed to argue that a particular belief is likely to be true, one must be accessible in a way that would satisfy at least weak internalism, while the second can be (and normally will be) purely external. Here the internalist will respond that this hybrid view is of no help at all in meeting the objection that the belief is not held in the rational, responsible way that justification intuitively seems to require, for the believer in question, lacking one crucial premiss, still has no reason at all for thinking that his belief is likely to be true.

An alternative to giving an externalist account of epistemic justification, one which may be more defensible while still accommodating many of the same motivating concerns, is to give an externalist account of knowledge directly, without relying on an intermediate account of justification. Such a view will obviously have to reject the justified true belief account of knowledge, holding instead that knowledge is true belief which

satisfies the chosen externalist condition, *e.g.* is a result of a reliable process (and perhaps further conditions as well). This makes it possible for such a view to retain an internalist account of epistemic justification, though the centrality of that concept to epistemology would obviously be seriously diminished.

Such an externalist account of knowledge can accommodate the commonsense conviction that animals, young children, and unsophisticated adults possess knowledge, though not the weaker conviction (if such a conviction even exists) that such individuals are epistemically justified in their beliefs. It is also at least less vulnerable to internalist counterexamples of the sorts discussed above, since the intuitions involved there pertain more clearly to justification than to knowledge. What is uncertain is what ultimate philosophical significance the resulting conception of knowledge is supposed to have. In particular, does it have any serious bearing on traditional epistemological problems and on the deepest and most troubling versions of scepticism, which seem in fact to be primarily concerned with justification, rather than knowledge?

A rather different use of the terms “internalism” and “externalism” has to do with the issue of how the content of beliefs and thoughts is determined: according to an internalist view of content, the content of such intentional states depends only on the non-relational, internal properties of the individual’s mind or brain, and not at all on his physical and social environment; while according to an externalist view, content is significantly affected by such external factors. Here too a view that appeals to both internal and external elements is standardly classified as an externalist view.

As with justification and knowledge, the traditional view of content has been strongly internalist in character. The main argument for externalism derives from the philosophy of language, more specifically from the various phenomena pertaining to natural kind terms, indexicals, *etc.* that motivate the views that have come to be known as “direct reference” theories. Such phenomena seem at least to



show that the belief or thought content that can be properly attributed to a person is dependent on facts about his environment – e.g. whether he is on Earth or Twin Earth, what in fact he is pointing at, the classificatory criteria employed by the experts in his social group, etc. – not just on what is going on internally in his mind or brain. (See Putnam, 1975; Burge, 1979.)

An objection to externalist accounts of content is that they seem unable to do justice to our ability to know the contents of our beliefs or thoughts “from the inside”, simply by reflection. If content is dependent on external factors pertaining to the environment, then knowledge of content should depend on knowledge of these factors – which will not in general be available to the person whose belief or thought is in question.

The adoption of an externalist account of mental content would seem to support an externalist account of justification in the following way: if part or all of the content of a belief is inaccessible to the believer, then both the justifying status of other beliefs in relation to that content and the status of that content as justifying further beliefs will be similarly inaccessible, thus contravening the internalist requirement for justification. An internalist must insist that there are no justification relations of these sorts, that only internally accessible content can either be justified or justify anything else; but such a response appears lame unless it is coupled with an attempt to show that the externalist account of content is mistaken.

See also CONCEPTS; EPISTEMIC VIRTUES; ETHICS AND EPISTEMOLOGY; INFINITE REGRESS ARGUMENT; SCEPTICISM, CONTEMPORARY; and RECENT WORK ON THE INTERNALISM–EXTERNALISM CONTROVERSY in Part I.

## BIBLIOGRAPHY

- Alston, W. P.: *Epistemic Justification* (Ithaca, NY: Cornell University Press, 1989).
- Armstrong, D. M.: *Belief, Truth and Knowledge* (London: Routledge and Kegan Paul, 1968).
- BonJour, L.: *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1985), ch. 3.
- Burge, T.: “Individualism and the Mental,” *Midwest Studies in Philosophy* 4 (1979), 73–121.
- Foley, R.: “What’s Wrong with Reliabilism?,” *The Monist* 68 (1985), 188–202.
- Fumerton, R.: “The Internalism/Externalism Controversy,” in *Philosophical Perspectives* 2 (1988), 442–59.
- Goldman, A.: *Epistemology and Cognition* (Cambridge, MA: Harvard University Press, 1986).
- Luper-Foy, S.: “The Reliabilist Theory of Rational Belief,” *The Monist* 68 (1985), 203–25.
- Nozick, R.: *Philosophical Explanations* (Cambridge, MA: Harvard University Press, 1981), ch. 3.
- Pollock, J.: *Contemporary Theories of Knowledge* (Totowa, NJ: Rowman and Littlefield, 1986), ch. 4.
- Putnam, H.: “The Meaning of ‘meaning,’” in his *Mind, Language and Reality* (Cambridge: Cambridge University Press, 1975), 215–71.
- Swain, M.: *Reasons and Knowledge* (Ithaca, NY: Cornell University Press, 1981).

LAURENCE BONJOUR

# F

---

**fact/value** The distinction between facts and values has outgrown its name: it applies not only to matters of fact vs. matters of value, but also to statements that something is, vs. statements that something *ought* to be, the case. Roughly, factual statements – “is-statements” in the relevant sense – represent some state of affairs as obtaining, whereas normative statements – evaluative and deontic ones – attribute goodness to something, or ascribe, to an agent, an obligation to act. Neither distinction is merely linguistic. Specifying a book’s monetary value is making a factual statement, though it attributes a kind of value: “That is a good book” expresses a value judgement, though the term “value” is absent (nor would “valuable” be synonymous here with “good”). Similarly, “We are morally obligated to fight” superficially expresses an is statement, and “By all indications it ought to rain” makes a kind of *ought*-claim; but the former is an *ought*-statement, the latter an (*epistemic*) is-statement.

Theoretical difficulties also beset the distinction. Some have absorbed values into facts, holding that all value is instrumental, roughly: to have value is to contribute – in a *factually* analysable way – to something further which is (say) deemed desirable. Others have suffused facts with values, arguing that facts (and observations) are “theory-impregnated” and contending that values are ineliminable in theory choice. But while some philosophers doubt that fact/value distinctions can be sustained, there persists a sense of a deep difference between evaluating, or attributing an obligation and, on the other hand, saying how the world is.

Fact/value distinctions may be defended by appeal to the notion of intrinsic value: value a thing has in itself and thus independently of its consequences. Roughly, a value statement

(proper) is an ascription of intrinsic value: one to the effect that a thing is to some degree good in itself. This leaves open whether *ought*-statements are implicitly value statements; but even if they imply that something has intrinsic value – e.g. moral value – they can be independently characterized, say by appeal to rules that provide (justifying) *reasons* for action. One might also ground the fact value distinction in the attitudinal (or even motivational) component apparently implied by the making of valuational or deontic judgements; thus, “It is a good book, but that is no reason for a positive attitude towards it” and “You ought to do it, but there is no reason to” seem inadmissible, whereas substituting, respectively, “an expensive book” and “you will do it” yield permissible judgements. One might also argue that factual judgements are the kind which are in principle appraisable scientifically, and thereby anchor the distinction on the factual side. This line is plausible, but there is controversy over whether scientific procedures are “value-free” in the required way.

Philosophers differ regarding the sense, if any, in which epistemology is normative (roughly, valuational). But what precisely is at stake in this controversy is no clearer than the problematic fact/value distinction itself. Must epistemologists as such make judgements of value or epistemic responsibility? If epistemology is naturalizable, then even epistemic principles simply articulate under what conditions – say, appropriate perceptual stimulations – a belief is justified, or constitutes knowledge. Its standards of justification, then, would be like standards of, e.g., resilience for bridges. It is not obvious, however, that the appropriate standards can be established without independent judgements that, say, a certain kind of evidence is

good enough for justified belief (or knowledge). The most plausible view may be that justification is like intrinsic goodness: though it supervenes (*see* EPISTEMIC SUPERVENIENCE) on natural properties, it cannot be analysed wholly in factual statements.

*See also* NATURALIZED EPISTEMOLOGY.

## BIBLIOGRAPHY

- Dewey, J.: *Theory of Valuation* (Chicago, IL: University of Chicago Press, 1939).  
 Frankena, W. K.: *Ethics*, 2nd edn (Englewood Cliffs, NJ: Prentice-Hall, 1973).  
 Goldman, A. I.: *Epistemology and Cognition* (Cambridge, MA: Harvard University Press, 1986).  
 Hempel, C. G.: *Aspects of Scientific Explanation* (New York: Free Press, 1965).  
 Moore, G. E.: *Principia Ethica* (Cambridge: Cambridge University Press, 1903).

ROBERT AUDI

**fallibilism** In the broadest sense of the term, an anti-dogmatic intellectual stance or attitude: an openness to the possibility that one has made an error and an accompanying willingness to give a fair hearing to arguments that one's belief is incorrect (no matter what that belief happens to be about). So understood, fallibilism's central insight is that it is possible to remain open to new evidence and arguments while also reasonably treating an issue as settled for the purposes of current inquiry and action. Fallibilism, so construed, was given its most influential formulation – and its name – by C. S. Peirce, though it was advocated by earlier philosophers as well, particularly the later ancient skeptic Philo of Larissa and perhaps also Hume. Contemporary epistemologists almost universally agree in endorsing this intellectual stance; it is part of the undisputed framework within which contemporary epistemological theorizing takes place.

In some recent discussions, the term “fallibilism” stands for the thesis that human beings are fallible about everything (or just about everything) they believe (Haack,

1979). More commonly, however, “fallibilism” is used as a name for a thesis about knowledge and justification: that we can have fallible justifications for our beliefs, and that it is possible to know that something is the case even if one has only a fallible justification for believing it. Differences of formulation and emphasis aside, this thesis is widely accepted, though there is also significant dissent.

According to standard usage, a *fallible* justification is one which is compatible with the falsity of the belief in question, in that having this justification does not necessitate or entail the truth of the belief. A justification can thus be said to be fallible even if there is a law of nature to the effect that, whenever someone has a justification of that type, if the person had the belief in question, it would be true. Assuming that knowledge requires justified belief, fallibilism about knowledge can be formulated as follows:

For some P, it is possible for one to know that P even if one could have exactly the same justification for believing P when P is false.

This thesis is sometimes formulated in terms of the notion of evidence:

For some P, it is possible for one to know that P even if one's evidence for P does not necessitate or entail the truth of P.

However, this latter formulation is unnecessarily narrow, since a fallibilist could deny that knowledge requires justification by evidence.

According to another influential formulation, fallibilist views of knowledge allow that one can know that p even if one is not in a position to “eliminate” or “rule out” every possible alternative to the truth of p (Lewis, 1996). This characterization is not clearly equivalent to the previous one. If we treat “eliminate” and “rule out” as requiring a position which *entails* or *necessitates* the falsity of the possible alternative in question, then the two characterizations are arguably equivalent. However, if “eliminate” or “rule out”

just means “know not to obtain” (as Dretske (1981) proposes), then this characterization of fallibilism is not equivalent to the previous one. Fallibilists (in the previous sense) about knowledge will treat “ruling out” or “eliminating” as a fallible relation. They will consequently hold that one’s position regarding *p* can be fallible even if one is able to rule out or eliminate every possible alternative to the truth of *p*. For this reason, they can hold that knowledge requires that one be in a position to rule out or eliminate every possible alternative to the truth of one’s belief.

Some externalist or reliabilist accounts of knowledge dispense with any justification requirement altogether. However, such accounts can still be fallibilist in a closely related sense of the term. Any account of knowledge will require that the belief in question be true. It will also require the satisfaction of certain other requirements, some having to do with the person and his or her beliefs, some having to do with the surrounding circumstances. We can use the term “epistemic position” as a label for the person’s position in relation to these latter requirements. An account of knowledge can then be said to be fallibilist if it allows that one can have knowledge even if one’s epistemic position does not entail the truth of the belief in question – that is, even if satisfaction of those further requirements (over and above the truth and the belief requirements) does not necessitate or entail the truth of the belief.

These differences of formulation create some difficulty in determining whether certain theories of knowledge are fallibilist or not. In the end, however, it is relatively unimportant whether one calls a given theory “fallibilist” or “infallibilist”. What matters is that one is clear about what exactly one means in calling it one thing or the other.

Fallibilism faces a further important problem of formulation. It is a commonplace in recent discussions of a priori knowledge that we can have *fallible* justification for believing necessary truths (BonJour, 1997; Casullo, 2003). However, this commonplace cannot be accommodated within the standard characterization of fallibilism. A necessary truth

can’t be false and is (on standard accounts) entailed by everything. It consequently is not possible to believe a necessary truth on the basis of a justification which does not entail its truth.

Recent attempts to solve this problem have suggested that one has fallible knowledge that *p* if one’s belief that *p* could have failed to be knowledge – either because it could have been false or for some other reason (Hetherington, 1999; Reed, 2002). This response is not fully satisfying. Epistemologists who speak of fallible a priori knowledge have in mind the possibility that one could know something a priori even though there is, in some sense, a gap between one’s evidence or justification and the truth of the known proposition. No one has yet fully explained what it means to talk of a gap here.

#### MOTIVATIONS FOR FALLIBILISM

External world skepticism has provided a significant motivation for the widespread acceptance of fallibilism about knowledge of the external world. Since – it is said – we cannot attain an infallible position regarding most or all of what we believe about the world, many epistemologists have held that infallibilism leads to skepticism, and so they have taken fallibilism to be the only acceptable option. Some epistemologists (Nozick, 1981; Dretske, 1970, 1981) have taken skepticism to arise from the principle that knowledge is closed under known entailment (roughly: if you know *p* and you know that *p* implies *q*, then you know *q*), and so have taken denial of this principle to be necessary to a fallibilist response to skepticism. The resulting view holds that we can know things about the world without knowing that we are not dreaming, not brains in vats, etc. However, a fallibilist response to skepticism need not require denial of the closure principle. A fallibilist could accept, for instance, that in order to know anything about the world one must know that one is not dreaming; the fallibilist will simply have to explain how one can have fallible knowledge that one is not dreaming.

It has recently been argued that an infallibilist conception of knowledge does not generate external world skepticism because we can have infallible justifications for our beliefs about the world (McDowell, 1994, 1995; Neta, 2002). For instance, you might justify your belief that there is a table in front of you by claiming that you see that there is a table in front of you, which of course entails that there is a table in front of you. However, it is doubtful that this move succeeds by itself. From the viewpoint of the infallibilist skeptic, the question will simply become what justification you have for believing that you see that there is a table there, and it is doubtful that you can provide an infallible, non-question-begging justification for this belief.

Even if we can have infallible justifications for some beliefs, a second motivation for fallibilism comes from consideration of ordinary cases of knowledge. Most of our scientific knowledge, and much of our everyday knowledge, seems to arise through reasoning and inference from evidence that does not entail or necessitate the truth of our beliefs. Attempts to interpret such cases within an infallibilist framework are bound to seem forced. Similarly, in ordinary circumstances it is indisputable that one knows who is president, but it is hard to see how one could ordinarily be in an infallible position regarding such matters. (Your epistemic position usually doesn't entail or necessitate that the president didn't die in the past five minutes.) In the end, careful attention to the details of our ordinary practices of knowledge attribution provides the best basis for accepting fallibilism (Austin, 1961; Leite, 2004).

#### OBJECTIONS, PROBLEMS, AND PROSPECTS

It is sometimes thought that, on a fallibilist view, one cannot rationally assert or conclude "p", but only something like "my evidence supports p" or "so far as I know, p". However, this criticism is incorrect. The fallibilist view is that one can know the truth of a proposition on the basis of grounds that do not necessitate its truth. Consequently, even if there is a norm to the effect that one can't

properly assert p unless one knows that p, fallibilism does not imply that one cannot rationally assert or conclude "p".

It is likewise sometimes thought that a fallibilist view precludes us from being able to know that we know that p. This is because, on a fallibilist view, there will be no introspectible difference between a case in which one has knowledge and some possible case in which one does not have knowledge. However, this worry arises from a certain preconception of what would be involved in knowing that one knows. An appropriately designed fallibilist theory of knowledge will not preclude us from knowing that we know, so long as the theory is consistently applied at both the first and second order (Feldman, 1981).

The slogan "If you know p, then you can't be wrong" has been taken to raise particular problems for fallibilism (Austin, 1961). However, this slogan is susceptible to multiple interpretations. If we interpret it as "Necessarily, if one knows p, then p is true", then it simply states that knowledge requires truth, which fallibilism does not deny. If we interpret it as asserting that when one knows that p there is a tight causal, nomological, or counterfactual connection between one's belief state and the truth, then fallibilism need not deny it. If we interpret it as asserting that whenever one knows that p there must be something about one's belief state that entails or necessitates the truth of one's belief, then of course fallibilism denies it. But, so interpreted, the slogan is hardly a truism.

Some philosophers (Austin, 1961; Kaplan, 2006) have worried that fallibilism licenses apparently incoherent assertions such as "I know that p, but I might be mistaken". The worry arises as follows. Fallibilism licenses one to assert "I know that p" when one's epistemic position regarding p does not entail or necessitate p's truth. In such circumstances, one's epistemic position is compatible with possible ways in which one could be wrong about p. Recognizing this, it seems that one should be able to assert "I might be mistaken about p". Unfortunately, putting the two together results in an assertion that seemingly both claims knowledge and repudiates that very claim.



Some fallibilists have suggested that the resulting assertion's oddness can be explained away by appealing to the pragmatic (non-semantic) commitments one makes when one claims to know something (Austin, 1961; similar ideas in Rysiew, 2001). However, a fallibilist doesn't have to grant that such an assertion would ever be true or appropriately made. The term "might" in the problematic assertion is most naturally interpreted in terms of epistemic possibility. It is plausible that something is not epistemically possible for you if you know it to be false or if its falsity is obviously entailed by something you know. Consequently, a fallibilist can hold that any such assertion would be unacceptable because it is false (Stanley, 2005): to assert "I might be mistaken about p" is to assert that it is *epistemically possible* that p is false, which contradicts the claim that one knows that p. Moreover, fallibilism does not in fact require or license one to assert "I might be mistaken" whenever one knows that p on the basis of a fallible position. A well-designed fallibilist theory will allow that one is sometimes entitled to assert "I know that p". If, whenever one asserts "I know that p", one cannot (on pain of contradiction) assert "I might (epistemic possibility) be mistaken", a fallibilist theory can thus avoid any difficulties on this score. It is true that on any fallibilist view one will be entitled to assert "I know that p, but my justification/evidence/epistemic position regarding p does not entail or necessitate the truth of p". However, such an assertion is not equivalent to "I know that p, but it's epistemically possible that I'm mistaken", and any objection to it will merely reassert a prior commitment to an infallibilist view of knowledge.

A number of writers have recently argued that fallibilist accounts of knowledge are inevitably subject to Gettier-type counter-examples (Owens, 2000; Zagzebski, 1994; Neta, 2002; McDowell, 1995, 2002). According to these writers, the failing in Gettier-type cases is that the person's belief is true only by accident or luck (relative to his or her evidence, justification, or epistemic position more broadly construed). The charge, then, is that fallibilist views can't preclude cases in which

a person satisfies whatever requirements the fallibilist might impose but the person's belief is none the less true only by accident or luck: just imagine a case in which, despite the person's evidence, justification, etc., the belief is false, and then modify the case so that good fortune intervenes to render the belief true.

This argument fails. A fallibilist can avoid the objection by specifying that one doesn't have knowledge unless something like the following condition is met: if one's belief is true, its truth isn't an accident or lucky (relative to one's evidence, justification, or epistemic position). Even if a belief's truth is not entailed by one's evidence, justification, or epistemic position, it doesn't follow that its truth is an accident or lucky (relative to one's evidence, justification, or epistemic position). It is therefore possible to satisfy such a requirement, possess a fallible justification that is adequate for knowledge, and yet fail to have a true belief. But a view which incorporates some such requirement will not be vulnerable to Gettier-type counter-examples (Ryan, 1996; Howard-Snyder and Feit, 2003).

A closely related worry is that, on a fallibilist view, one's possession of a true belief will always be a matter of accident or good luck relative to one's evidence, justification, or epistemic position (McDowell, 1995, 2002). If having a belief which meets the fallibilist requirements is not inviolably connected with having a true belief, then it would seem that an element of luck intervenes in every case: either bad luck yielding a false belief (unless that bad luck is in turn counteracted by further good luck) or good luck yielding a true belief. However, this is not so. To put the point in the most general terms, the absence of bad luck is not always itself a matter of luck. In the ordinary run of things, it's not a lucky accident that my belief that my yard has moles – based, as it is, on the sorry state of my lawn – is true. However, my evidence, justification, or epistemic position more broadly construed is still plausibly held not to entail or necessitate the truth of that belief. It is a mistake to think that the fallibilist must understand the relation between one's epistemic position and the truth of one's belief as resulting from the operation of a chance mechanism.

Fallibilism can be characterized as the idea that, in some cases in which a person knows that *p*, there is something about the person and his or her circumstances which does not entail the truth of *p* but which, in combination with the truth of *p*, makes it the case that the person knows that *p*. Understood in this way, fallibilism assumes that in some cases a person's knowing that *p* is a complex state or state of affairs involving the person, his or her belief, his or her circumstances, and the fact (or true proposition) *p*. Timothy Williamson (2000) has recently challenged this conception of knowledge, arguing that knowledge that *p* is a non-analyzable factive mental state – a mental state that entails the truth of *p*. If this view is right, then formulations like the above characterization of fallibilism will fail. However, an approach like Williamson's allows for the possibility that, even though knowledge that *p* is a non-analyzable factive mental state, there are necessary conditions for being in this state. One such condition could be that one have adequate evidence for *p*, and it could be allowed that this condition can be met even if one's evidence does not entail or necessitate the truth of *p*. Consequently, even a Williamson-style approach to knowledge can be developed in a way that is reasonably regarded as fallibilist. Though such an approach has some benefits, Williamson's arguments for it are inconclusive (Leite, 2005).

Lotteries and cases of statistical reasoning pose particular problems for fallibilism. The challenge here is to explain how one can ever have knowledge through statistical reasoning though one can't, it seems, know on merely statistical grounds that one's ticket has lost in a fair lottery (Vogel, 1990; Cohen, 1988; DeRose, 1996; Hawthorne, 2004; Harman and Sherman, 2004). These problems are compounded for fallibilist views which accept epistemic closure (Vogel, 1990; Hawthorne, 2004). Solving these problems is arguably the most important challenge currently facing fallibilism's detailed development.

See also LOTTERY PARADOX; PROPOSITIONAL KNOWLEDGE; SCEPTICISM, CONTEMPORARY.

## BIBLIOGRAPHY

- Austin, J.: "Other Minds," in *Philosophical Papers*, 3rd edn, ed. J. O. Urmson and J. G. Warnock (Oxford: Oxford University Press, 1961).
- BonJour, L.: *In Defense of Pure Reason* (Cambridge: Cambridge University Press, 1997).
- Casullo, A.: *A Priori Justification* (Oxford: Oxford University Press, 2003).
- Cohen, S.: "How to Be a Fallibilist," *Philosophical Perspectives*, 2, *Epistemology* (1988), ed. J. Tomberlin (Atascadero, CA: Ridgeview).
- DeRose, K.: "Knowledge, Assertion, and Lotteries," *Australasian Journal of Philosophy* 74 (1996), 568–80.
- Dretske, F.: "The Pragmatic Dimension of Knowledge," *Philosophical Studies* 40 (1981), 363–78.
- Dretske, F.: "Epistemic Operators," *Journal of Philosophy* 67 (1970), 1007–23.
- Feldman, R.: "Fallibilism and Knowing That One Knows," *The Philosophical Review* 90 (1981), 266–82.
- Haack, S.: "Fallibilism and Necessity," *Synthese* 41 (1979), 37–63.
- Harman, G. and Sherman, B.: "Knowledge, Assumptions, Lotteries," *Philosophical Issues* 14 (2004) (a supplement to *Nous*), ed. E. Sosa and E. Villanueva (Boston, MA: Blackwell), 492–500.
- Hawthorne, J.: *Knowledge and Lotteries* (Oxford: Oxford University Press, 2004).
- Hetherington, S.: "Knowing Fallably," *Journal of Philosophy* 96 (1999), 565–87.
- Howard-Snyder, D., Howard-Snyder, F., Feit, N.: "Infallibilism and Gettier's Legacy," *Philosophy and Phenomenological Research* 66 (2003), 304–27.
- Kaplan, M.: "If You Know, You Can't Be Wrong," in *Epistemology Futures*, ed. S. Hetherington (Oxford University Press, 2006), 180–98.
- Leite, A.: "On Williamson's Arguments That Knowledge Is a Mental State," *Ratio (new series)*, 18 (2005), 165–75.
- Leite, A.: "Is Fallibility an Epistemological Shortcoming?" *The Philosophical Quarterly* 54 (2004), 232–51.
- Levi, I.: *The Enterprise of Knowledge: An Essay on Knowledge, Creedal Probability and Chance* (Cambridge, Mass.: MIT Press, 1980).
- Lewis, D.: "Elusive Knowledge," *Australasian Journal of Philosophy* 74 (1996), 649–67.
- McDowell, J.: "Knowledge and the Internal Revisited," *Philosophy and Phenomenological Research* 64 (2002), 97–105.
- McDowell, J.: "Knowledge and the Internal," *Philosophy and Phenomenological Research* 55 (1995), 877–93.

- McDowell, J.: "Knowledge by Hearsay," in *Knowing from Words: Western and Indian Philosophical Analysis of Understanding and Testimony*, ed. Bimal Krishna Matilal and Arindam Chakrabarti, Synthese Library, 230 (Dordrecht and Boston, Mass.: Kluwer, 1994), 195–224.
- Neta, R.: "S Knows That P," *Nous* 36 (2002), 663–81.
- Nozick, R.: *Philosophical Explanations* (Cambridge, MA: Harvard University Press, 1981).
- Owens, D.: *Reason without Freedom* (London: Routledge, 2000).
- Reed, B.: "How to Think about Fallibilism," *Philosophical Studies* 107 (2002), 143–57.
- Ryan, S.: "Does Warrant Entail Truth?," *Philosophy and Phenomenological Research* 56 (1996), 183–92.
- Rysiew, P.: "The Context-Sensitivity of Knowledge Attributions," *Nous* 35 (2001), 477–514.
- Stanley, J.: "Fallibilism and Concessive Knowledge Attributions," *Analysis* 65 (2005), 126–31.
- Vogel, J.: "Are There Counterexamples to the Closure Principle?," in *Doubting: Contemporary Perspectives on Skepticism*, ed. M. Roth and G. Ross (Dordrecht: Kluwer, 1990), 13–27.
- Williamson, T.: *Knowledge and Its Limits* (Oxford: Oxford University Press, 2000).
- Zagzebski, L.: "The Inescapability of Gettier Problems," *Philosophical Quarterly* 44 (1994), 65–73.

ADAM LEITE

**feminist epistemology** Feminist epistemologies are at once political and revisionist. Their focus on *practices* of knowledge construction rather than on formal, necessary and sufficient conditions for "knowledge in general" moves the question "Whose knowledge are we talking about?" to a central analytical position. The very act of posing this question challenges the governing ideals of the post-positivist empiricist epistemology that has occupied the standard-setting and monitoring position in Anglo-American philosophy. It opens the way towards showing that behind the mask of objectivity and value-neutrality received epistemology presents to the world are complex hierarchical structures of exclusion and inclusion, dominance and subjugation.

Born of the consciousness-raising practices of the 1960s, feminist enquiry has exposed gaps, cognitive dissonances, and

incongruities between women's diverse experiences and the theories, categories, and conceptual schemes that purport to explain – to *know* – the physical-social-material world, and to know women's and other Others' experiences as part of that world. Hence, by uncovering the assumptions that have allowed the epistemologies of the mainstream to establish and sustain their hegemonic authority, feminists have effected shifts in the tasks of epistemology as such. Eschewing aims of determining how ideal, abstract, interchangeable epistemic agents ought to know, feminists are producing ever more sophisticated critical-constructive analyses of how historically and materially "situated", diversely embodied knowers can and do know, and of how normative epistemic requirements can and should be fulfilled by real knowers (cf. Haraway, 1988; Alcoff, 1996).

In the 1980s and early 1990s, feminist epistemologists tended to occupy one of two main positions which coalesced out of "second-wave" feminism: feminist empiricism and feminist standpoint theory. Some accorded biology and the social sciences the centrality, if not precisely the paradigmatic status, that post-positivist epistemology accorded the physical sciences (Harding, 1986; Longino, 1990), while others drew more heavily on mainstream epistemology than on its philosophy-of-science offspring (cf. Alcoff and Potter, 1993). In view of cross-disciplinary fertilizations, and increasingly sophisticated scholarship, feminist epistemology in the early twenty-first century can no longer be mapped neatly within these categories, with some version of post-modernism – the third category in the initial taxonomy – taking care of the remainder; even though, in a significant sense, the post-modern import of the whole endeavour is apparent. Yet because these are the sites where the conceptual tools were forged that have shaped subsequent feminist knowledge projects, a sketch of empiricism and standpoint theory, and of the putative postmodern contrast, offers a useful point of entry into these issues.

Feminist empiricists have argued that, far from being as neutral and objective as their

self-presentation indicates, traditional empiricists are caught in the androcentricity of their speaking positions. Hence, the goal of feminist-informed knowledge production, and *a fortiori* of science, is to produce knowledge cleansed of androcentrism, sexism, and (latterly) of racist, classist, and other “partialities”. Feminists insist – and show – that an unabashedly value-laden yet rigorous empiricism can produce more adequate knowledge than standard methods, ignorant of their specificity and their complicity in a sex/gender system, can produce. In short, an informed political commitment can generate a better empiricism. Thus, Rae Langton (2000) demonstrates, convincingly, that when it comes to (received) knowledge “women get left out” and/or “women get hurt”: historically and still today, women are, variably, excluded from communities of accredited knowers. They are hurt because so often they are objectified in the self-fulfilling prophecies of received knowledge. In addressing these effects, separately and together, Langton shows how feminist-informed epistemology can expose and work to counter the invisible sub-texts of standard empiricist impartiality.

Longino (1990, 2002) has developed a contextual empiricism for which knowledge construction is a thoroughly social practice; hence, incorporating political commitments and values into enquiry does not require an indiscriminate tolerance of individual interests. The theory retains a rigorous respect for evidence, yet evidential reasoning is context-dependent, and data count *as* evidence only in relation to background assumptions and hypotheses. Objectivity is ensured by critical discourse to which products of enquiry are submitted, within and across diverse communities. The diversity requirement is intended to guard against an uncontested homogeneity in points of view: to unmask androcentricity and other constitutive background assumptions, even in “good” science and other enquiry, and even from an admittedly interested position, itself open to criticism. Code’s (1991, 1995) approach is residually empiricist in its realism, but it departs from canonical empiricism (see EMPIRICISM) in maintaining that subjectivity is socially constituted, and in

proposing that knowing other people affords a better model of cognitive activity than knowing medium-sized physical objects. It departs more sharply still in advocating a *negotiated* empiricism, of which epistemic deliberation is an integral component.

In the 1980s, some exponents of feminist empiricism (cf. Harding, 1986) argued that, despite the theory’s promise, its feminism undermines its empiricism. The fact that a feminist empiricist *qua* knowing subject cannot be an abstract, a-historical, disembodied and dislocated individual so flagrantly violates a basic empiricist tenet that the “empiricism” label sits uneasily with the epistemic practice. Others proposed ways of reconciling feminism and empiricism. For Nelson (1990) and Duran (1990), the empiricism of Quinean naturalized epistemology, which demands neither the stark individualism nor the theory-neutrality of classical theories, is a valuable resource for feminists. Departing from pre-occupations with determining whether knowledge is possible, naturalized epistemology (see NATURALIZED EPISTEMOLOGY) assumes that people can and do have knowledge. Drawing upon cognitive science, it abandons transcendence to examine how people actually know, individually and socially; and because it adheres to principles of scientific objectivity it offers effective tools for producing knowledge of the physical and social world. At the beginning of the twenty-first century, however, some critics (cf. Code, 2006) have contested the laboratory’s claims to count as a “natural” site of knowledge production, and are sceptical of the extent to which principles derived there can translate into everyday epistemic practices.

According to feminist-standpoint theorists (cf. Hartsock, 1983), neither orthodox nor feminist empiricists can adequately address the multiple historical and material conditions out of which people produce knowledge. Authoritative, standard-setting knowledge in Western societies has been derived from and tested against the social experiences of a limited segment of the population – white, middle-class, educated men – while women and other Others (like the proletariat of Marxist theory) are oppressed in marginal, underclass



epistemic positions. Science as practice has created an esoteric discourse to which few women gain ready access, and the hierarchical structures of more secular knowledge often work to silence or discount women's knowledge claims, in both cases explaining their limited epistemic success with "scientifically proven facts" about women's intellectual incompetence. Such "facts" may have been established by methods not explicitly designed to oppress women and other disadvantaged groups, yet their effects are none the less oppressive. Marginality, however, can be turned into an epistemic advantage for, simply in order to survive, members of the underclass have to know their oppressors and negotiate the systems that legitimate their power, often better than the oppressors know themselves (see Wylie, 2003; Collins, 1990). Moreover, as Harding has shown (1993), because standpoint theory must cast its evidential net so widely in order to explain and engage with the complexity of diverse experiences and circumstances, the standard of objectivity it works with is "stronger" than formal objectivity, traditionally conceived. In consequence, the knowledge produced from an informed feminist standpoint, grounded in women's historical-material circumstances, has an explanatory, transformative, and emancipatory potential which more modest empiricist projects cannot realize.

Its critics have objected that, because there can be no single, unified feminist perspective, standpoint theory obliterates differences and hence fails by its own standards. Others challenge its claims to epistemic privilege, arguing that its "locatedness" produces a perspective on social reality as limited as any other. Whatever the pertinence of such charges, the importance of standpoint theories' commitments to producing faithful, if often critical, analyses of women's experiences, together with their analyses of forms of oppression legitimated – indeed, "naturalized" – by hegemonic epistemic values, cannot be gainsaid.

Both feminist empiricist and standpoint theories attest to the effects of postmodern criticisms of the epistemological projects of the Enlightenment, even if their practitioners resist the label. In challenging humanistic

conceptions of subjectivity and taking the particularities of knowers into account, these epistemologies endorse the anti-essentialism of postmodernism. Like postmodernists, standpoint and empiricist feminists are (variously) critical of attempts to tell one true story, to develop a single master narrative. Yet many feminists are wary of opting for a postmodernism that would require an indifferent tolerance of any and all viewpoints, for such an approach, *ex hypothesi*, could not contest the androcentrism, ingrained sexism, or any other form of dominance or subjugation. Hence, Fricker (2000), for example, advocates a "pluralism without postmodernism": a position realist in its metaphysics, and committed in its epistemology to enabling different perspectives to gain a hearing, not merely for the sake of quasi-aesthetic self-expression, but as contributors to ongoing, and ethical, critical discursive practices.

Feminist critiques, thus, contest entrenched assumptions that "reason is alike in all men", and that knowing subjects can be represented as undifferentiated placeholders in a formal process, whose minds convert information into knowledge, leaving it untouched by its passage through them. The "whose knowledge?" question unsettles the epistemic imaginary for which knowers are self-reliant; free from the influence of locational peculiarities; immune to the workings of the power structures that shape the social orders where they are constituted as knowers – or not. Indeed, feminist exposures of striking coincidences between the traits attributed to ideal knowers in science-oriented Western societies and the norms of male psychosexual development in those same societies have made feminist epistemology possible (cf. Bordo, 1987; Keller, 1985). These coincidences, which attest to persistent historical practices of defining reason, rationality, and objectivity through the exclusion of attributes and traits commonly associated with femininity (cf. Lloyd, 1993), leave no doubt that the invisible knower in philosophical theories of knowledge is implicitly male. It is hardly surprising that the knowledge he produces is androcentric: it derives from typically male experiences.



The “whose knowledge?” question thus demands rethinking the conceptions of subjectivity mainstream epistemologists tacitly take for granted. The idealized “individual” Subject of modern epistemology (Longino, 2002) – the neutral, disinterested, detached, self-reliant spectator of the world – is exposed as an epistemic fiction, an artefact of the very epistemology whose ascendancy feminists seek to counter. He is of a piece with the autonomous moral agent and political actor in the dramas of modernity, in an overarching veneration of autonomous individualism that generates many of the impediments to feminist and other projects of socializing epistemology. His place is taken over by multiply situated and engaged knowers whose subjectivity, corporeal specificities and experiences are produced, constructed out of social–political–racial–class–ethnic–cultural–religious circumstances. When epistemologists ask how people know, how they can best negotiate their specific circumstances and the larger situations in which their lives are embedded; when abstract individualism gives way to specifically located and invested subjectivities, then the ideal of transcending the limitations of “partial perspective” to achieve a view from nowhere is drained of its explanatory potential. Knowledge construction is reconfigured as a social, communal, and power-infused activity where people must learn how to attend to “the evidence” before they can make knowledge from it, where such learning is necessarily a social – hence socially variant – activity, and where epistemic communities play a decisive part as critics and arbiters in the justification of knowledge claims and in determining whose projects gain recognition, whose knowledge counts.

Once knowledge-production and epistemology are socially-politically relocated, objectivity (see OBJECTIVITY), too, has to be reconstructed as a socially produced and mediated value, achieved as much in consequence of inter-subjective, communal criticism – the secular counterpart of peer review in the academy – as in consequence of scrupulous attention to “the evidence” (cf. Longino, 1990, 2002). Evidence counts *as* evidence within settings determined as much by social-political as by

purely observational criteria. Hence burdens of proof are redistributed *laterally* (across communities of enquirers) and not just vertically (from transcendent observer to the data). It becomes as vital to evaluate the credibility of knowledge-claimants, their critics and interlocutors as it is to know how, empirically, to verify a claim “on its own merits” (cf. Code, 1987). Indeed, feminists argue that only people in positions of power, with the resources to believe that they can transcend and control their circumstances, would imagine such detachment from materiality and circumstance as even a theoretical option. Enquiry comes out of, and is intricately with, human purposes, whether at commonsensical or esoteric, scientific levels; and no epistemic community or sub-group can be assumed, before the fact, to be disinterested or benign. Thus, the purposes that animate inquiry have to be monitored morally, politically, epistemologically, if knowledge of the world and other people is to be appropriately truth-conducive; if it is to do epistemic justice to the subjects and objects of inquiry, and to realize its emancipatory potential, locally and further afield (cf. Longino, 1990; Code, 1991).

Reliance on a purified, value-neutral, physical-science model of knowledge has allowed mainstream epistemologists to hold that every act of knowing properly so called is replicable by every other knower in identical observation conditions: replicability is virtually synonymous with objectivity. But these very assumptions have played a definitive part in theorists of knowledge, historically, deeming women’s and other traditional arts and skills “unscientific”, hence unworthy of the label “knowledge” (Alcoff and Dalmya, 1993). Analogous exclusions operate in psychology, anthropology, history and sociology, which have worked with objectivist epistemologies indifferent to the partiality and frequent imperialism that structure even their claims to neutrality (cf. Harding, 1986; Haraway, 1988; Longino, 1990; Nelson, 1990). Collaborative knowledge projects tend in this approach to count as no more than the sum of their parts, with every participant singly and separately

accountable to the evidence. Results (= knowledge) become his individual achievement, monologically presented to an objective and disinterested public, and both collaborative and ethical-political issues are rendered invisible.

But once knowledge-production is reconceived as a social practice, engaged in by embodied, gendered and historically, racially and culturally located knowers, and produced out of a division of intellectual labour, its products cannot fail to bear the marks of their makers. Rooney (2003) argues that employing a *verb-sense* of “epistemology” (as distinct from a *noun-sense*) highlights the critically reflexive perspective feminists bring to epistemology, while helping to resolve an “uneasy alliance” between feminist epistemology and naturalized epistemology. Working constructively away from the laboratory model, Code (2006) proposes “ecological naturalism” as a theory of knowledge indebted to Quinean naturalism, drawing on the methods and commitments of ecological science, literally and metaphorically (cf. Carson, 1962; Shrader-Frechette, 2001), as a model for knowledge-making. The constitutive part played by epistemic locations – physical, social, material, cultural and myriad others – in making knowledge possible is central to the argument; as is the requirement that knowledge be accountable to the particularities of place, its inhabitants, and their interconnections. Yet the constructivism implied in diverse claims that knowledge is *made*, not found, is constrained by the intransigence – indeed, the agency – of a reality that has always to be negotiated. The case is elegantly made in Karen Barad’s *agential realism*, which recognizes the agency, and hence the responsibility, of both observer and observed in generating phenomena (Barad, 2007). It avoids the essentialist/constructivist impasse that tends to paralyse feminist theory, while working past situations where cognitive options are limited by a stubborn conservatism of traditions, practices, institutions and social structures that resist wishful negation or reconstruction. Here, cognitive agents are at once significantly free in their projects of constructing knowledge, yet intra-actively

engaged with “the universe”, and always accountable for the knowledge they produce.

So long as epistemology is conceived as a project of determining necessary and sufficient conditions for justifying knowledge claims and refuting scepticism, the product of feminist enquiry will not be a feminist epistemology. Yet the vital significance of epistemology for feminist projects cannot be exaggerated; for only by achieving epistemic authority, credibility, and responsibility can women attain the emancipation feminist politics are committed to realizing. Hence feminist epistemologists work to determine criteria for adjudicating knowledge claims when foundational appeals are no longer permissible; devise methods for analysing knowledge that is at once socially constructed and bears the marks of its makers, and constrained by a world significantly independent of those makers; and develop epistemological analyses cognisant of the specificities of their subject matters (their intractability to reductive analysis) yet open to critical debate across a plurality of locations, participants, and methods.

*See also* CONTINENTAL EPISTEMOLOGY.

#### BIBLIOGRAPHY

- Alcoff, L.: *Real Knowing: New Versions of the Coherence Theory* (Ithaca, NY: Cornell University Press, 1996).
- Alcoff, L and Potter, E., eds: *Feminist Epistemologies* (New York: Routledge, 1993).
- Alcoff, L. and Dalmiya, V.: “Are ‘Old Wives’ Tales’ Justified?,” in Alcoff and Potter (eds), 1993.
- Barad, K.: *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning* (Durham, NC: Duke University Press, 2007).
- Bordo, S.: *The Flight to Objectivity* (Albany, NY: SUNY Press, 1987).
- Carson, R.: *Silent Spring* (New York and Boston, MA: Houghton Mifflin, 1962).
- Code, L.: *Ecological Thinking: The Politics of Epistemic Location* (New York: Oxford University Press, 2006).
- Code, L.: *Rhetorical Spaces: Essays on Gendered Locations* (New York: Routledge, 1995).
- Code, L.: *What Can She Know? Feminist Theory and the Construction of Knowledge* (Ithaca, NY: Cornell University Press, 1991).

- Code, L.: *Epistemic Responsibility* (Hanover, NH: University Press of New England, 1987).
- Collins, P. H.: *Black Feminist Thought* (New York: Routledge, 1990).
- Duran, J.: *Toward a Feminist Epistemology* (Savage, MD: Rowman and Littlefield, 1990).
- Fricker, M.: "Feminism in Epistemology: Pluralism without Postmodernism," in *The Cambridge Companion to Feminism in Philosophy*, ed. M. Fricker and J. Hornsby (Cambridge: Cambridge University Press, 2000).
- Haraway, D.: "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective," *Feminist Studies* 14 (1988), 575–99.
- Harding, S.: *The Science Question in Feminism* (Ithaca, NY: Cornell University Press, 1986).
- Harding, S.: "Rethinking Standpoint Epistemology: What Is 'Strong Objectivity'?", in Alcoff and Potter (eds), 1993.
- Hartsock, N.: "The Feminist Standpoint: Developing the Ground for a Specifically Feminist Historical Materialism," in *Discovering Reality: Feminist Perspectives on Epistemology, Methodology, and Philosophy of Science*, ed. S. Harding and M. B. Hintikka (Dordrecht: Reidel, 1983).
- Keller, E. F.: *Reflections on Gender and Science* (New Haven, CT: Yale University Press, 1985).
- Langton, R.: "Feminism in Epistemology: Exclusion and Objectification," in *The Cambridge Companion to Feminism in Philosophy* eds. M. Fricker and J. Hornsby (Cambridge: Cambridge University Press, 2000).
- Lloyd, G.: *The Man of Reason: "Male" and "Female" In Western Philosophy* (London: Routledge, 1984; 2nd edn 1993).
- Longino, H.: *Science as Social Knowledge* (Princeton, NJ: Princeton University Press, 1990).
- Longino, H.: *The Fate of Knowledge* (Princeton, NJ: Princeton University Press, 2002).
- Nelson, L. H.: *Who Knows: From Quine to a Feminist Empiricism* (Philadelphia, PA: Temple University Press, 1990).
- Rooney, P.: "Feminist Epistemology and Naturalized Epistemology: An Uneasy Alliance," in *Feminist Interpretations of W. V. Quine*, ed. L. H. Nelson and J. Nelson (University Park, PA: Penn State University Press, 2003).
- Shrader-Frechette, K.: "Ecology," in *A Companion to Environmental Philosophy*, ed. D. Jamieson (Oxford: Blackwell, 2001).
- Wylie, A.: "Why Standpoint Matters," in *Science and Other Cultures*, ed. R. Figueroa and S. Harding (New York: Routledge, 2003).

LORRAINE CODE

**first philosophy** The term "First Philosophy" was used by Aristotle for metaphysics or the study of being *qua* being. But when a modern philosopher such as Quine (see QUINE) describes his naturalism as "abandonment of the goal of a first philosophy", he takes this to involve the denial that our knowledge is "answerable to any supra-scientific tribunal" (1981, p. 72). The paradigm of what is thus repudiated is found in Descartes' *Meditations on First Philosophy* (see DESCARTES), which undertakes to refute scepticism and explain the legitimacy of the sciences relying only on what is absolutely certain or indubitable. A first philosophy would thus be a philosophical investigation which was prior to, and more secure than, investigations in the special sciences.

## BIBLIOGRAPHY

- Descartes, R.: *Meditations on First Philosophy* (1641), in *The Philosophical Writings of Descartes*, trans. J. Cottingham, R. Stoothoff and D. Murdoch (Cambridge: Cambridge University Press, 1985).
- Quine, W. V.: *Theories and Things* (Cambridge, MA: Harvard University Press, 1981).

CHRISTOPHER HOOKWAY

**Firth, Roderick (1917–87)** American philosopher. Firth wrote his dissertation under C. I. Lewis (see LEWIS), and taught at Harvard from 1953 until his death. He published an influential article on Ideal Observer theories in ethics, but the bulk of his work was in epistemology.

Firth, like Lewis, was a staunch defender of foundationalism (see FOUNDATIONALISM). He believed that empirical knowledge ultimately rests on self-warranted beliefs about sensory experience, and that beliefs about physical objects can be justified without appeal to any principles beyond ordinary inductive inference. His defence of phenomenalism (1950) (see PHENOMENALISM) against attacks like Chisholm's (see CHISHOLM) is subtle and cogent; it has been neglected rather than refuted. Firth contends that critics have

ignored the fact that the phenomenologists' theory of meaning entails that the *sentences* available to express *statements* about the external world will inevitably be much less numerous than the statements themselves. When a phenomenologist who initially says "Physical-object statement *p* entails sensedatum statement *s*" is led (by some version of the "argument from perceptual relativity") to say "*s* could be false and *p* true", Firth contends that the second assertion should not be interpreted as strictly inconsistent with the first. What is actually going on, Firth claims, is that the sentence "*p*" is being used to express different (though closely related) statements in the two assertions: the statement expressed by the first occurrence of "*p*" is in fact false if "*s*" is false, but the state-of-affairs which makes the statement expressed by "*p*" false is none the less more appropriately expressed by "*p*" than by "not-*p*".

#### WRITINGS

"Radical Empiricism and Perceptual Relativity," *Philosophical Review* 59 (1950), 164–83, 319–31.

"Coherence, Certainty, and Epistemic Priority," *Journal of Philosophy* 61 (1964), 545–57.

"The Anatomy of Certainty," *Philosophical Review* 76 (1967), 3–27.

"Are Epistemic Concepts Reducible to Ethical Concepts?," in *Values and Morals*, ed. A. Goldman and J. Kim (Dordrecht: Reidel, 1978), 215–30.

*Selected Philosophical Writings* (Cambridge, MA: Harvard University Press, forthcoming).

JOHN TROYER

**Foucault, Michel (1926–84)** French philosopher, born in Poitiers. If knowledge is the product of historically specific, contingent modes of inquiry, what effect has this had on our own knowledges? Foucault attempted to answer this question for bodies of knowledge in the human sciences, principally psychiatry, penology and sexology. Influenced by Nietzsche, Bachelard and Canguilhem, Foucault's project was to reveal the "positive unconscious of knowledge", or

those hidden but constitutive elements in the background of the process of knowing. These elements remain unexamined when the key to meaning is assumed to be the "referent" of claims or the intentions of a believing subject. Foucault proposed that at a deeper level rules for the constitution of beliefs operate to determine whether statements have meaning and can function within specific discourses. Briefly, a discourse is the conglomeration of statements, concepts, objects and practices produced by an *episteme*, which is the total set of relations or discursive regularities that set out the possibilities of meaning and truth. These rules are immanent in that they are simple regularities between the elements of a discourse, and they are contingent in that they do not arise from transcendental facts about language or human beings. Foucault's method of "archaeology" attempts to identify these internal relations between discursive elements without reference to intentions or representations.

Around 1970 Foucault introduced into this analysis a new formative element, power, and developed a new "genealogical" method to describe what he called "power/knowledge". He conceived of power as existing not in individuals but in social relations and as involved as much with domination and constraint as with the production of knowledges, pleasures and subjectivities. He formulated the dyad power/knowledge to indicate that each is always implicated in the other, in the sense that the negotiations and strategic movements of power create the open spaces where discourses can emerge, but that power is exercised *through* knowledge. Genealogy, an approach borrowed from Nietzsche, is the examination of the relationship between power and specific knowledges.

Foucault's project was political in that it was motivated towards dislodging our dogmatic attachments to present categories and concepts, by revealing their genesis in the mire of contingent conceptual transformations, historical conflict and political struggle. His exploration of the history of madness, for example, had as its prime directive not an illumination of the past but a liberation of the present, via a demonstration of the contingent



ways madness has been conceptualized and treated and the variable ways the distinction between the mad and the sane has been drawn. Given such contingencies, the operations of power are necessary to explain the emergence of all knowledge systems.

Because he often claimed that not only theories are contingent but also the criterion of distinction between true and false, Foucault's work has been called self-defeating and incapable of sustaining the political critique he himself desired. His proponents have argued that his claims were regional or specific and too modest to be able to undermine all knowledge claims, and thus capable of sustaining his own. They also argue that his historical narratives have successfully called into question many current categories of analysis, and that herein lies his political impact. His primary influence on epistemology, if he is to have any, will most likely be in the introduction of power as a salient ingredient in all conceptualizations of knowledge and truth.

See also CONTINENTAL EPISTEMOLOGY; NIETZSCHE; SOCIAL SCIENCES.

#### WRITINGS

- Language, Counter-Memory, Practices*, ed. D. Bouchard (Ithaca, NY: Cornell University Press, 1977).
- L'Archéologie du Savoir* (Paris: 1969); trans. A. Sheridan as *The Archaeology of Knowledge* (New York: Pantheon, 1972).
- Les Mots et les Choses: Une Archéologie des Sciences Humaines* (Paris: 1966); trans. A. Sheridan as *The Order of Things* (New York: Random House, 1970).
- Naissance de la Clinique: Une Archéologie du Regard Medical* (Paris: 1963); trans. A. Sheridan as *The Birth of the Clinic: An Archaeology of Medical Perception* (New York: Vintage, 1973).
- Power/Knowledge: Selected Interviews and Other Writings, 1972–1977*, ed. C. Gordon (New York: Pantheon, 1980).
- Surveiller et Punir: Naissance de la Prison* (Paris: 1975); trans. A. Sheridan as *Discipline and Punish: The Birth of the Prison* (New York: Pantheon, 1977).

LINDA ALCOFF

## foundationalism

### INITIAL ACCOUNT

Foundationalism is a view concerning the *structure* of the system of justified belief possessed by a given individual. Such a system is divided into “foundation” and “superstructure”, so related that beliefs in the latter depend on the former for their justification but not vice versa. However, the view is sometimes stated in terms of the structure of *knowledge* rather than of justified belief. If knowledge is true justified belief (plus, perhaps, some further condition), one may think of knowledge as exhibiting a foundationalist structure by virtue of the justified belief it involves. In any event, I will construe the doctrine as concerned primarily with justified belief, though I will feel free to speak of knowledge instead from time to time.

The first step towards a more explicit statement of the position is to distinguish between *mediate* (indirect) and *immediate* (direct) justification of belief. To say that a belief is *mediately* justified is to say that it is justified by some appropriate relation to other justified beliefs, e.g., by being inferred from other justified beliefs that provide adequate support for it, or, alternatively, by being *based on* adequate reasons. Thus if my reason for supposing that you are depressed is that you look listless, speak in an unaccustomedly flat tone of voice, exhibit no interest in things you are usually interested in, etc., then my belief that you are depressed is justified, if at all, by being adequately supported by my justified belief that you look listless, speak in a flat tone of voice. . . .

A belief is *immediately* justified, on the other hand, if its justification is of some other sort, e.g. if it is justified by being based on experience or if it is “self-justified”. Thus my belief that you look listless may not be based on anything else I am justified in *believing* but just on the way you look to me. And my belief that  $2 + 3 = 5$  may be justified not because I infer it from something else I justifiably believe, but simply because it seems obviously true to me.

In these terms we can put the thesis of foundationalism by saying that all *mediately* justified beliefs owe their justification



ultimately to immediately justified beliefs. To get a more detailed idea of what this amounts to it will be useful to consider the most important argument for foundationalism, the regress argument. Consider a mediately justified belief that *p* (we are using lower case letters as dummies for belief contents). It is, by hypothesis, justified by its relation to one or more other justified beliefs, *q* and *r*. Now what justifies each of these, e.g. *q*? If it too is mediately justified that is because it is related appropriately to one or more further justified beliefs, e.g. *s*. By virtue of what is *s* justified? If it is mediately justified, the same problem arises at the next stage. To avoid both circularity and an infinite regress, we are forced to suppose that in tracing back this chain we arrive at one or more immediately justified beliefs that stops the regress, since their justification does not depend on any further *justified belief*. (For more details, see INFINITE REGRESS ARGUMENT.) More accurately, since each mediately justified belief may be supported by more than one other justified belief, the general picture is that of a multiply branching tree structure, with the original belief at its base. In those terms, foundationalism can be formulated as the view that every mediately justified belief is at the origin of such a structure, at the tip of each branch of which is an immediately justified belief. (See Alston, 1989, chs 1, 2.)

#### ALTERNATIVE VERSIONS

I have been presenting foundationalism as a view concerning the structure *that is in fact exhibited* by the justified beliefs of a particular person. It has sometimes been construed in ways that deviate from each of the italicized phrases in the previous sentence. Thus it is sometimes taken to characterize the structure of “our knowledge” or “scientific knowledge”, rather than the structure of the cognitive system of an individual subject. As for the other phrase, foundationalism is sometimes thought of as concerned with how knowledge (justified belief) is acquired or built up, rather than with the structure of what a person finds herself with at a certain

point. Thus some people think of scientific inquiry as starting with the recording of observations (immediately justified observational beliefs), and then inductively inferring generalizations. Again, foundationalism is sometimes thought of not as a description of the finished product or of the mode of acquisition, but rather as a proposal for how the system could be *reconstructed*, an indication of how it *could* all be built up from immediately justified foundations. This last would seem to be the kind of foundationalism we find in Descartes (see DESCARTES). However foundationalism is most usually thought of in contemporary Anglo-American epistemology as an account of the structure actually exhibited by an individual’s system of justified belief.

It should also be noted that the term is used with a deplorable looseness in contemporary literary circles, and even in certain corners of the philosophical world, to refer to anything from realism – the view that reality has a definite constitution regardless of how we think of it or what we believe about it – to various kinds of “absolutism” in ethics, politics, or wherever, and even to the truism that truth is stable (if a proposition is true, it stays true!). This essay will continue to focus on foundationalism as it was explained above.

#### VIEWS CONCERNING THE FOUNDATIONS

Since foundationalism holds that all mediate justification rests on immediately justified beliefs, we may divide variations in forms of the view into those that have to do with the immediately justified beliefs, the “foundations”, and those that have to do with the modes of derivation of other beliefs from these, how the “superstructure” is built up. The most obvious variation of the first sort has to do with what modes of immediate justification are recognized. Many treatments, both pro and con, are parochially restricted to one form of immediate justification – self-evidence, self-justification (self-warrant), justification by a direct awareness of what the belief is about, or whatever. It is then unwarrantedly assumed by critics that disposing of that one form will dispose of foundationalism generally

(Alston, 1989, ch. 3). The emphasis historically has been on beliefs that simply “record” what is directly given in experience (Lewis, 1946) and on self-evident propositions (Descartes’ “clear and distinct perceptions” and Locke’s “perception of the agreement and disagreement of ideas”). But self-warrant has also recently received a great deal of attention (Alston 1989, ch. 11; Chisholm, 1977, ch. 2), and there is also a reliabilist version (see RELIABILISM), according to which a belief can be immediately justified just by being acquired by a reliable belief-forming process that does not take other beliefs as inputs (BonJour, 1985, ch. 3).

Foundationalisms also differ as to what further constraints, if any, are put on foundations. Historically it has been common to require of the foundations of knowledge that they exhibit certain “epistemic immunities”, as we might put it, immunity from error (see INFALLIBILITY), refutation (see INCORRIGIBILITY), or doubt (see INDUBITABILITY). Thus Descartes, along with many other seventeenth- and eighteenth-century philosophers, took it that any knowledge worthy of the name would be based on cognitions the truth of which is guaranteed (infallible), that were maximally stable, immune from ever being shown to be mistaken (incorrigible), and concerning which no reasonable doubt could be raised (indubitable). Hence the search in the *Meditations* for a divine guarantee of our faculty of rational intuition. Criticisms of foundationalism have often been directed at these constraints (Lehrer, 1974; Will, 1974; both responded to in Alston, 1989, ch. 2). It is important to realize that a position that is foundationalist in a distinctive sense can be formulated without imposing any such requirements on foundations. (See the next section.)

As for the second dimension of variation, modes of derivation of the superstructure from the foundations, Descartes, along with Locke and many other early modern philosophers, took the hard line that deductive inference from the foundations is the only way to get mediate knowledge in a strict sense, for nothing else would *guarantee* that truth is preserved in the process. Locke was also

interested in weaker forms of support, in his account of Judgment and Probability in Bk IV of the *Essay*, but not much progress was made toward a systematic account (see INTUITION AND DEDUCTION). In our century foundationalists have widely recognized that the Cartesian demands are too limiting, but there is little consensus on what to put in their place. There is general agreement that inductive and probabilistic inference are required, and inference to the best explanation (see INFERENCE TO THE BEST EXPLANATION) has been popular of late (see e.g. Moser, 1989), but the discussions have tended to bog down in numerous difficulties concerning these modes of inference. Chisholm (1977, ch. 4) (see CHISHOLM) is well known for attempting to cut the Gordian knot by boldly saying that we have to countenance whatever modes of derivation enable us to get from what direct knowledge we have to whatever else it is we are justified in believing.

#### TYPES OF FOUNDATIONALISM

There are various ways of distinguishing types of foundationalist epistemology by the use of the variations we have been enumerating. Plantinga (1983) has put forward an influential conception of “classical foundationalism”, specified in terms of limitations on the foundations. He construes this as a disjunction of “ancient and medieval foundationalism”, which takes foundations to comprise what is self-evident and “evident to the senses”, and “modern foundationalism” that replaces “evident to the senses” with “incorrigible”, which in practice was taken to apply only to beliefs about one’s present states of consciousness. Plantinga himself developed this notion in the context of arguing that items outside this territory, in particular certain beliefs about God, could also be immediately justified. A popular recent distinction is between what is variously called “strong” or “extreme” foundationalism and “moderate”, “modest” or “minimal” foundationalism, with the distinction depending on whether various epistemic immunities are required of foundations. Finally I have distinguished

“simple” and “iterative” foundationalism (Alston, 1989, ch. 1), depending on whether it is required of a foundation only that it be immediately justified, or whether it is also required that the higher level belief that *the former belief is immediately justified* is itself immediately justified. In the essay just referred to I suggest that the plausibility of the stronger requirement stems from a “level confusion” between beliefs on different levels.

The classic opposition is between foundationalism and coherentism (see COHERENTISM). Coherentism denies any immediate justification. It deals with the regress argument by rejecting “linear” chains of justification and, in effect, taking the total system of belief to be epistemically primary. A particular belief is justified to the extent that it is integrated into a coherent system of belief. More recently, pragmatists (see PRAGMATISM) like John Dewey (see DEWEY) have developed a position known as contextualism, which avoids ascribing any overall structure to knowledge. Questions concerning justification can only arise in particular contexts, defined in terms of assumptions that are simply taken for granted, though they can be questioned in other contexts, where other assumptions will be privileged.

#### CRITICISMS OF FOUNDATIONALISM

Foundationalism can be attacked both in its commitment to immediate justification and in its claim that all mediated justified beliefs ultimately depend on the former. Though, in my opinion, it is the latter that is the position’s weakest point, most of the critical fire has been directed to the former. As pointed out above, much of this criticism has been directed against some particular form of immediate justification, ignoring the possibility of other forms. Thus much anti-foundationalist artillery has been directed at the “myth of the given”, the idea that facts or things are “given” to consciousness in a preconceptual, pre-judgemental mode, and that beliefs can be justified on that basis (Sellars, 1963). The most prominent general argument against immediate justification is a “level ascent”

argument, according to which whatever is taken to immediately justify a belief can only do so if the subject is justified in supposing that the putative justifier has what it takes to do so. Hence, since the justification of the original belief depends on the justification of the higher level belief just specified, the justification is not immediate after all (BonJour, 1985, ch. 2). In my view, we lack adequate support for any such higher level requirement for justification; and if it were imposed we would be launched on an infinite regress, for a similar requirement would hold equally for the higher level belief that the original justifier was efficacious.

See also INFINITE REGRESS ARGUMENT; PERCEPTUAL KNOWLEDGE; PRIVATE LANGUAGE ARGUMENT; SELLARS; and FOUNDATIONS AND COHERENCE in Part I.

#### BIBLIOGRAPHY

- Alston, W. P.: *Epistemic Justification* (Ithaca, NY: Cornell University Press, 1989).
- BonJour, L.: *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1985).
- Chisholm, R. M.: *Theory of Knowledge*, 2nd edn (Englewood Cliffs, NJ: Prentice-Hall, 1977).
- Lehrer, K.: *Knowledge* (New York: Oxford University Press, 1974).
- Lewis, C. I.: *An Analysis of Knowledge and Valuation* (La Salle, IL: Open Court, 1946).
- Moser, P.: *Knowledge and Evidence* (New York: Cambridge University Press, 1989).
- Plantinga, A.: “Reason and Belief in God,” in *Faith and Rationality*, ed. A. Plantinga and N. Wolterstorff (Notre Dame, IN: University of Notre Dame Press, 1983).
- Sellars, W.: “Empiricism and the Philosophy of Mind,” in his *Science, Perception, and Reality* (London: Routledge and Kegan Paul, 1963).
- Will, F.: *Induction and Justification* (Ithaca, NY: Cornell University Press, 1974).

WILLIAM P. ALSTON

**Frege, Gottlob (1848–1925)** German philosopher and mathematician. Frege, professor of mathematics at the University of Jena, devoted his intellectual life to a single,

narrowly circumscribed project: the reduction of arithmetic to pure logic (*see* LOGICISM). In the attempt to implement this programme, however, Frege found himself increasingly forced to confront the broader philosophical problems which it raised: problems concerning the nature of logic, language, meaning and the mind. As a result he formulated theories concerning identity, truth, validity, existence, sense, reference, generality, logical form, number, objects, concepts, thoughts and judgements. His importance extends, therefore, far beyond the narrow confines of the foundations of arithmetic: his influence on Carnap, Russell and Wittgenstein was formative, for example, and his significance within contemporary Anglo-American philosophy is second to none. He is now widely viewed as the father of analytic philosophy.

Broadly speaking, Frege's significance for epistemology is twofold. On the one hand, the revolution that he inaugurated inevitably assigns a less central role to epistemological considerations than that typically assigned to them in post-Cartesian philosophy. As Dummett has emphasized, one of the hallmarks of post-Fregean, "analytic" philosophy is that questions concerning the nature of our knowledge or the justification of our beliefs are displaced from the centre of the philosophical stage by questions concerning meaning and understanding (*see* Dummett, 1973, pp. 665–70). For Frege and those who follow him, in other words, one must first ask how it is possible even to mean, or say, or grasp the thought that *p*; and only when an adequate answer to this fundamental question has been formulated is one justified in turning to such intrinsically subsidiary questions as whether we genuinely know that *p*, whether our belief that *p* is justified, and so forth. From within this perspective, then, issues concerning the nature of thoughts – their structure, their truth-value, their expressibility in language and their intelligibility – are prior to and independent of issues in epistemology. More radically, indeed, they are also prior to and independent of issues in psychology and the philosophy of mind; for although thinking and judging are mental acts, there is, in Frege's view, nothing subjective or psy-

chological about *what* we think and judge (*see* PSYCHOLOGISM).

On the other hand, Frege made a number of contributions within epistemology, most notably via the analysis he provided of our cognitive access to abstract objects in general, and to numbers in particular. Abstract objects are self-subsistent entities that possess no spatio-temporal properties and are devoid of causal powers. They must, it seems, be imperceptible; and if so, then the problem immediately arises as to how we can have any contact with, or knowledge of them.

Frege's solution was this. Our knowledge of abstract objects is not concerned with "objects which we come to know as something alien from without through the medium of the senses, but with objects given directly to our reason and, as its nearest kin, utterly transparent to it" (Frege, 1884, p. 115). The most fundamental factor in our knowledge of any object, whether abstract or concrete, is our grasp of its identity condition – our knowledge of which object it is. In the case of abstract objects, then, we must provide identity conditions which, while accessible to reason, do not depend upon any sensory awareness of those objects. Accordingly, Frege introduces a procedure of abstraction which is designed to transform an equivalence relation (*i.e.* one that is transitive, reflexive and symmetrical) holding between empirical entities, into a strict identity holding between abstract objects (*see* Frege, 1884, p. 74). If *R* is an equivalence relation, and if *a* and *b* are empirical objects, then we can transform the statement "*aRb*" into one of the form "*f(a) = f(b)*", where the terms "*f(a)*" and "*f(b)*" will refer to abstract objects. For example, if line *a* is parallel to line *b*, then we are entitled to say: the direction of *a* is identical with the direction of *b*. Or again, if there are as many red things on the table as there are blue, then we can say: the number of red things is identical with the number of blue things. These identity statements specify identity conditions for abstract objects (directions and numbers, respectively); and these statements are intelligible because they make the same claim as the statements from which they are abstracted, and which mention only perceptible objects and the

empirically ascertainable relations that hold between them. In this way, according to Frege, we can begin to have knowledge of abstract objects in spite of the fact that we can have no intuitions or perceptions of them.

See also ANALYTICITY; INTUITION AND DEDUCTION.

#### WRITINGS

*Die Grundlagen der Arithmetik* (Breslau: 1884); trans. J. L. Austin as *The Foundations of Arithmetic* (Oxford: Blackwell, 1974).

“Über Sinn und Bedeutung,” *Zeitschrift für Philosophie und philosophische Kritik* 100 (1892), 25–50; trans. P. T. Geach and M. Black as “On Sense and Reference” in *Translations from the Philosophical Writings of Gottlob Frege* (Oxford: Blackwell, 1960), 56–78.

“Der Gedanke. Eine logische Untersuchung,” *Beiträge zur Philosophie des deutschen Idealismus* 1 (1918), 58–77; trans. P. T. Geach and R. H. Stoothoff as “Thoughts” in *Logical Investigations* (Oxford: Blackwell, 1977), 1–30.

#### BIBLIOGRAPHY

Bell, D.: *Frege's Theory of Judgement* (Oxford: Clarendon Press, 1979).

Dummett, M. A. E.: *Frege: Philosophy of Language* (London: Duckworth, 1973).

Wright, C.: *Frege's Conception of Numbers as Objects* (Aberdeen: Aberdeen University Press, 1983).

DAVID BELL

**Freud, Sigmund (1856–1939)** Austrian founder of psychoanalysis, who worked in Vienna until the Nazi occupation of 1938; he died in London. Psychoanalysis can be linked with the Kantian tradition in epistemology, as indicating how hypotheses about mental representation cast light on general features and limits of knowledge. Freud held that the rationally connected motives of which we are aware (including beliefs and desires) are underlain and shaped by others, apparently formed in infancy, which are pre-rational, only proto-realistic, and concerned with

basic bodily processes. These motives cause *wish-fulfilments*, that is, representations of situations in which they are satisfied; and their operation is traceable in terms of *unconscious phantasy* (cf. Segal, 1987, ch. 2; and for italicized terms generally, see Laplanche and Pontalis, 1987).

For example, some common epistemic phantasies can be illustrated by the following dream. A teacher had been greatly surprised when a devoted pupil – who had made a special effort to be taught by him, and was trying hard to master his ideas – suddenly revealed a desire to suck his penis. That night he dreamt that *a lamb had come to suck milk from his finger*. . . . It was clear to him that the suckling in his dream represented his teaching, the lamb his pupil who had come to imbibe his ideas, and his milk-giving finger the penis this pupil had wanted to suck.

The dream can be seen as representing the fulfilment both of a specific sexual wish, and also of phantasies in which the dreamer compared his teaching with bodily activities, both sexual and nurturing, and his ideas with semen and milk. In particular, in this context he represented his penis as fulfilling the role of a feeding breast. This was a phantasy related to his rivalry with women; and it was plainly continuous with many of his conscious beliefs, including a conviction he often emphasized, that women were inferior because less creative than men.

At a personal level such phantasies are maintained not because they correspond with reality but because they obscure it; and they in turn sustain a variety of other motives which share this function. Phantasies of this kind can implement *projection*, since they not only wishfully satisfy, but also enable one to represent oneself as lacking and others as having aspects and feelings which are painful to contemplate, envious, destructive, and so forth. Projective phantasies seem to play a particularly significant role in development, and hence in the unconscious organization of adult psychic and social life. (Thus the activities of teacher and pupil above were apparently based on phantasies about the phallic transmission of knowledge, which could also serve to project feelings of lack, envy, etc.



into women, as well as on their conscious beliefs.)

The role of projection thus contrasts with the traditional idea that we construct our representation of other minds (*see* OTHER MINDS) by analogy with what we know of ourselves through introspection (*see* ARGUMENT FROM ANALOGY). Rather, a main part of our construction consists in systematically attributing to others motives which we will *not* recognize in ourselves. Such projection, however, can both subserve social co-ordination and attain a measure of accuracy. This is sometimes fairly clearly exemplified in the attitudes of rival groups, which achieve internal cohesion on the basis of similar projective phantasies, and are hence disposed to mutual distrust. Where their projections mirror one another closely, each group may construct a very full picture of the malignancy of the other, while thereby remaining ignorant as to its own. (Hence, paradoxically, it is often only when such motives have been sufficiently disowned that they can be acted on, with the full ferocity justified by imaginary innocence and purity.)

Phantasy and belief are mutually sustaining, so that beliefs are apt to serve purposes of misrepresentation even when true, and to be maintained for this role alone when false. Hence the study of phantasy, and the avoidance of error based on it, seems a seri-

ous epistemological concern. More generally, psychoanalysis suggests that our picture of the world, like the language we speak, is in good part that which an elaborate co-ordination of psychological mechanisms (including projection) requires us to construct. Hence the nature and working of our representations are best understood in light of the mechanisms underlying them; and such accuracy as they do achieve is often inexplicable, except as the result of a co-ordination of representations and objects established by evolution.

#### WRITINGS

*Standard Edition of the Complete Psychological Works of Sigmund Freud* (London: Hogarth Press, 1955).

#### BIBLIOGRAPHY

- Laplanche, J. and Pontalis, J.-B.: *The Language of Psychoanalysis* (London: Hogarth Press, 1987).  
 Neu, J.: *The Cambridge Companion to Freud* (Cambridge: Cambridge University Press, 1991).  
 Segal, H.: *Introduction to the Work of Melanie Klein* (London: Hogarth Press, 1987).  
 Wollheim, R.: *Freud* (London: Fontana, 1971); a philosophically informed account of Freud's work.

JIM HOPKINS

# G

---

## **Gadamer, Hans-Georg (1900–2002)**

German philosopher. Hermeneutics traditionally was a theory of the interpretation of meaning principally concerned with biblical exegesis. Gadamer's project has been to develop a more generalized philosophical hermeneutics (*see* HERMENEUTICS) which can give an account of the conditions of all interpretation. And to the extent that all understanding involves an act of interpretation, this account will apply to understanding in general. Gadamer takes as his point of departure a recognition of the embeddedness of all inquiry in language and in history. Knowers operate within what he calls a horizon which includes their prejudices or pre-judgements, explicit and implicit conceptual commitments, and their situated frame of reference, and is both the enabling condition for and the constraint upon knowledge. Given this, the meaning of both reason and truth must be rethought. Reason cannot be counterposed to prejudice or the authority of tradition if we are incapable of transcending tradition (in part because we are incapable of transcending linguisticity) and if we can engage in knowing only within the confines of tradition's historical trajectory (even while this trajectory is constantly being transformed). And truth cannot be the appropriation of external, timeless facts when it only emerges in dialogic encounters between specific elements, each of which has a horizon that contributes to its formation. On Gadamer's view, then, the Cartesian ontology of the knowing process presents an objectivist picture (*see* OBJECTIVITY) which is most vulnerable to the arbitrariness of authority, since it assumes a degree of control and isolation on the part of the subject that simply does not obtain. An acknowledgement of our historical and linguistic embeddedness will require a new

ontology of knowing and a new account of truth, but will improve our knowledge at least about the nature of knowledge itself.

*See also* CONTINENTAL EPISTEMOLOGY; HEGEL; SOCIAL SCIENCES.

## WRITINGS

*Kleine Schriften*, 3 vols (Tübingen: J. C. B. Mohr, 1967); partially trans. D. Linge as *Philosophical Hermeneutics* (Berkeley, CA: University of California Press, 1976).

*Wahrheit und Methods* (Tübingen: J. C. B. Mohr, 1960); *Truth and Method* (New York: Seabury Press, 1975).

LINDA ALCOFF

**genetic epistemology** The study of the epistemic transitions from less adequate epistemic states to more advanced ones. Construed even more broadly as a theory of the development, evolution, genesis or history of knowledge, one can find important contributions to genetic epistemology by Aristotle, Descartes, Locke, Hume, Vico, Condorcet, Kant, Fichte, Hegel, Marx, Spencer, Comte, Bergson, Dewey, Cassirer, Wittgenstein, Brunschvicg, Husserl, Popper, Kuhn, Lakatos, Toulmin and a wealth of contemporary philosophers. But aside from the work of some nonphilosophers, there has been no systematic attempt to outline the major features of this theory of knowledge and to define its distinctive nature.

All epistemologies are concerned with how knowledge changes over time. But no one yet has adequately distinguished the characteristics marking off the differences between genetic epistemology, evolutionary

epistemology (see EVOLUTIONARY EPISTEMOLOGY), developmental epistemology and historical epistemology. Furthermore, it remains unclear how current philosophy of science (see NATURAL SCIENCE), which is largely committed to an historically-based approach (and hence concerned with the growth of scientific knowledge) is related to the traditional epistemology of the individual subject whose knowledge changes over time. In addition, it remains unclear how both of these branches of the theory of knowledge are related to the psychology of knowledge, the sociology of knowledge (see SOCIOLOGY OF KNOWLEDGE) and the history of knowledge. In short, as a distinct and well-defined discipline, genetic epistemology remains inchoate.

#### THE FOUNDERS OF GENETIC EPISTEMOLOGY

The first explicit concern with the creation of a field called genetic epistemology can be ascribed to the philosopher-psychologist James Mark Baldwin (1861–1934). In addition to writing several works on the psychological growth of knowledge in the individual, his genetic epistemology is contained in his four-volume *magnum opus: Thought and Things: A Study of the Development and Meaning of Thought, or Genetic Logic* (1906–15), especially vol. 1 (*Genetic Theory of Knowledge*) and vol. 3 (*Genetic Epistemology*).

Although Baldwin was perhaps the originator of “genetic epistemology”, it was surely the views of Jean Piaget (1896–1980) that have been most closely identified with this field. In fact, during his long career he wrote several major treatises explicitly devoted to characterizing the nature of genetic epistemology and its relation to philosophy and psychology; his most notable philosophical works include his 3-volume *Introduction à L'Épistémologie génétique* (1950), *Études Sociologiques* (1967), his articles in *Logique et Connaissance Scientifique* (1967) and *Insight and Illusions of Philosophy* (1971). Not surprisingly, many of the views of Baldwin can also be found in Piaget; unlike Baldwin, however, Piaget has been much more explicitly

concerned with the philosophical basis of genetic epistemology (Kitchener, 1986).

#### CONTEMPORARY GENETIC EPISTEMOLOGY

Although contemporary epistemologists and philosophers of science would not use the term genetic epistemology to characterize their views, many of them can be said to be engaged in constructing a theory of genetic epistemology. This applies to the work of individuals doing more traditional post-Gettier epistemology, to those concerned with BAYESIAN EPISTEMOLOGY “belief kinematics”, to those working in evolutionary epistemology (see EVOLUTIONARY EPISTEMOLOGY), and to those philosophers of science concerned with constructing a theory of “the growth of knowledge”. All of these contemporary philosophers are concerned, in one way or another, with describing and explaining how epistemic states (beliefs, theories, conceptual schemes, etc.) change over time. In so doing, they are tacitly concerned with constructing a genetic epistemology.

#### NATURE AND SCOPE OF GENETIC EPISTEMOLOGY

Serious conceptual issues surround the definition, nature, and scope of genetic epistemology. Perhaps no one would take issue over genetic epistemology, taken as the study of the growth of knowledge, but serious disagreements would arise over the philosophical analysis of various details of this definition and over the connection between philosophy and psychology and the relation between ontogenesis and the history of science. Thus construed, the scope of genetic epistemology is so broad as to encompass all examples of the development of knowledge. This would include: *Ontogenesis* (the growth of knowledge in the individual person from birth to maturity), *History of Science* (the growth of [scientific] knowledge in historical cultures), *The anthropology (anthropogenesis) of knowledge* (the growth of knowledge in pre-literate

['primitive']) cultures, and *The biology (phylogenesis) of knowledge* (the growth of knowledge during phylogenesis [from amoeba to humans]). It goes without saying that no one has constructed an adequate genetic epistemology to cover all these fields.

#### EPISTEMIC KINEMATICS AND DYNAMICS

By and large, the history of epistemology has been committed to the thesis that a complete description of the nature of knowledge consists of a set of logical relations between a belief (or set of beliefs) and an evidential statement (or set of evidential statements). This kind of logicism is still prevalent in twentieth-century epistemology (e.g. the traditional analysis of "S knows that *p*"; see TRIPARTITE DEFINITION OF KNOWLEDGE), and philosophy of science (e.g. Carnap's theory of confirmation, and Hempel's theory of explanation; see CARNAP; HEMPEL; EXPLANATION).

Genetic epistemology is fundamentally opposed to a logicist epistemology and, in place of it, advances a historical, historicist or dynamic theory of knowledge. Adopting the distinction in physics between "kinematics" and "dynamics", we can say that genetic epistemology includes, first, an *epistemic kinematics* and, secondly, an *epistemic dynamics*.

Any historically oriented epistemology is concerned with how knowledge changes over time. Epistemic kinematics is concerned with *describing* the succession of these epistemic states over time:  $S_1(t_1)$ ,  $S_2(t_2)$ , . . .  $S_n(t_n)$ . Each state can be construed as a stage in a larger sequence, characterized as a *stage law* of epistemic development: knowledge proceeds via stages of epistemic development, each stage possessing an underlying logical structure related to the preceding and succeeding stages in logically structured ways such that the entire sequence manifests progress in knowledge.

It is a fundamental assumption of genetic epistemology that changing epistemic states abide by certain rational (morphological) constraints, limiting the logical form any such trajectory may take. These constraints are both internal (constraints limiting its logical

form) and external (constraints due to a changing evidential or informational basis).

Epistemic dynamics is concerned with explaining epistemic stage laws by reference to rational, epistemic factors – what are called (in the history of science) internalist explanations. (Psychological dynamics, by contrast, would explain such epistemic change by mentioning purely psychological facts and hence would be providing externalist explanations.)

Virtually all individuals concerned with the dynamics of epistemic change have offered (basically) the same explanation, whether it is called equilibrium, lack of dissonance, consistency or coherence (see COHERENTISM). Most frequently it has been called 'equilibrium'. Most generally we can say:

An epistemic state is in *equilibrium* if (1) there are actual or potential internal or external disturbances and (2) there are epistemic operator transformations, such that these disturbances are compensated for and there is a return to the initial epistemic state (or goal-tending state) or such that these disturbances are averted.

(Clearly, there are degrees of equilibrium corresponding to the degree of adequacy of compensatory mechanisms.)

Equilibration can be used to explain epistemic change in the following sense: epistemic state  $S_1$  at time  $t_1$  changes into epistemic state  $S_2$  at time  $t_2$  because  $S_2$  has more equilibrium than  $S_1$  has. The same notion can be used to characterize the concept of rationality:

an epistemic state is *rational* if it is in equilibrium under all epistemic disturbances. (An epistemic state is rational to a degree corresponding to its degree of equilibrium.)

And

an epistemic stage law is *rational* if there is a progressive increase in its degree of equilibration.

Genetic epistemology is committed to providing epistemic explanations of belief revision; as such it offers internalist explanations, not externalist ones. Such explanations appear to

be inescapably normative in nature. At the same time, however, genetic epistemology is thoroughly committed to a some version of naturalistic epistemology (see NATURALIZED EPISTEMOLOGY). How to reconcile these positions is one central issue facing contemporary genetic epistemologists and other naturalistic epistemologists. A closely related issue concerns the connection between psychology and epistemology (see PSYCHOLOGY AND EPISTEMOLOGY). Genetic epistemologists are committed to the thesis that empirical psychology is *relevant* to normative epistemology although precisely how remains controversial since it involves bridging the fact/value distinction (see FACT/VALUE) in some way or reducing value to fact. All of these issues are clearly related.

## CRITICAL EVALUATION

Finally, something must be said of the current status of the programme of genetic epistemology. Most epistemologists have paid scant attention to it largely because they have mistakenly thought Piaget was merely doing child psychology, because they were unfamiliar with his major epistemological works in French, and because they held on to a strict separation of psychology and epistemology. Much of this is now changing, with the result that the programme of a genetic epistemology now seems more plausible. As mentioned earlier, this programme is very sketchy and needs further discussion, criticism and development.

Current issues that seem especially important in clarifying and developing this programme include:

1. The problematic nature of epistemic stages.
2. The tacit commitment to some kind of teleological notion of progress and to a cumulativeness condition for the growth of knowledge.
3. The relevance of empirical psychology to normative epistemology.
4. The question of the social dimensions of knowledge and the reliance on a biological theory of knowledge.

5. The conceptual adequacy of key genetic epistemological concepts such as equilibrium.
6. The empirical adequacy of the psychological theory underlying genetic epistemology and the question of whether another psychological theory, e.g., contemporary AI-based cognitive psychology, is not more adequate.

Although several individuals have raised serious objections to genetic epistemology on a variety of conceptual and empirical grounds, these objections do not seem insurmountable. Although still in its germinal stage, genetic epistemology appears to be an epistemological programme with considerable philosophical promise.

See also EVOLUTIONARY EPISTEMOLOGY.

## BIBLIOGRAPHY

- Broughton, J. M. and Freeman-Moir, D. J., eds: *The Cognitive-Developmental Psychology of James Mark Baldwin* (Norwood, NJ: Ablex, 1981).
- Hamlyn, D.: *Experience and the Growth of Understanding* (London: Routledge and Kegan Paul, 1978).
- Haroutunian, S.: *Equilibrium in the Balance: A Study of Psychological Explanation* (New York: Springer-Verlag, 1983).
- Jean Piaget Archives, ed.: *Bibliographie Jean Piaget* (Geneva: Fondation Archives Jean Piaget, 1989); a complete bibliography of Piaget's works.
- Kitchener, R. F.: "Bibliography of Philosophical Work on Piaget," *Synthese* 65 (1985), 139–52.
- Kitchener, R. F.: *Piaget's Theory of Knowledge: Genetic Epistemology and Scientific Reason* (New Haven, CT: Yale University Press, 1986).
- Mischel, T., ed.: *Cognitive Development and Epistemology* (New York: Academic, 1971).
- Rotman, B.: *Jean Piaget: Psychologist of the Real* (Ithaca, NY: Cornell University Press, 1977).
- Russell, J.: *The Acquisition of Knowledge* (New York: St Martin's Press, 1978).
- Seltman, M. and Seltman, P.: *Piaget's Logic: A Critique of Genetic Epistemology* (London: Allen and Unwin, 1985).

RICHARD F. KITCHENER



**genetic fallacy** The tactic of attacking the origin or original context of some practice, concept or argument that is instrumental to an opponent's point of view, in an attempt to discredit that point of view unfairly. For example: "The wedding ring originated from the ankle chain used by men to confine their wives; therefore, wearing a wedding ring is a sexist (bad) practice." Arguments based on origins are not always fallacious, but this kind of argument can be used as a sophistical tactic to try to transfer negative connotations in lieu of going into the real merits of an argument or practice.

See *Argumentum ad hominem* in INFORMAL FALLACIES.

#### BIBLIOGRAPHY

- Damer, T. E.: *Attacking Faulty Reasoning* (Belmont, CA: Wadsworth, 1980).  
 Pirie, M.: *The Book of the Fallacy* (London: Routledge, 1985).

DOUGLAS N. WALTON

**geometry** At the hands of the early Greeks, especially Euclid (c.300 BC), geometry achieved a degree of perfection unmatched by any other science prior to the modern period. Ancient and medieval astronomy, although extensively pursued, was little more than applied geometry, for it lacked an adequate dynamical explanation of the phenomena it studied. Indeed, not until Newton's *Principia* (1687) did any scientific body of knowledge equal Euclid's *Elements* in fertility, scope, rigour, or systematic development. For almost exactly two millennia, Euclidean geometry constituted the paradigm of systematic scientific knowledge. The import of this fact for the history of philosophy would be hard to overrate.

Geometry did not originate with Euclid; as every school-child knows, the ancient Egyptians used geometry as a practical art in building pyramids and surveying land. At some time – perhaps about 600 BC – the Greeks began to transform this art into a

rigorous deductive science. Certainly by Plato's time (c.429–347 BC) (see PLATO) it had reached a high level of sophistication, and Plato was duly impressed. According to legend, a sign at the entrance to his academy (see ACADEMY) read, "Let no one enter here who does not know geometry."

Plato realized that, strictly speaking, geometry is not *about* physical things; its subject matter consists of such abstract entities as perfect straight lines, perfect circles, and perfect triangles. In relation to his famous metaphor of the divided line, the objects of geometry belong, not to the material world of Becoming, but to the ideal world of Being. Geometry thus provides the entrance to Plato's ideal realm of the forms. Knowledge of the forms comes, not from sense experience, but from human thought. However, since physical things "participate" in the forms, geometry provides information that is useful for practical affairs. Thus, Plato offered an early version of the doctrine, later formulated by Kant, that there is such a thing as synthetic a priori knowledge (see KANT).

Euclid apparently originated the idea that all of the truths of geometry could be deduced from a small number of basic assumptions; in so doing, he invented the axiomatic method, which still pervades mathematics and logic. We do not know how Euclid regarded his postulates (whether they are called "postulates" or "axioms" is unimportant); it is clear that many of his successors through the ages took them to be self-evident truths. The picture of scientific knowledge that emerges is a body of propositions all of which are rigorously deduced from premises that are evident to reason (see AXIOMATIZATION, AXIOMATICS; RATIONALISM).

It is often said that Descartes (see DESCARTES) is the first modern philosopher; with the invention of analytic geometry, he also furnished an epoch-making contribution to mathematics. The affinity of his philosophical thought to geometry is manifest. In his work on method, the procedure he recommends is just what would be appropriate for discovering and proving geometrical propositions, but it bears hardly any resemblance to the methods of the empirical sciences. In the *Meditations*, where he builds his philosophical

system from scratch, he seeks and finds an *indubitable* proposition, the cogito (“I think, therefore, I exist”) (see COGITO), on which to found it. He appeals to principles *evident to the natural light of reason* in his a priori “proofs” of the existence of God. He emphasizes the certainty of clear and distinct ideas. All of this strongly suggests the kind of self-evident certainty that was often accorded to the axioms of geometry. His metaphysical view that *extension* is the essence of matter also reveals the deep influence of geometry upon his philosophy.

When Kant (see KANT) took up the question of synthetic a priori knowledge, he did not have to linger over the question of whether synthetic a priori knowledge (see A PRIORI/A POSTERIORI) exists. Euclidean geometry obviously filled the bill. Its propositions can be established by pure reason, and they inform us about the structure of the physical space of our universe and of the spatial relationships among the entities within it. The serious question was how such knowledge is possible. He saw Euclidean geometry as a necessary form of spatial intuition – that is, a framework for the visualization of things and events in the physical world. If scientific knowledge is to be possible at all, this framework is indispensable.

Euclid’s geometry was based on five postulates:

- P1: A straight line can be drawn between any two points.
- P2: A finite straight line can be extended continuously in a straight line.
- P3: A circle can be drawn with any centre and any radius.
- P4: All right angles are equal to one another.
- P5: Given a straight line and a point not on that line, there is one and only one line through that point parallel to the given line. (This form, known as Playfair’s axiom, is not Euclid’s fifth postulate but is equivalent to it; it has the advantage of focusing directly on the question of parallelism.)

Throughout the centuries, from antiquity to the beginning of the nineteenth century, there had been numerous attempts to prove the fifth postulate – perhaps it was not

considered quite as self-evident as the other four. At about the time of Kant’s death, in the early years of the nineteenth century, various mathematicians began to realize that the parallel postulate is not a necessary truth, and the development of non-Euclidean geometry began. In all probability, Carl Friedrich Gauss was the first to take this step, but he did not publish this work. His excuse was that he did not want to listen to the braying of the asses – i.e. the cries of outrage that would result were he to make public his denial of the parallel postulate. About twenty years later, Johann Bolyai and Nikolai Ivanovich Lobachevski developed and published systems of geometry in which Playfair’s axiom was replaced by the postulate that through the given point there is more than one parallel. Although Euclid’s first four postulates are sufficient to prove that there is at least one parallel, Georg F. B. Riemann, about the middle of the nineteenth century, showed how to construct a non-Euclidean geometry without parallels by making minor modifications of the first four postulates.

By the mid-nineteenth century, then, there existed three types of geometry: (1) no parallels, (2) one parallel, and (3) more than one parallel. The question of the logical consistency of the non-Euclidean geometries naturally arises, and it was answered by a proof of relative consistency. Although it is impossible to demonstrate the absolute consistency of any of these geometries it was shown that all three types stand or fall together. If either type of non-Euclidean geometry contains a contradiction so does Euclidean geometry.

The discovery of non-Euclidean geometries that are on a par with Euclidean geometry with respect to logical consistency constituted an intellectual revolution perhaps as profound as the Copernican Revolution. No longer could one point to a unique geometry as the only possible representation of physical space. It should be emphasized, however, that Kant’s doctrine of the synthetic a priori was *not* thereby refuted. If Euclidean geometry were the only logically consistent geometry, then it would seem to be *analytic a priori* rather than *synthetic a priori*. Kant’s view was not that Euclidean geometry is logically privileged,

but rather, that it is epistemologically privileged (as a *necessary* form of visualization).

Various philosophers addressed the epistemological status of geometry. Around the turn of the twentieth century, Henri Poincaré argued that the choice of a geometry to describe physical space has a large component of convention; he believed, however, that we will always choose Euclidean geometry to describe our world because it is simpler than the non-Euclidean ones. Later in this century, when Einstein employed non-Euclidean geometry in formulating his general theory of relativity, he advocated the rejection of Euclidean geometry on the ground that the total system including both geometry and physics is simpler, even though Euclidean geometry by itself is simpler than non-Euclidean geometry. The classic philosophical work on this topic is by Hans Reichenbach (see REICHENBACH).

Another result of the emergence of non-Euclidean geometry was a careful study of the nature of axiomatic systems. Although Euclid's work was truly remarkable it did not meet modern standards of logical rigour. David Hilbert, for example, formalized both Euclidean and non-Euclidean geometries according to much more exacting standards. Hilbert's work led to the position known as *formalism* in the philosophy of mathematics.

In the light of developments since the discovery of non-Euclidean geometries, we must distinguish between *pure* geometry and *applied* geometry. A pure geometry is simply an uninterpreted axiomatic system that, in and of itself, has no bearing on the nature of physical space. It is, however, *a priori*. An applied geometry is a description of some aspect of physical reality, but it is not *a priori*. It is, however, *synthetic*. There is no geometry that is both *synthetic* and *a priori*.

See also ANALYTICITY; MATHEMATICAL KNOWLEDGE.

#### BIBLIOGRAPHY

Cajori, F.: *Sir Isaac Newton's Mathematical Principles of Natural Philosophy and His System of the*

*World* (Berkeley, CA: University of California Press, 1947).

Friedman, M.: *Foundations of Space-Time Theories* (Princeton, NJ: Princeton University Press, 1983).

Grünbaum, A.: *Philosophic Problems of Space and Time*, 2nd edn (Dordrecht: Reidel, 1974).

Heath, Sir Thomas L.: *The Thirteen Books of Euclid's Elements* (New York: Dover Publications, 1956).

Hilbert, D.: *Foundations of Geometry*, 2nd edn (LaSalle, IL: Open Court, 1971).

Poincaré, H.: *Science and Hypothesis* (New York: Dover, 1952).

Reichenbach, H.: *The Philosophy of Space and Time* (New York: Dover, 1958).

Salmon, W.C.: *Space, Time, and Motion: A Philosophical Introduction* (Encino, CA: Dickenson, 1975).

WESLEY C. SALMON

**Gettier problem** The so-called "standard analysis" of propositional knowledge, suggested by Plato and Kant among others, implies that if one has a justified true belief that *p*, then one knows that *p* (see TRIPARTITE DEFINITION OF KNOWLEDGE). In 1963 Edmund Gettier published two counterexamples to this implication of the standard analysis. In essence, they are:

- (1) Smith and Jones have applied for the same job. Smith is justified in believing that (a) Jones will get the job, and that (b) Jones has ten coins in his pocket. On the basis of (a) and (b) Smith infers, and thus is justified in believing, that (c) the person who will get the job has ten coins in his pocket. As it turns out, Smith himself will get the job, and he also happens to have ten coins in his pocket. So, although Smith is justified in believing the true proposition (c), Smith does not know (c).
- (2) Smith is justified in believing the false proposition that (a) Jones owns a Ford. On the basis of (a) Smith infers, and thus is justified in believing, that (b) either Jones owns a Ford or Brown is in Barcelona. As it turns out, Brown is in Barcelona, and so (b) is true. So although Smith is justified in believing the true proposition (b), Smith does not know (b).

Gettier's counterexamples are thus cases where one has a justified true belief that *p*, but lacks knowledge that *p*. The *Gettier problem* is the problem of finding a modification of, or an alternative to, the standard justified-true-belief analysis of knowledge that avoids counterexamples like Gettier's. Some philosophers have suggested that Gettier-style counterexamples are defective owing to their reliance on the false principle that false propositions can justify one's belief in other propositions. But there are examples much like Gettier's that do not depend on this allegedly false principle. Here is one example inspired by Keith Lehrer and Richard Feldman:

- (3) Suppose Smith knows the following proposition, *m*: Jones, whom Smith has always found to be reliable and whom Smith has no reason to distrust now, has told Smith, his office-mate, that *p*: He, Jones, owns a Ford. Suppose also that Jones has told Smith that *p* only because of a state of hypnosis Jones is in, and that *p* is true only because, unknown to himself, Jones has won a Ford in a lottery since entering the state of hypnosis. And suppose further that Smith deduces from *m* its existential generalization, *q*: There is someone, whom Smith has always found to be reliable and whom Smith has no reason to distrust now, who has told Smith, his office-mate, that he owns a Ford. Smith, then, knows that *q*, since he has correctly deduced *q* from *m*, which he also knows. But suppose also that on the basis of his knowledge that *q*, Smith believes that *r*: Someone in the office owns a Ford. Under these conditions, Smith has justified true belief that *r*, knows his evidence for *r*, but does not know that *r*.

Gettier-style examples of this sort have proven especially difficult for attempts to analyse the concept of propositional knowledge (see PROPOSITIONAL KNOWLEDGE).

The history of attempted solutions to the Gettier problem is complex and open-ended; it has not produced consensus on any solution.

Many philosophers hold, in light of Gettier-style examples, that propositional knowledge requires a fourth condition, beyond the justification, truth and belief conditions. Although no particular fourth condition enjoys widespread endorsement, there are some prominent general proposals in circulation. One sort of proposed modification, the so-called *defeasibility analysis*, requires that the justification appropriate to knowledge be "undefeated" in the general sense that some appropriate subjunctive conditional concerning genuine defeaters of justification be true of that justification. One straightforward defeasibility fourth condition, for instance, requires of Smith's knowing that *p* that there be no true proposition *q*, such that if *q* became justified for Smith, *p* would no longer be justified for Smith (see the articles by Lehrer and Paxson, and by Swain, in Pappas and Swain, 1978). A different prominent modification requires that the actual justification for a true belief qualifying as knowledge not depend in a specified way on any falsehood (see Armstrong, 1973). The details proposed to elaborate such approaches have met with considerable controversy.

My own proposed solution to the Gettier problem relies on a fourth condition of evidential truth-sustenance. More specifically, for a person, *S*, to have knowledge that *p* on justifying evidence *e*, *e* must be truth-sustained in this sense: for every true proposition *t* that, when conjoined with *e*, undermines *S*'s justification for *p* on *e*, there is a true proposition, "*t*", that, when conjoined with *e* & *t*, restores the justification of *p* for *S* in a way that *S* is actually justified in believing that *p*. The gist of my solution, put roughly, is that propositional knowledge requires justified true belief that is sustained by the collective totality of truths. I have argued in *Knowledge and Evidence* that this approach handles not only such Gettier-style examples as (1)–(3), but various others as well.

Three features of my proposed solution merit emphasis. First, it avoids a subjunctive conditional in its fourth condition, and so escapes some difficult problems facing the use of such a conditional in an analysis of knowledge. Second, it allows for non-deductive

justifying evidence as a component of propositional knowledge. An adequacy condition on an analysis of knowledge is that it not restrict justifying evidence to relations of deductive support. Third, my proposed solution is sufficiently flexible to handle cases describable as follows:

- (4) Smith has a justified true belief that *p*, but there is a true proposition, *t*, which undermines Smith's justification for *p* when conjoined with it, and which is such that it is either physically or humanly impossible for Smith to be justified in believing that *t*.

Examples represented by (4) suggest that we should countenance varying strengths in notions of propositional knowledge. These strengths are determined by accessibility qualifications on the set of relevant knowledge-precluding underminers. A very demanding concept of knowledge assumes that it need only be logically possible for a knower to believe a knowledge-precluding underminer. Less demanding concepts assume that it must be physically or humanly possible for a knower to believe knowledge-precluding underminers. But even such less demanding concepts of knowledge need to rely on a notion of truth-sustained evidence if they are to survive a threatening range of Gettier-style examples. Given my solution, the needed fourth condition for a notion of knowledge is not a function simply of the evidence a knower actually possesses.

The highly controversial aftermath of Gettier's original counterexamples has left some philosophers doubtful of the real philosophical significance of the Gettier problem. Such doubt, however, seems misplaced. One fundamental branch of epistemology seeks understanding of the nature of propositional knowledge. And our understanding exactly what propositional knowledge is essentially involves our having a Gettier-resistant analysis of such knowledge. If our analysis is not Gettier-resistant, we will lack an exact understanding of what propositional knowledge is. It is epistemologically important, therefore, to have a defensible solution to the Gettier

problem, however demanding such a solution is.

See also CAUSAL THEORIES; COHERENTISM; PROPOSITIONAL KNOWLEDGE; RELIABILISM.

#### BIBLIOGRAPHY

- Armstrong, D. M.: *Belief, Truth, and Knowledge* (Cambridge: Cambridge University Press, 1973).  
 Chisholm, R. M.: *Theory of Knowledge*, 3rd edn (Englewood Cliffs, NJ: Prentice-Hall, 1989).  
 Conee, E.: "Why Solve the Gettier Problem?," in *Philosophical Analysis*, ed. D. Austin (Dordrecht: Kluwer, 1988), 55–8.  
 Feldman, R.: "An Alleged Defect in Gettier Counter-examples," *Australasian Journal of Philosophy* 52 (1974), 68–9.  
 Foley, R.: "A Trial Separation between the Theory of Knowledge and the Theory of Justified Belief," in *Ernest Sosa and His Critics*, ed. John Greco (Oxford: Blackwell, 2004), 59–71.  
 Gettier, E. L.: "Is Justified True Belief Knowledge?," *Analysis* 23 (1963), 121–3.  
 Lehrer, K.: *Theory of Knowledge* (Boulder, Colo.: Westview Press, 1990).  
 Lycan, W.: "On the Gettier Problem Problem," in *Epistemology Futures*, ed. Stephen Hetherington (Oxford: Oxford University Press, 2006), 148–68.  
 Moser, P. K.: *Knowledge and Evidence* (Cambridge: Cambridge University Press, 1989).  
 Moser, P. K. and van der Nat, A., eds: *Human Knowledge: Classical and Contemporary Approaches* (New York: Oxford University Press, 1987).  
 Pappas, G. S. and Swain, M., eds: *Essays on Knowledge and Justification* (Ithaca, NY: Cornell University Press, 1978).  
 Roth, M. D. and Galis, L., eds: *Knowing: Essays in the Analysis of Knowledge* (New York: Random House, 1970).  
 Shope, R. K.: *The Analysis of Knowing* (Princeton, NJ: Princeton University Press, 1983).  
 Shope, R. K.: "Conditions and Analyses of Knowing," in *The Oxford Handbook of Epistemology*, ed. Paul Moser (New York: Oxford University Press, 2002), 25–70.

PAUL K. MOSER

**given, the** The concept of the given refers to the immediate apprehension of the contents of sense experience, expressed in



first-person, present-tense reports of appearances. Apprehension of the given is seen as immediate both in a causal sense, since it lacks the usual causal chain involved in perceiving real qualities of physical objects, and in an epistemic sense, since judgements expressing it are justified independently of all other beliefs and evidence. Some proponents of the idea of the given maintain that its apprehension is absolutely certain: infallible, incorrigible and indubitable (see CERTAINTY; INFALLIBILITY; INCORRIGIBILITY; INDUBITABILITY). It has been claimed also that a subject is omniscient with regard to the given: if a property appears, then the subject knows this.

The doctrine dates back at least to Descartes, who argued in second Meditation that it was beyond all possible doubt and error that he seemed to see light, hear noise, and so on (see DESCARTES). The empiricists added the claim that the mind is passive in receiving sense impressions, so that there is no subjective contamination or distortion here (even though the states apprehended are mental). The idea was taken up in twentieth-century epistemology by C. I. Lewis (see LEWIS) and A. J. Ayer (see AYER), among others, who appealed to the given as the foundation for all empirical knowledge (see EMPIRICISM). Since beliefs expressing only the given were held to be certain and justified in themselves, they could serve as solid foundations.

There are two main arguments to the conclusion that foundations for knowledge are required, and one argument that such foundations must be certain. The first argument for foundations is the infinite regress argument (see INFINITE REGRESS ARGUMENT). It points out that beliefs justified only in relation to others are justified only if those others are also. This fact is held to create a regress that can be terminated only by beliefs justified independently of all others. The argument ignores the possibility of mutually supportive beliefs, however, and so it is not sound.

The second argument for the need for foundations is sound. It appeals to the possibility of incompatible but fully coherent systems of belief, only one of which could be completely true. In light of this possibility, coherence cannot suffice for complete justi-

fication (see COHERENTISM). Without some independent indication that some of the beliefs within a coherent system are true, coherence in itself is no indication of truth. Fairy stories can cohere. But our criteria for justification must indicate to us the probable truth of our beliefs. Hence, within any system of beliefs there must be some privileged class with which others must cohere to be justified. In the case of empirical knowledge, such privileged beliefs must represent the point of contact between subject and world: they must originate in perception. When challenged, we justify our ordinary perceptual beliefs about physical properties by appeal to beliefs about appearances. Such beliefs seem more suitable as foundations than beliefs about objective properties. In the case of beliefs about appearances, there is no class of more certain perceptual beliefs to which we appeal for their justification.

The argument that foundations must be certain was offered by Lewis (1946) (see LEWIS). He held that no propositions can be probable unless some are certain. If the probability of all propositions or beliefs were relative to evidence expressed in others, and if these relations were linear, then any regress, it seemed to Lewis, would have to terminate in propositions or beliefs that are certain. But he showed neither that such relations must be linear nor that regresses cannot terminate in beliefs that are merely probable or justified in themselves without being certain or infallible.

Arguments against the idea of the given originate with Kant (see KANT), who argues in book I of the *Transcendental Analytic* that perceptions without concepts do not yet constitute any form of knowing. Being non-epistemic, they presumably cannot serve as epistemic foundations. Once we recognize that we must apply concepts of properties to appearances and formulate beliefs utilizing those concepts before the appearances can play any epistemic role, it becomes more plausible that such beliefs are fallible. The argument was developed in this century by Wilfrid Sellars (1963) (see SELLARS). According to him, the idea of the given involves a confusion between sensing particulars (having sense impressions), which is non-epistemic, and

having non-inferential knowledge of propositions referring to appearances. The former may be necessary for acquiring perceptual knowledge, but it is not itself a primitive kind of knowing. Its being non-epistemic renders it immune from error, but also unsuitable for epistemological foundations. The latter, non-inferential perceptual knowledge, is fallible, requiring, according to Sellars, concepts acquired through trained responses to public physical objects.

Cast in the form of a dilemma, the argument is the following: either sensations or experiences are non-conceptual, in which case they cannot justify beliefs, or they are similar to beliefs in being conceptualized as having certain properties, in which case their contents themselves require justification. In neither case can they serve as foundations. Most contemporary foundationalists seize the first horn of Sellars's dilemma by taking unconceptualized perceptual experience to justify either beliefs about objective properties or beliefs about appearances, either by causing the beliefs (Haack, 1993) or by directly confronting or being compared to the beliefs (BonJour, 2001).

There is contemporary debate, however, about whether perceptual experience has non-conceptual content. Unconceptualized experiences have representational contents whose descriptions may use predicates not known to the subject. The most widely accepted argument for such content is that perceptual experience is too rich and detailed to be fully captured by concepts at any one time (Peacocke, 2001). Those who deny such content reply that demonstrative concepts can serve to capture any content (McDowell, 1994). A better argument for non-conceptual content is that, if experience is to be a source of concepts, it must appear certain ways before concepts of those properties are acquired. How could one acquire a full concept of red unless redness first appeared to one? Another argument appeals to the fact that we may conceptualize experience wrongly or inconsistently. If we can fail to apply the correct concept to an experience, the concept we apply must fail to match the way we formerly conceptualized the non-conceptual content.

But this last argument, while it implies that non-conceptual content exists, calls into question the role that contemporary foundationalists want it to play. Simply causing a belief will not justify it, since unconceptualized input might cause any sort of belief. Sellars's point remains that experience cannot provide reasons for belief until it is conceptualized correctly or consistently, until it is given justified propositional contents. One's applying a predicate to an appearance may not be explicitly comparative, but its correct application requires that one remember its prior applications. Ayer (1950) had argued earlier that grasping the meaning of a proposition expressing the given suffices for knowing the truth of that proposition. But this will be so only if the proposition uses only purely demonstrative terms or concepts, only if the current reference serves merely to pick out whatever properties now appear. But such purely demonstrative meaning, while avoiding the problem of misclassification, cannot produce foundational beliefs to which one can appeal in justifying other beliefs.

Any proponent of the given now faces a new dilemma. If the terms used in statements expressing its apprehension are purely demonstrative, then such statements (assuming they are statements) are certain, but fail to express beliefs that could serve as foundations for knowledge. If what is expressed is not awareness of genuine reinstantiable properties, then the awareness does not justify its subject in believing anything else. But if statements about what appears use genuine predicates that apply to reinstantiable properties, then beliefs expressed cannot be infallible and understanding will not suffice for truth or knowledge. Coherentists (*see* COHERENTISM) would add that such genuine beliefs stand in need of justification themselves and so cannot be foundations.

Can one, then, simply compare a belief about an appearance with the experience itself to see whether one has conceptualized the experience correctly? Can non-conceptual contents directly confront beliefs in this way? It seems that once experienced contents are conceptualized, once they are scrutinized as reasons for beliefs, one cannot return to the

preconceptual stage of perception to compare the stages. And, more importantly, even if one could regain a naked eye, this would not reveal whether the experience had been conceptualized correctly, where “correctly” means at least consistently with prior applications of the relevant concept.

Coherentists will claim that a subject requires evidence that he applies concepts consistently, that he is able, for example, consistently to distinguish red from other colours that appear. Beliefs about red appearances could not then be justified independently of other beliefs expressing that evidence. To save that part of the doctrine of the given that holds beliefs about appearances to be self-justified, we require an account of how such justification is possible, how some beliefs about appearances can be justified without appeal to evidence.

The best strategy here is to tie an account of self-justification to a broader exposition of epistemic warrant. One such account sees justification as a kind of inference to the best explanation. A belief is shown to be justified if its truth is shown to be part of the best explanation for why it is held. A belief is self-justified if the best explanation for it is its truth alone (Goldman, 1988). The best explanation for my belief that I am appeared to redly may be simply that I am. The explanation for my being able to form such a belief is that I am applying my concept of red consistently, albeit fallibly. Such accounts seek to ground knowledge in perceptual experience without appealing to an infallible given, now universally dismissed.

*See also* ARGUMENT FROM ILLUSION; CERTAINTY; EXPERIENCE; HEIDEGGER; HERMENEUTICS; KNOWLEDGE BY ACQUAINTANCE/BY DESCRIPTION; NIETZSCHE; PROTOCOL SENTENCES; SENSATION/COGNITION.

#### BIBLIOGRAPHY

- Ayer, A. J.: “Basic propositions,” in *Philosophical Analysis*, ed. M. Black (Englewood Cliffs, NJ: Prentice-Hall, 1950), 60–74.  
 Bonjour, L.: “Toward a Defense of Empirical Foundationalism,” in *Resurrecting Old-Fashioned*

- Foundationalism*, ed. M. DePaul (Lanham, MD: Rowman and Littlefield, 2001).  
 Goldman, A.: *Empirical Knowledge* (Berkeley, Calif.: University of California Press, 1988), part 2.  
 Haack, S.: *Evidence and Inquiry* (Oxford: Blackwell, 1993).  
 Kant, I.: *Critique of Pure Reason*, trans. N. Kemp Smith (London: Macmillan, 1964).  
 Lewis, C. I.: *An Analysis of Knowledge and Valuation* (LaSalle, IL: Open Court, 1946), ch. 7.  
 McDowell, J.: *Mind and World* (Cambridge, MA: Harvard University Press, 1994).  
 Peacocke, C.: “Does Perception Have Nonconceptual Content?,” *The Journal of Philosophy* 98 (2001), 239–64.  
 Sellars, W.: “Empiricism and the Philosophy of Mind,” in his *Science, Perception, and Reality* (London: Routledge and Kegan Paul, 1963).

ALAN H. GOLDMAN

**Goodman, Nelson (1906–98)** American philosopher. Goodman maintains that epistemology comprehends understanding or cognition in all its modes. It does not restrict itself to the theory of knowledge.

#### SYSTEMS

Goodman repudiates analyticity, necessity and certainties grounded in immediate experience. But even though any conviction can be abandoned, some are initially credible (Goodman, 1972, pp. 60–8). We construct systems of thought around these, revising or relinquishing them only as required to maximize overall credibility. Typically, any of several modifications yields a maximally credible system. Each such system is acceptable. Pluralism results, for maximally credible systems need neither reduce to nor supervene on a single base. In particular, the tenability of phenomenalism (*see* PHENOMENALISM) does not turn on its being reducible to or the basis for physicalism (Goodman, 1951).

In (1988), Goodman and Elgin suggest that something even more modest than initial credibility may suffice – current adoption. In theorizing, we order, emend, elaborate and extend considerations currently under adoption, whether credible or not. And we judge

their enduring epistemic worth by the cognitive efficacy of the systems that emerge.

Initially credible statements and current adoptions typically underdetermine individuation and classification. For example, our initially credible statements about stars do not settle whether black holes are stars. Systems answering to the same initially credible statements may decide such matters differently. One might consider black holes stars; another, the residue of extinguished stars. Both systems may be maximally credible. Then both answers are right.

Obviously, we cannot countenance contradiction. So, Goodman concludes, rightness is relative to acceptable system or world-version. Relative to one such version, black holes are stars. Relative to another, they are not. Still, it does not follow that anything goes. Relative to no acceptable system are black holes both stars and not stars (Goodman, 1978).

Cognitive rightness, Goodman suggests, consists in fitting and working. Right considerations intertwine to form a system that fits our initially credible statements and works to further our cognitive objectives. Truth is neither necessary nor sufficient for lightness. Truths are often wrong because irrelevant, trivial, convoluted or uninformative; falsehoods often right because relevant, illuminating, fruitful and/or elegant. Boyle's Law, although literally false, enhances understanding of gases as no true description of the phenomena can (Goodman and Elgin, 1988, pp. 153–66).

## PROJECTION

The *grue paradox* demonstrates the importance of rightness of categorization (see PROJECTION, PROJECTIBILITY). Something is *grue* if examined before future time *t* and green, or not so examined and blue. Even though all emeralds in our evidence class are *grue*, we ought not infer that all emeralds are *grue*. For “*grue*” is unprojectible. It cannot transmit credibility from known to unknown cases. Only projectible predicates are right for induction.

Goodman considers entrenchment the key to projectibility. Having a long history of successful projections, “green” is entrenched; lacking such a history, “*grue*” is not. A hypothesis is projectible, Goodman suggests, only if its predicates (or suitably related ones) are much better entrenched than its rivals’.

Past successes do not assure future ones. Induction remains a risky business. The rationale for favouring entrenched predicates is pragmatic. Of the possible projections from our evidence class, the one that fits with past practice enables us to best utilize our cognitive resources. Its prospects of being true are no worse than its competitors’ and its cognitive utility is greater.

Respect for entrenchment does not preclude conceptual innovation. A term like “quark” may be introduced where no entrenched predicate serves. Having no history of projection, it lacks earned entrenchment. Its projectibility derives from entrenchment inherited from related terms, e.g. “electron”. When competing hypotheses are equal in earned entrenchment, marked differences in inherited entrenchment determine projectibility.

Despite their entrenchment, predicates whose projection leads regularly from true premises to false conclusions are unprojectible. The history of failed Newtonian projections drained classical physical categories of their projectibility, making way for relativistic categories. Novel predicates thus become projectible by fitting into working inductive systems or into replacements for ineffective ones (see INDUCTION; PROBLEMS OF INDUCTION).

## ART

According to Goodman, aesthetics is a branch of epistemology. The arts enhance understanding and aesthetics explains how they do so (Goodman, 1964).

Works of art, he maintains, belong to symbol systems. To understand a work is not to “appreciate” or enjoy it or find it beautiful, but to interpret it correctly – to recognize what and how it symbolizes and how what

it symbolizes bears on other aspects of our worlds. An encounter with art can engender new modes of apprehension, provoking novel classifications that cut across stale categories to reveal hitherto unrecognized patterns and discrepancies. Not all are literal. Metaphor, allusion and more complex forms of reference often make connections that available literal terminology cannot capture (Goodman, 1984, pp. 55–77).

One prominent mode of symbolization in the arts is exemplification, whereby a symbol refers to some of its own properties. An early Picasso might literally exemplify blue and metaphorically exemplify melancholy. But exemplification is not peculiar to art. Being the relation of samples to what they sample, it is a staple of science and commerce as well.

A sample affords epistemic access to the stuff it samples. From a fabric swatch, one can infer the character of the corresponding fabric. Likewise, when works of art exemplify, they supply epistemic access to some of their features and to things that share those features. Largely through literal and metaphorical exemplification, *Guernica* enables us to understand something of the horror of war (Goodman, 1964, pp. 45–95). With samples, as with other symbols, rightness of interpretation is crucial.

Goodman does not deny the emotional impact of art, but relocates its importance. In the arts, he maintains, emotions function cognitively. We use our reactions to a work as tools for exploring it. And the more refined our sensibilities, the more insight emotion provides (*ibid.*, pp. 248–52).

Merit too functions as means. Rather than attending to a work to determine its worth, Goodman believes, we should use evaluation to focus attention. That seemingly similar sculptures differ in value provokes a search for differences. It may lead to discovery of features we could not previously discern, thereby extending our perceptual capacities (Goodman, 1972, pp. 120–1).

Goodman's epistemology is thus a wide-ranging inquiry into cognitive excellences and the conditions that foster them. It comprehends the arts as well as the sciences. And it considers the contributions to the advance-

ment of understanding made by perception, emotion, theory, practice, and symbolizing.

#### WRITINGS

- Fact, Fiction, and Forecast*, 4th edn (Cambridge, MA: Harvard University Press, 1983).  
*Languages of Art*, 2nd edn (Indianapolis, IN: Hackett, 1976).  
*Of Mind and Other Matters* (Cambridge, MA: Harvard University Press, 1984).  
*Problems and Projects* (Indianapolis, IN: Hackett, 1972).  
*The Structure of Appearance*, 3rd edn (Dordrecht: Reidel, 1977).  
*Ways of Worldmaking* (Indianapolis, IN: Hackett, 1978).  
 with Elgin, C. Z.: *Reconceptions* (Indianapolis, IN: Hackett, 1988).

#### BIBLIOGRAPHY

- Elgin, C. Z.: *With Reference to Reference* (Indianapolis, IN: Hackett, 1983).  
 Mitchell, W. J. T.: *Iconology* (Chicago, IL: University of Chicago Press, 1986), ch. 2.  
 Scheffler, I.: *The Anatomy of Inquiry* (Indianapolis, IN: Hackett, 1981), chs 8–10.  
 Schwartz, R.: "The Power of Pictures," *Journal of Philosophy* 82 (1985), 711–20.

CATHERINE Z. ELGIN

**Grice, H. Paul (1913–88)** British philosopher. Grice, who worked first at Oxford and then at the University of California at Berkeley, is best-known for his work on philosophy of language and mind, but in later years he worked extensively in ethics and metaphysics. Most of this work remained unpublished at his death.

Grice's contributions to epistemology are two – one direct and one indirect. The direct contribution is his attempt to provide a new twist on causal theories of perception; such theories can be found at least as early as Ockham (*see* OCKHAM) and more recently in Price, but they were out of favour at the time Grice wrote. Grice's defence of the theory utilizes an important methodological distinction between the domain of philosophy and that of



the specialized sciences. It is thus antithetical to current naturalized approaches to epistemology (see NATURALIZED EPISTEMOLOGY), even though it shares with such approaches an emphasis on causal connections. He argued that philosophical analyses of knowledge should leave one or more blanks to be filled in by psychology or neuro-psychology, and that the philosophical aspect of the account should be sufficiently neutral to allow particular causal accounts to fit in. His very tentative account was in terms of an appropriate causal relation (details to be specified by the specialist) between a state of affairs and a true present tense sense-datum report.

His indirect contribution is his theory of conversational implicature. The conversational implications of a statement are (roughly) those propositions that cannot be inferred from the explicit literal statement, but that can be inferred from the fact that it was made at all, or from the way the statement was made (choice of words, tone, etc.). One specific philosophical point which Grice wished to counter with the distinction was the claim that, for example, we do not know that we have two hands under normal circumstances because it would be odd to assert in those circumstances that we know that we have two hands. Grice's distinction enables us

to say why it is odd, namely because the statement is obviously true and there is a general conversational injunction not to waste everyone's time by stating obvious truths. A second epistemological use he envisaged for the theory, related to the first point above, was to define sense-data theories against the argument that it is odd to say that one "seems to see a dog" under normal circumstances. Grice continued to believe that sense-data theories were defensible after most sense-data theorists had abandoned approach. His suggestion was that sense-data theories be reformulated in linguistic terms that would enable them to be ontologically neutral, but he never developed this approach in detail.

*See also* CAUSAL THEORIES IN EPISTEMOLOGY; PSYCHOLOGY AND EPISTEMOLOGY; REPRESENTATIVE REALISM; SENSE-DATA.

#### BIBLIOGRAPHY

- "The Causal Theory of Perception," *Proceedings of the Aristotelian Society* supp. vol. (1961), 121–52.  
*Studies in the Ways of Words* (Harvard, MA: Harvard University Press, 1989).

RICHARD E. GRANDY

# H

---

**Habermas, Jürgen (1929– )** German philosopher and sociologist, the leading contemporary representative of Critical Theory, a tradition of Marxist social philosophy initiated by the Frankfurt School in the 1930s. The aim of Critical Theory is to explain the development of modern capitalist society, but also to indicate ways in which it may be released from relations of domination and exploitation. Unlike other forms of Marxism (*see* MARXISM), Critical Theory has been acutely aware of the need to examine its own cognitive status and has therefore given great weight to epistemological questions. Defining itself in contrast both to “metaphysics” and to “scientism”, it has repudiated the search for timeless philosophical foundations, while insisting that there are legitimate theoretical and moral pursuits beyond those of the natural sciences. In particular, it deplores the modern tendency of confining rationality (*see* RATIONALITY) to “instrumental reason”, the efficient marshalling of means in the services of ends which are left unjustified.

Habermas has tried to provide Critical Theory with a new and systematic epistemological basis by developing, first, a broad conception of reason with ethical implications, and second, a non-positivist methodology of the social sciences (*see* SOCIAL SCIENCES). His main theme is a switch from the “subject–object scheme” of post-Cartesian epistemology (the “philosophy of consciousness”) to a theory of intersubjective communication.

*Knowledge and Human Interests* (1968) contains what Habermas calls a “critical self-reflection” on the foundations of knowledge which results in a kind of transcendental pragmatism. Habermas distinguishes three types of inquiry – the “empirical analytical sciences”, i.e. those that provide nomological causal knowledge, the “historical–

hermeneutic sciences”, and the “critically oriented sciences” (psychoanalysis, critical social theory). He claims that each is guided by its own “cognitive interest”, respectively, the “technical” interest in the prediction and control of nature, the “practical” interest in understanding other human beings, and the “emancipatory” interest in freedom and the overcoming of unconscious compulsion. These cognitive interests are “general orientations” or “cognitive strategies” with a “quasi-transcendental” status. On the one hand, they have their basis in the productive labour and communication, and hence ultimately in the “natural history of the human species”. On the other hand, they are not merely permanent features of human history, but are virtually transcendental preconditions of human knowledge. For example, empirical science does not investigate an antecedently given domain of objects: empirical reality is “constituted” in the course of our pursuit of the technical interest.

In spite of the idealist terminology (*see* IDEALISM), this need not mean that we create nature, but only that there are no true empirical statements independent of our pursuit of the technical interest. However, the nature of this dependency remains obscure: as part of our natural history cognitive interests might *de facto* be invariant, but they remain contingent, and hence cannot be transcendental in Kant’s sense. Moreover, it remains unclear how a Kantian “critique”, an examination of the preconditions of knowledge, can underpin a Marxist critique of social structures.

In response to the objection that his assimilation of knowledge and human interests reduces epistemology in a naturalistic (*see* NATURALISM) or instrumentalist way, Habermas has distinguished between “action”, which is guided by pragmatic interests, and

“discourse”, an intersubjective search for knowledge which is guided solely by the aim of reaching a rational consensus. He has also distinguished between “critical sciences”, which investigate the possibilities of emancipation, and “reconstructive sciences”, the name he has given to sciences like linguistics, which aim to make explicit in a theoretical form the various capacities needed for human interaction. Among the latter, pride of place goes to “universal pragmatics”, a theory that is modelled on Chomsky’s reconstruction of linguistic competence, but aims to accommodate the pragmatic and intersubjective aspects of language highlighted by speech-act theories. Universal pragmatics examines the universal conditions of communication and interaction. Habermas argues that it provides a basis for Critical Theory, since the universal preconditions of linguistic communication include normative commitments with ethical implications. When we engage in a discourse, we “counterfactually anticipate” an “ideal speech situation” characterized by “equality” and “reciprocity of participation”. Perhaps these sibylline pronouncements can be taken to mean that in an argumentative discourse we are entitled to assume that all participants seek an agreement not on the basis of coercion or deceit, but only through reasoning.

Habermas’s theory of communication includes a “consensus theory of truth”, according to which a statement is true if and only if it would be accepted by all in an ideal speech situation. Several reformulations notwithstanding, this approach commits a genetic fallacy: that a statement is accepted as true by any number of speakers, however free they may be from coercion or prejudice, does not entail that it is true.

Habermas’s methodology of the social sciences is fully presented in *The Theory of Communicative Action* (1981). Against positivism he insists that the social sciences cannot adopt the “objectivizing” attitude of causal explanation, but must seek to understand human practices from the perspective of potential participants, by focusing on the explanations the participants themselves would give for their actions. But against the relativistic tendencies of hermeneutical and

Wittgensteinian sociologists he insists that social science can nevertheless criticize these practices on theoretical and ethical grounds. Perhaps the most interesting contribution of Habermas’s theory of discourse is a kind of transcendental argument (see TRANSCENDENTAL ARGUMENTS) against relativism (see RELATIVISM). In order to contribute to a discourse with their opponents, the arguments relativists advance on behalf of “primitive” cultures have to abide by standards of argumentation that are absent from the cultures they defend.

Habermas’s work is eclectic and often obscure. But his untiring endeavours to provide an account of rationality that avoids both objectivism (see OBJECTIVITY) and relativism (see SOCIOLOGY OF KNOWLEDGE) are impressive.

See also CONTINENTAL EPISTEMOLOGY; SOCIAL SCIENCES.

#### WRITINGS

- Knowledge and Human Interests* (London: Heinemann, 1971).  
*Communication and the Evolution of Society* (London: Heinemann, 1979), ch. 1.  
*The Theory of Communicative Action*, 2 vols (Cambridge: Polity, 1984, 1987).  
*Moral Consciousness and Communicative Action* (Cambridge: Polity, 1990), ch. 1.  
*Truth and Justification* (Cambridge: MIT Press, 2003).

#### BIBLIOGRAPHY

- McCarthy, T.: *The Critical Theory of Jürgen Habermas* (Cambridge: Polity, 1984).  
Thompson, J. B. and Held, D., eds: *Habermas: Critical Debates* (London: Macmillan, 1982).

HANS-JOHANN GLOCK

**Hegel, Georg Wilhelm Friedrich (1770–1831)** Though widely misunderstood, Hegel’s theory of knowledge is rich and often insightful. The standard view of Hegel is that he sought to overcome the sceptical character of Kant’s distinction between appearances and things in themselves by rejecting

epistemology and “purifying” Kant’s transcendental idealism to dispense with things in themselves. Instead, Hegel criticized Kant’s arguments for idealism, and his epistemology has great contemporary relevance. Hegel was the first epistemologist to realise that a socially and historically based epistemology is consistent with realism (see REALISM). His epistemology is anti-foundationalist (see FOUNDATIONALISM); he rejects non-conceptual knowledge and the ideal of certainty (see CERTAINTY), especially for alleged “elementary” beliefs or experiences. He holds a correspondence analysis of truth, though not a correspondence criterion of truth, and he defends a fallibilist account of justification.

Hegel’s theory of justification contains both externalist (see EXTERNALISM/INTERNALISM) and coherentist (see COHERENTISM) elements. He recognizes that some *prima facie* justification is provided by percepts and beliefs being generated reliably by our interaction with the environment. Hegel contends that full justification additionally requires a self-conscious, reflective comprehension of one’s beliefs and experiences which integrates them into a systematic conceptual scheme (outlined in his *Logic*) which provides an account for them which is both coherent and reflexively self-consistent.

Rationalist elements (see RATIONALISM) appear in Hegel’s epistemology in his theses that knowledge of particulars requires applying conceptions to them, that observation terms and formal logic are insufficient for empirical knowledge, and that statements of laws of nature are conceptual constructs which express actual structures of nature. He also holds the rationalist ideal that everything worth knowing is rationally comprehensible, which he calls “absolute knowledge”.

Naturalist elements (see NATURALISM) appear in Hegel’s epistemology in his theses that biological needs (one root of consciousness) involve elementary classification of objects, that the contents of consciousness are derived from a public world, and that classificatory thought presupposes natural structures in the world. Hegel insists that philosophy is grounded in the empirical sciences: “Not only must philosophy accord with the experience

nature gives rise to; in its *formation* and in its *development*, philosophic science presupposes and is conditioned by empirical physics” (*Encyclopedia* § 246 Remark). He holds similar views about grounding social philosophy in both the nascent cultural sciences of his day and in political economy.

Hegel contends that the corrigibility of conceptual categories is a social phenomenon. Our partial ignorance about the world can be revealed and corrected because one and the same claim or principle can be applied, asserted and assessed by different people in the same context or by the same person in different contexts. Hegel’s theory of justification requires that an account be shown to be adequate to its domain and to be superior to its alternatives. In this regard, Hegel is a fallibilist (see FALLIBILISM) according to whom justification is provisional and ineluctably historical, since it occurs against the background of less adequate alternative views.

Realism in epistemology (see REALISM) requires two things: that there be things whose characteristics do not depend upon our thoughts or language, and that those things be knowable; it requires that there be no metaphysical distinction between appearance and reality which blocks knowledge of reality. Hegel’s “idealism” is in fact such a realism; it is a kind of ontological holism, and not the view typically associated with “absolute idealism” (see IDEALISM). According to Hegel, the causal characteristics of things are essential to their identity conditions, and the individual properties of things obtain only as members of contrastive sets of properties. Hence the causal interdependence of particulars, along with the constitutive similarities and differences among their properties, establish the mutual dependence of their identity conditions. The result is two-fold. On the one hand, particulars have their ground in the whole world-system, because their characteristics obtain only in and through contrast with opposed characteristics of other things and because they are generated and corrupted through their causal interaction with other things. On the other hand, Hegel analyses “the concept” (*der Begriff*) as an ontological structure. Hegel’s “concept” is a principle of the

constitution of characteristics through contrast; it exists only in and as the interconnection of things and their properties in the world. Hegel's "idea" is the instantiation of this conceptual structure by worldly things and phenomena. Hegel describes particular things as "ideal" because they are not individually self-sufficient, and thus not ultimately real. He characterizes the world-system as "spirit" because he believes it has a normative *telos* towards which it develops historically. Part of this *telos* is self-knowledge, which the world-system gains through human knowledge of the world.

The sceptical view that things are the unsensed causes of sensory experience has been popular from Protagoras to Putnam (see PUTNAM); it appears in Locke's "thing I know not what" and Kant's unknowable "thing in itself". Hegel's analysis of forces and scientific laws responds to this view and provides support for his holistic ontology. Hegel objects to the hypothetico-deductive model of explanation in ways which have only recently become commonplace. He defends a "phenomenological" account of laws of nature. (This account is distinct from Hegel's "phenomenological" method.) According to such an account, laws of nature are relations among manifest phenomena. This view was prominent throughout the nineteenth century in German and British physics. Hegel purports to show that nothing more can be attributed to any force or set of forces than precisely the array of manifest phenomena which they are postulated to explain, so that ultimately there is nothing more to "forces" than the conceptual interrelation of manifest phenomena. These interrelations are, on Hegel's view, objective features of those phenomena, and the aim of conceiving those phenomena is to formulate those interrelations accurately. Because the interrelations among and within natural phenomena are not strictly speaking perceptible, but none the less are objective features of those phenomena, those interrelations are conceptual and concepts are structures of nature.

Hegel develops various aspects of his epistemology in different parts of his philosophical system. The *Phenomenology of Spirit*

presents a sophisticated meta-epistemology which responds to Sextus Empiricus' problem of the criterion (the problem of establishing standards of assessment without circularity or dogmatism) (see PROBLEM OF THE CRITERION) and defends an outline of a substantive epistemology against a wide range of sceptical, relativist and subjectivist views (see SCEPTICISM; RELATIVISM; SUBJECTIVISM). Hegel defends his views by criticizing opposed views internally, on the basis of the principles and examples cited in those views. Accordingly, a core element in his meta-epistemology is a subtle account of self-criticism, used to explain his method of internal criticism and to avoid problems of question-begging.

Hegel's "System of Philosophical Science", comprising his *Logic*, *Philosophy of Nature* and *Philosophy of Spirit*, takes up a wide range of substantive epistemological issues. The *Logic* examines the ontological and cognitive roles of ontological categories (e.g. being, existence, quantity, essence, appearance, relation, thing, cause) and principles of logic (e.g. identity, excluded middle, noncontradiction). His *Logic* also analyses syllogism, judgement and principles of scientific explanation (mechanical, chemical and organic or teleological functions) in accordance with which we are able to know the world. The *Philosophy of Nature* treats these principles of explanation in connection with a wide range of examples drawn from the sciences of his day, about which he was quite informed.

Hegel's philosophical psychology is deeply naturalist and draws heavily from Aristotle. The first part of his *Philosophy of Spirit*, the "Philosophy of Subjective Spirit", treats psychological topics pertinent to epistemology, including sensibility, feeling and habit under the heading "anthropology"; the conscious phenomena of sense-perception, intellect, and desire under the heading of "phenomenology"; and theoretical intelligence, including intuition, representation, memory, imagination and thought under the heading "psychology".

See also CONTINENTAL EPISTEMOLOGY; DIALECTIC (HEGEL); IDEALISM; IN ITSELF/FOR ITSELF; RATIONALISM.



## WRITINGS

*Die Phänomenologie des Geistes* (Bamberg and Würzburg: 1807); trans. A. V. Miller as *The Phenomenology of Spirit* (Oxford: Clarendon Press, 1977).

*Wissenschaft der Logik*, 2 vols (Nürnberg: 1812–16; Berlin: 1831); trans. A. V. Miller as *Hegel's Science of Logic* (London: Allen and Unwin, 1969).

*Enzyklopädie der philosophischen Wissenschaften im Grundrisse*, 3 vols (Heidelberg: 1817, 1827, 1830); vol. 1 trans. T. F. Geraets, H. S. Harris and W. A. Suchting as *The Encyclopedia of Logic* (Indianapolis, IN: Hackett, 1991); vol. 2 ed. and trans. M. J. Petry as *Hegel's Philosophy of Nature* (London: Allen and Unwin, 1970); vol. 3 ed. and trans. M. J. Petry as *Hegel's Philosophy of Subjective Spirit* (Dordrecht: Reidel, 1978).

## BIBLIOGRAPHY

Beaumont, B.: "Hegel and the Seven Planets," *Mind* 62 (1954), 246–8.

Beiser, F., ed.: *The Cambridge Companion to Hegel* (Cambridge: Cambridge University Press, 1992).

Beiser, F.: *Hegel* (New York and London: Routledge, 2005).

Buchdahl, G.: "Hegel's Philosophy of Nature and the Structure of Science," in *Hegel*, ed. M. Inwood (Oxford: Oxford University Press, 1985), 110–36.

deVries, W.: *Hegel's Theory of Mental Activity* (Ithaca, NY: Cornell University Press, 1988).

Fulda, H. F.: *Das Problem einer Einleitung in Hegels Wissenschaft der Logik*, 2nd edn (Frankfurt: Kostermann, 1975).

Pippin, R. B.: *Idealism as Modernism: Hegelian Variations* (Cambridge: Cambridge University Press, 1977).

Westphal, K. R.: *Hegel's Epistemology* (Indianapolis, IN: Hackett, 2003).

Westphal, K. R.: *Hegel's Epistemological Realism: A Study of the Aim and Method of Hegel's Phenomenology of Spirit* (Dordrecht: Kluwer, 1989).

Will, F. L.: *Beyond Deduction: Ampliative Aspects of Philosophical Reflection* (London: Routledge, 1988).

KENNETH R. WESTPHAL

**Heidegger, Martin (1889–1976)** German philosopher. From the start of his career Heidegger defined his project as answering the question: What is the meaning of being? This metaphysical concern was motivated by the conviction that the dominant movements

at the turn of the century – naturalism and Kantian theories of science (see NATURALISM; KANT) – are loaded down with uncritical assumptions about reality and our place in it. Traditional epistemology assumes: (1) a picture of the self as essentially a mind or subject trying to correctly represent objects in the external world (the subject/object model), and (2) a conception of our everyday beliefs as needing a philosophical account to show how they are possible (foundationalism).

Heidegger's phenomenology (see PHENOMENOLOGY) of average everydayness – the description of everyday agency prior to reflection and theorizing – aims at dissolving the assumptions built into epistemology. According to this description, Dasein (or human existence) is "being-in-the-world", a unified totality consisting of practical "dealings" with things and the meaningful equipmental contexts of the familiar life-world. This "unified phenomenon" is revealed in Heidegger's description of hammering in a workshop. When everything is running smoothly in such activities, what initially shows up for us is not a brute hammer-thing invested with a use, but rather "hammering" which is for realizing a project. The hammer is "ontologically defined" by its relations of "in-order-to", "for-which", "by-means-of" and "for-the-sake-of-which" as these are woven together by our practices. This totality of the "ready-to-hand" makes up the "world" as the tacit "dwelling" for our involvements. At the same time, Dasein's own identity as an agent of a particular type is defined by the specific worlds in which it finds itself. Given this reciprocal interdependence of self and context, there is no way to distinguish a subject from the brute objects it supposedly seeks to represent.

The world of everydayness is always a public world. Its intelligibility is constituted by the linguistic articulations of a historical community (the "they"). As the context of standardized interpretations, the public pre-defines possible ways of understanding humans and equipment. Because we are "outside ourselves", dealing with things along the guidelines of a communally attuned "preunderstanding of being", we are caught in a hermeneutic

circle (see HERMENEUTICS) with no access to an uninterpreted given (see GIVEN). But the fact that the world and our involvements are already intelligible suggests that there is no need for a philosophical account of our beliefs and practices. The concern with foundations, like the subject/object model, arises only when there is a “breakdown” in our everyday dealings with things. Only when entities obtrude as brute, meaningless “present-at-hand” objects can we come to regard ourselves as mere spectators collecting data about items independent of us. Yet, since this specialized stance of theoretical reflection is derivative from being-in-the-world, it can have no broader implications for grasping our everyday epistemic predicament.

The later Heidegger discards the vestiges of humanism in *Being and Time* and describes both Dasein and worldly entities as emerging through manifestations of being itself. Epistemology is treated as part of technology, the Western tendency to treat reality as a “world-view” on hand for our inspection and use. The antidote to technology is described as a special sort of thoughtful “letting-be” which no longer tries to achieve mastery over things.

#### WRITINGS

- Sein und Zeit* (Tübingen: 1927); trans. J. Macquarrie and E. Robinson as *Being and Time* (New York: Harper and Row, 1962).  
 “Die Zeit des Weltbildes,” *Vorträge und Aufsätze* (Pfullingen: 1954); trans. W. Lovitt as “The Age of the World Picture,” in *The Question Concerning Technology and Other Essays* (New York: Harper Colophon, 1977), 115–54.

#### BIBLIOGRAPHY

- Guignon, C.: *Heidegger and the Problem of Knowledge* (Indianapolis, IN: Hackett, 1983).  
 Okrent, M.: *Heidegger's Pragmatism: Understanding, Being, and the Critique of Metaphysics* (Ithaca, NY: Cornell University Press, 1988).  
 Olafson, F. A.: *Heidegger and the Philosophy of Mind* (New Haven, CT: Yale University Press, 1987).

CHARLES GUIGNON

**Hempel, Carl Gustav (1905–97)** Born in Oranienburg, Germany, Hempel studied mathematics and physics as well as philosophy, at the universities of Göttingen, Heidelberg and Berlin. After a brief period of research in Europe, Hempel migrated to the United States in 1937 and taught at Yale, Princeton and Pittsburgh.

Hempel was strongly influenced by the positivist philosophers of the early twentieth century, notably Reichenbach, Schlick and Carnap, and became a prominent representative of the logical positivist (see LOGICAL POSITIVISM) perspective on knowledge, language and science. Although he stayed largely within the empiricist and scientific framework of positivism, he was also a highly effective internal critic of the positivist excesses, and his many writings, well known for their lucidity and judicious argumentation, contributed greatly to the process of transforming the positivist movement and integrating it with the mainstream philosophy of the English-speaking world.

Hempel's early critical work on the positivist criterion of meaning (see VERIFICATIONISM) helped to liberalize the restrictive positivist doctrine on what is “cognitively significant” and what is “cognitively meaningless”. His mature views on meaning came to incorporate an important holistic element: whole scientific theories must be taken as the ultimate units of cognitive significance, and it is only when a theory is taken together with its “interpretative system” (i.e. a set of statements in which both “theoretical” and “observational” terms occur) that one can meaningfully speak of its empirical content.

Hempel was among the first to develop precise definitions of “evidence *e* confirms hypothesis *h*” and related concepts. His approach was largely formal and syntactical. His celebrated “Raven Paradox” concerning confirmation of generalizations by “positive instances” inspired an active debate in confirmation theory for a number of years (see HEMPEL'S PARADOX OF THE RAVENS).

Hempel's most influential work by far was on the nature of scientific explanation. According to his “covering-law” conception of explanation, the occurrence of an event is

explained by subsuming, or "covering", that event under a general law. When the covering-laws are deterministic, the explanation takes the form of a deductive argument with laws and statements of antecedent conditions as premises and an appropriate statement describing the event to be explained as its conclusion. Explanations conforming to this deductive model are called "deductive-nomological explanations". Hempel applied the model to explanations in history, explanations of human actions and functional explanations in biology and the social sciences.

Hempel's covering-law approach also allowed "statistical explanations", explanations in which statistical or probabilistic laws are used to show that the event to be explained is made highly probable, rather than deductively necessitated, by the explanatory premisses. The nature of the precise constraints to be placed on this "statistical model" has become a topic of much productive discussion in philosophy of science during the last three decades, spawning numerous alternative models.

Underlying Hempel's work on explanation are the following two central ideas: first, explanation, or scientific understanding, is not merely a matter of "intellectual satisfaction" but must have an objective, testable basis; second, this testability condition is to be implemented by the requirement that an acceptable explanation must show that the occurrence of the phenomenon to be explained can be rationally expected on the basis of the information contained in the explanatory premisses.

See also BAYESIAN EPISTEMOLOGY in Part I; EXPLANATION.

#### WRITINGS

*Fundamentals of Concept Formation in Empirical Science* (Chicago, IL: Chicago University Press, 1952).  
*Aspects of Scientific Explanation* (New York: Free Press, 1965). This volume includes most of Hempel's important philosophical papers.  
*Philosophy of Natural Science* (Englewood Cliffs, NJ: Prentice-Hall, 1966).

JAEGWON KIM

**Hempel's paradox of the ravens** In a pioneering study of confirmation, Carl G. Hempel (1945) introduced a paradox that raises fundamental questions about what counts as confirming evidence for a universal hypothesis. To generate the paradox three intuitive principles are invoked:

1. *Nicod's Principle* (after Jean Nicod, 1930): Instances of *As* that are *Bs* provide confirming evidence for the universal hypothesis that all *As* are *Bs*; while instances of *As* that are non-*Bs* provide disconfirming evidence. For example, instances of ravens that are black constitute confirming evidence for the hypothesis "All ravens are black", while instances of non-black ravens are disconfirming.
2. *Equivalence Principle*: If *e* is confirming evidence for hypothesis *h1*, and if *h1* is logically equivalent to hypothesis *h2*, then *e* is confirming evidence for *h2*. For example, if instances of ravens that are black are confirming evidence that all ravens are black, they are also confirming evidence that all non-black things are non-ravens, since the latter hypothesis is logically equivalent to the former.
3. *A Principle of Deductive Logic*: A sentence of the form "All *As* are *Bs*" is logically equivalent to one of the form "All non-*Bs* are non-*As*".

Using these principles, the paradox is generated by supposing that all the non-black things so far observed have been non-ravens. These might include white shoes, green leaves and red apples. By Nicod's principle, this is confirming evidence for the hypothesis "All non-black things are non-ravens". (In the schematic version of Nicod's principle, let *As* be non-black things and *Bs* be non-ravens.) But by principle (3) of deductive logic, the hypothesis "All non-black things are non-ravens" is logically equivalent to "All ravens are black". Therefore by the equivalence principle (2), the fact that all the non-black things so far observed have been non-ravens is confirming evidence for the hypothesis that all ravens are black. That is, instances of white shoes, green leaves and

red apples count as evidence for this hypothesis, which seems absurd. This is Hempel's ravens paradox.

Various solutions have been proposed, of which three will be noted. The first is due to Hempel himself, who accepts all three generating principles and the conclusion that follows from them. Hempel's solution or resolution consists in an attempt to show that the conclusion is not in fact paradoxical, although it may seem so. One reason the conclusion may seem paradoxical is that the hypothesis "All ravens are black" appears to be about ravens, and not about shoes, leaves or apples. Therefore, information about the latter may appear to be irrelevant. But, says Hempel, the hypothesis that all ravens are black is in fact about *everything*, including shoes, leaves and apples. It claims that everything is such that if it is a raven then it is black. It is what logicians call a universal quantification of the form "for all  $x$ , if  $x$  is a raven then  $x$  is black". A second reason the conclusion may seem paradoxical is that if we know in advance that a certain item, which is white, is a shoe, then we know in advance that it is not a raven. So this would not be a genuine test, since information is presupposed that could not falsify the hypothesis. But, Hempel notes, if this additional information is unavailable beforehand – if we know only that it is white, and then determine by observation whether or not it is a raven – we have a genuine test. In short, one can test the hypothesis that all ravens are black by finding a raven and determining whether it is black, or by finding a non-black thing and determining whether it is a non-raven.

Many commentators have felt, however, that testing the hypothesis by examining ravens is better than doing so by examining non-black things. Solutions of this sort usually appeal to the idea of class size. It is provable that if the class of ravens is smaller than the class of non-black things, then finding a raven that is black increases the probability of the hypothesis that all ravens are black more than does finding a non-black non-raven. (For a proof, and a general defence of this approach, see Hosiasson-Lindenbaum, 1940; somewhat similar proposals were made by

Pears, 1950; von Wright, 1957; Alexander, 1958; and Good, 1960.) This solution is closely associated with the idea that information is confirming evidence for a hypothesis if and only if it increases the probability of the hypothesis. So assuming the class of ravens is smaller than the class of non-black things, information that examined ravens are black is better confirming evidence for the hypothesis that all ravens are black than is information that examined non-black things are non-ravens.

Some writers on evidence, however, reject the principle that information is evidence for a hypothesis if and only if it increases its probability (see Achinstein, 1983; Glymour, 1980). One approach (adopted by Achinstein) is to assume that a necessary but not a sufficient condition for information  $e$  to be evidence for hypothesis  $h$  is that  $h$ 's probability on  $e$  be "high", say, greater than  $1/2$ . (Nicod's principle thus requires modification to satisfy this necessary condition.) The solution to the ravens paradox then proceeds by showing that, in normal circumstances, the probability that all ravens are black, given that all observed ravens are black, will be high; whereas the probability that all non-black things are non-ravens, given that all observed non-black things are non-ravens, will not be high. The reason is that in the latter case the selection procedure for choosing non-black things to observe will usually be strongly biased in favour of non-ravens, whereas this will not normally happen when choosing ravens to observe. A selection procedure (or rule)  $S$  for choosing  $A$ s to observe is strongly biased in favour of  $B$ s if and only if the probability that all observed  $A$ s are  $B$ s, given that  $A$ s are selected for observation in accordance with  $S$ , is close to 1; whereas the probability that all observed  $A$ s are  $B$ s (without this assumption) is not close to 1. In choosing non-black things to observe, say white shoes, one may well be following the procedure to select non-black things from inside one's house, or within one's immediate visual field, or that are inanimate, or that are known in advance to be non-ravens. All these selection procedures are strongly biased in favour of non-ravens.

Now a probability theorem is provable to the effect that if the selection procedure *S* for choosing *As* to observe is strongly biased in favour of *Bs*, then – given certain conditions that will normally be satisfied – the probability that all *As* are *Bs*, given that all observed *As* have been *Bs*, will not be high. Accordingly the high probability condition for confirming evidence mentioned earlier will be violated. In brief, if in choosing non-black things to observe (which turn out to be shoes, leaves or apples) one is following a selection procedure such as one of those mentioned above that is strongly biased in favour of non-ravens, then the fact that all the non-black things observed have been non-ravens would not be confirming evidence that all non-black things are non-ravens, i.e. that all ravens are black.

To be sure, there are selection procedures for non-black things that are not strongly biased in favour of non-ravens. But unlike the biased ones, they are difficult to follow and to know that we have followed. By contrast, there are simple, straightforward selection procedures for ravens that are not strongly biased in favour of blackness (e.g. "Select ravens from different locales, at different times of year"). Since they are unbiased, the probability of the general hypothesis "All ravens are black" can become high with the observation of more and more black ravens, thus satisfying the necessary condition for confirming evidence. This is why in testing the ravens hypothesis we should prefer to select ravens rather than non-black things.

## BIBLIOGRAPHY

- Achinstein, P.: *The Nature of Explanation* (New York: Oxford University Press, 1983).  
 Alexander, H. G.: "The Paradoxes of Confirmation," *British Journal for the Philosophy of Science* 9 (1958), 227–33.  
 Glymour, C.: *Theory and Evidence* (Princeton, NJ: Princeton University Press, 1980).  
 Good, I. J.: "The Paradox of Confirmation", parts 1 and 2, *British Journal for the Philosophy of Science* 11 (1960), 145–8; 12 (1961), 63–4.

- Hempel, C. G.: "Studies in the Logic of Confirmation," parts 1 and 2, *Mind* 54 (1945), 1–26, 97–121. Reprinted, with additions, in Hempel, *Aspects of Scientific Explanation* (New York: Free Press, 1965), 3–51.  
 Hosiasson-Lindenbaum, J.: "On Confirmation," *The Journal of Symbolic Logic* 5 (1940), 133–48.  
 Nicod, J.: *Foundations of Geometry and Induction* (London: K. Paul, Trench, Trübner, 1930).  
 Pears, D. F.: "Hypotheticals," *Analysis* 10 (1950), 49–63.  
 von Wright, G. H.: *The Logical Problem of Induction* (Oxford: Blackwell, 1957).

PETER ACHINSTEIN

**hermeneutics** The theory of interpretation, originating in older concerns with textual exegesis, today deals less with issues of the correctness of interpretations and more with the ontological question. What is the being of the entity which has an understanding of itself and its world? According to ontological hermeneutics, humans are self-constituting beings. We are what we make of ourselves in the course of our activities. Moreover, because we are already dealing with things in our practical affairs, we always have some "pre-understanding" of how things count in the familiar life-world. This tacit "know-how", embodied in our practices and made accessible in language, is prior to explicit propositional knowledge. There can be no exit from this all-pervasive shared background of understanding to gain access to brute facts or neutral data which might ground our interpretations. We are caught in a "hermeneutic circle": we grasp the world in terms of its components, but we can grasp things within the world only in terms of our prior mastery of the web of significance of the world as a whole. Nevertheless, as "insiders" initiated into the practices of a historical culture, the world is already intelligible to us. As a result, the questions of traditional epistemology are topics for specialized "regional" inquiries which have no wider consequences for making sense of our ordinary beliefs and practices.

See also SOCIAL SCIENCES.



## BIBLIOGRAPHY

- Gadamer, H.-G.: *Truth and Method* (1960), trans. J. Weinsheimer and D. G. Marshall (New York: Crossroad, 1989).
- Heidegger, M.: *Being and Time* (1927), trans. J. Macquarrie and E. Robinson (New York: Harper and Row, 1962).
- Hiley, D., Bohman, J. and Schusterman, R., eds: *The Interpretive Turn* (Ithaca, NY: Cornell University Press, 1991).

CHARLES GUIGNON

**Hintikka, Jaakko (1929– )** Finnish philosopher. Hintikka has made important contributions to many areas of philosophical research, including logic, epistemology, philosophy of language, philosophy of science and the history of philosophy.

In his early works *Distributive Normal Forms* (1953) and “Form and Content in Quantification Theory” (in *Two Papers on Symbolic Logic*, 1955), Hintikka developed two logical theories which he later applied to many different areas: the theory of distributive normal forms for quantification theory, and the theory of model sets which yields semantically motivated proof procedures for quantification theory and modal logics.

In his 1957 paper, “Modality as Referential Multiplicity”, Hintikka argued that intensional contexts involve a multiplicity of possible scenarios or worlds in which the referents of linguistic expressions can vary. This paper is one of the earliest formulations of the so-called possible worlds semantics. In his 1962 book *Knowledge and Belief* Hintikka developed an explicit modal logic for the concepts of knowledge and belief, employing the method of model sets, and applied it to various conceptual problems in epistemology, including the problem of knowing that one knows and Moore’s paradox (see MOORE’S PARADOX) of saying and disbelieving. This work opened a virtually new area of logical research, epistemic logic. This logic is the foundation of Hintikka’s theory of questions and answers, in which questions are construed as epistemic imperatives (or requests). He has applied this theory to the methodology

of knowledge acquisition, especially the scientific method; the resulting question-theoretic model of inquiry illuminates the role of inference, observation, experiment and strategy in inquiry (see KNOWLEDGE-SEEKING BY QUESTIONING; KK-THESIS).

By means of distributive normal forms Hintikka has defined new probability-measures for first-order languages which overcome some traditional difficulties in inductive logic. These measures can assign positive probabilities to non-trivial general hypotheses even in infinite universes and thus enable us to study inductive generalization within probabilistic confirmation theory. Likewise Hintikka has defined interesting new measures of information. He has shown that certain forms of deductive inference yield new information (in his perfectly objective sense of “information”), establishing thus an interesting distinction between “analytic” and “synthetic” deductive reasoning. In this way Hintikka has illuminated and partly vindicated Kant’s conception of the synthetic character of mathematical reasoning (see KANT; MATHEMATICAL KNOWLEDGE).

Much of Hintikka’s work in the semantics of natural language is based on the general theoretical framework called game-theoretical semantics. The philosophical precedent of this semantics is Wittgenstein’s idea (see WITTGENSTEIN) of language-game, but Hintikka conceives semantical games as games in the precise sense of the mathematical theory of games and as games of verification, not as games played by performing linguistic acts. Game-theoretical semantics has proved an immensely fruitful theoretical tool for the study of natural language, both as a general account of meaning and as a means of solving particular problems (for example, problems concerning coreference).

Although Hintikka has worked in a wide area, his work shows a great deal of conceptual and theoretical unity. This is partly due to the logical and semantical methods he uses, partly to the transcendental character (in the Kantian sense) of his philosophy. Hintikka has emphasized the role of rule-governed human activities in knowledge

acquisition and in cognitive representation; his game-theoretical approach to meaning is a case in point. The structures of such activities can be taken to provide the synthetic a priori features of our knowledge. In this respect Hintikka's philosophy is Kantian in spirit.

See also DIFFERENT CONSTRUCTIONS IN TERMS OF "KNOWS"; KNOWLEDGE-SEEKING BY QUESTIONING.

## WRITINGS

*Distributive Normal Forms in the Calculus of Predicates* (Acta Philosophica Fennica 6, 1953).  
*Two Papers on Symbolic Logic* (Acta Philosophica Fennica 8, 1955).

*Knowledge and Belief: An Introduction to the Logic of the Two Notions* (Ithaca, NY: Cornell University Press, 1962).

*Models for Modalities* (Dordrecht: Reidel, 1969).

*Logic, Language-Games, and Information* (Oxford: Oxford University Press, 1973).

*Knowledge and the Known: Historical Perspectives in Epistemology* (Dordrecht: Reidel, 1974).

*The Intentions of Intentionality and Other New Models for Modalities* (Dordrecht: Reidel, 1975).

*The Semantics of Questions and the Questions of Semantics* (Acta Philosophica Fennica, 28: 4, 1976).

with J. Kulas: *The Game of Language: Studies in Game-Theoretical Semantics and Its Applications* (Dordrecht: Reidel, 1983).

## BIBLIOGRAPHY

Bodgan, R. J., ed.: *Jaakko Hintikka* (Dordrecht: Reidel, 1987).  
*Philosophia* 11: 1–2 (1982).

RISTO HILPINEN

**historical knowledge** Knowledge acquired by the use of present evidence from which the knower draws conclusions about the past is called historical knowledge. This tripartite conception of historical knowledge – knower, evidence and known – gives rise to a great many problems.

## THE REALITY OF THE PAST

In historical knowledge, the object of knowledge is *transcendental*. That is, we never have experience with past events *as past*. This makes it impossible to verify directly the conclusions that we reach about the past on the basis of evidence. Bertrand Russell has put the problem in a particularly trenchant way by pointing out that it is possible that there is no past for historical knowledge to be about: "There is no logical impossibility in the hypothesis that the world sprang into being five minutes ago, exactly as it then was, with a population that 'remembered' a wholly unreal past" (Russell, 1921). One might think that this difficulty about knowing an unexperiencable past could be avoided by the use of induction in the following way: we use evidence to gain knowledge about past events which now-living people participated in and remember; once we find that our methods and types of evidence give us reliable knowledge, as certified by living memory, about these recent events, we are justified in extending these methods and types of evidence to the more distant past; we thus infer from this reliability in the case of one class of past events to reliability for all past events. But this will surely not work. This test for reliability makes essential use of memories which are themselves often unreliable and in critical need of independent verification. But we can never achieve the needed verification, even in the case of very recent memories, by directly comparing our memories to the past events remembered. Some thinkers have suggested that we regard historical knowledge as a construct – a way of making sense of *present* experience – rather than as knowledge about an unexperiencable object (Oakeshott, 1933; Croce, 1960; Meiland, 1965; Goldstein, 1976). Other thinkers believe that the hypothesis that there is a real past is the best explanation of present evidence being as it is; and they accordingly take historical statements to be true or false about that real past. (This dispute is structurally the exact equivalent of the dispute between scientific realism and constructive empiricism in the philosophy of science. See Van Fraassen, 1980.)

SUBJECTIVITY, INTERPRETATION,  
AND KNOWLEDGE

Let us suppose that there is a real past about which statements can be true or false. The evidence used by the historian often consists of texts created by participants in the events being studied or created by non-participant earlier historians after the events occurred. (Other types of evidence include works of art, rolls, records, speeches, letters, diaries, architectural remains, and monuments.) As in the case of all texts, the authors of these sources made *decisions*. They decided what was important to mention and what should be left out. This necessary process of selection is thought to introduce an inevitable element of *subjectivity* into these primary and secondary sources. Each source will, because of the necessity of selection, give a different account of the facts – of what the facts are, let alone what they mean – and consequently, each text will present an interpretation. For example, some sources may state that the prince was at the front of his troops throughout the battle, while other sources may state that he stayed in the rear. The historian's sources will thus disagree about what the facts are. On top of this, the historian must then make his or her own selection from the sources to construct a narrative or interpretation of what happened, thus telling us what the facts *mean*. This additional layer of selection introduces still more subjectivity into the historical work.

What guides this selectivity? The Scottish historian A. J. Youngson (1985) gives two accounts of the unsuccessful Jacobite rebellion of 1745–6 led by Charles Edward (Bonnie Prince Charlie), one from the Hanoverian point of view and the other from the Jacobite point of view. Each account includes all of the agreed-upon facts of the case. Each is an honest account, sticking strictly to the canons of historical inquiry and avoiding any element of propaganda. And yet the two accounts are very different from one another. Youngson says this about what guides the selection and interpretation represented in these accounts: "In the meantime it is clear only that many historians, whether they

recognize the fact or not, base their work on some prior set of principles or beliefs, and that what they write is an elaboration or illustration or defence of these principles or beliefs; and that all the others, however impartial they may try to be, have their personal preferences and their natural sympathies and therefore inevitably, although perhaps quite unconsciously, design their version of events to be consonant with their ideas about life in general" (Youngson, 1985, p. 14).

Of course, if we take this line, we shall find ourselves committed to a general scepticism about events – that is, a scepticism that includes complex present events too, since the same problems of selectivity and interpretation will apply to the present too. This is not specifically a problem about knowledge of the past.

## EXPLANATION AND UNDERSTANDING

Historical knowledge is often compared to scientific knowledge. Scientific knowledge is regarded as knowledge of the laws and regularities of nature which operate throughout past, present, and future. Some thinkers (e.g. the German historian Ranke) have argued that historical knowledge should be "scientific" in the sense of being based on research, on scrupulous verification of facts as far as possible, with an objective account being the principal aim. Others have gone further, asserting that historical inquiry and scientific inquiry have the same goal, namely providing explanations of particular events by discovering general laws from which (together with initial conditions) the particular events can be inferred (Hempel, 1942). This is often called "The Covering-Law Theory" of historical explanation. Proponents of this view usually admit a difference in direction of interest between the two types of inquiry: historians are more interested in explaining particular events, while scientists are more interested in discovering general laws. But the logic of explanation is said to be the same for both (*see* EXPLANATION; HEMPEL).

Yet a cursory glance at the articles and books that historians produce does not support

this view. Those books and articles focus overwhelmingly on the particular – e.g. the particular social structure of Tudor England, the rise to power of a particular political party, the social, cultural and economic interactions between two particular peoples. Nor is some standard body of theory or set of explanatory principles cited in the footnotes of history texts as providing the fundamental materials of historical explanation. In view of this, other thinkers have proposed that narrative itself, apart from general laws, can produce understanding, and that this is the characteristic form of historical explanation (Dray, 1957). If we wonder why things are the way they are – and, analogously, why they were the way they were – we are often satisfied by being told a story about how they got that way.

What we seek in historical inquiry is an understanding that respects the agreed-upon facts. A chronicle can present a factually correct account of a historical event without making that event intelligible to us – for example, without showing us why that event occurred and how the various phases and aspects of the event are related to one another. Historical narrative aims to provide intelligibility by showing how one thing led to another even when there is no relation of causal determination between them. In this way, narrative provides a form of understanding especially suited to a temporal course of events and alternative to scientific, or lawlike, explanation.

Another approach is understanding through knowledge of the purposes, intentions and points of view of historical agents. If we know how Julius Caesar or Leon Trotsky saw and understood their times and know what they meant to accomplish, then we can better understand why they did what they did. Purposes, intentions, and points of view are varieties of thought and can be ascertained through acts of empathy by the historian. R. G. Collingwood (1946) goes further and argues that those very same past thoughts can be *re-enacted* (and thereby made present) by the historian. Historical explanations of this type cannot be reduced to the covering-law model (Dray, 1957;

Atkinson, 1978) and allow historical inquiry to achieve a different type of intelligibility.

See also EXPLANATION; MEMORY; SOCIAL SCIENCES; VICO.

## BIBLIOGRAPHY

- Atkinson, R. F.: *Knowledge and Explanation in History* (London: Macmillan, 1978).  
 Collingwood, R. G.: *The Idea of History* (Oxford: Clarendon Press, 1946).  
 Croce, B.: *History: Its Theory and Practice* (New York: Russell and Russell, 1960).  
 Dray, W.: *Laws and Explanation in History* (Oxford: Oxford University Press, 1957).  
 Goldstein, L. J.: *Historical Knowledge* (Austin, Tex.: University of Texas Press, 1976).  
 Hempel, C. G.: "The Function of General Laws in History," *Journal of Philosophy* 29 (1942), 35–48.  
 Hobart, M. E.: "The Paradox of Historical Constructionism," *History and Theory* 28 (1989), 43–58.  
 Hurst, B. C.: "The Myth of Historical Evidence," *History and Theory* 20 (1981), 278–90.  
 Kempt, P.: "History as Narrative and Practice," *Philosophy Today* 29 (1985), 213–22.  
 Meiland, J. W.: *Scepticism and Historical Knowledge* (New York: Random House, 1965).  
 Mink, L. O.: *Historical Understanding* (Ithaca, NY: Cornell University Press, 1987).  
 Oakeshott, M.: *Experience and Its Modes* (Cambridge: Cambridge University Press, 1933).  
 Russell, B.: *The Analysis of Mind* (London: Allen and Unwin, 1921).  
 Van Fraassen, B. C.: *The Scientific Image* (Oxford: Clarendon Press, 1980).  
 Veyne, P.: *Writing History: Essay in Epistemology* (Middletown, CT: Wesleyan University Press, 1984).  
 White, H.: "Historical Pluralism," *Critical Inquiry* 12 (1986), 480–93.  
 Youngson, A. J.: *The Prince and the Pretender* (London: Croom Helm, 1985).

JACK W. MEILAND

**historicism** The view that the nature and value of some phenomenon can be understood only in the context of its historical development. Since our concepts and theories are themselves developing phenomena,

historicists usually deny that there is a trans-historical perspective from which to write history. Our theories must be evaluated in a process of immanent critique, and the history of that process is our final resource to understand the character of our present conceptions and their attraction for us. Historicists are therefore divided between those who deny that there are objective standards by which one historical stage may be assessed from the standpoint of another (Mannheim, Rorty (see RORTY)), and those who argue that history itself ultimately yields the perspective from which objective standards of criticism emerge. The latter view is clearest in Hegel (see HEGEL), who held that the end of history yields a state of “absolute knowledge” where the historical subject, the world spirit, attains an authentic understanding of the character of its own development. Historicist perspectives are usually contrasted with the methodology of natural science, which seeks to understand phenomena by reference to universal physical laws (the term initially gained currency in German debates about the relation between the *Geistes-* and *Naturwissenschaften*). It is thus odd that Popper (see POPPER) denounced historicism as the view that history is governed by principles analogous to scientific laws, though his critics may find a pleasing irony in his manifestly ahistorical use of the term.

DAVID BAKHURST

**Hobbes, Thomas (1588–1679)** English philosopher and political thinker. Hobbes's epistemology is not only an account of the nature and sources of knowledge; it is also a theory of the relative values of different types of knowledge, and especially of the surpassing value of scientific knowledge. The two principal types of knowledge that he recognizes are differently labelled in his various writings. Sometimes, as in the relatively early *Elements of Law* (1640), he speaks of “knowledge original” on the one hand and science on the other (Tonnies, 1889, p. 24); at other times, as in *Leviathan* (1651), he distinguishes between knowledge of fact and knowledge of consequence (Molesworth, 1839–45, vol. 3, p. 71). It is essentially the same distinction in

both places. Original knowledge or knowledge of fact is the registration by the senses of some accident or property of an object at a time; it is easily acquired knowledge, but it is also piecemeal, touching only this or that local object, and fleeting, lasting only as long as there are traces of it in the imagination. It can also be idiosyncratic, reflecting the peculiarities of one person's experience or vantage point. Original knowledge is not necessarily miscellaneous or inchoate: over time it enables people to recognize one accident or property as a sign of another, and thereby to register regularities in experience. But original knowledge does not alert us to any causal or necessary connections between properties or facts, and the range of regularities that it discloses is likely to be very narrow.

Scientific knowledge overcomes some of these limitations. It does hit on necessary connections. It is also general, synoptic, uncontroversial, well adapted to communication at a time and suitable for preservation over time. The key to its superiority over original knowledge is its linguistic medium. Whereas the medium of original knowledge is sense, scientific knowledge presupposes the existence of names, ways of combining names into propositions, and ways of combining true propositions into demonstrations (Molesworth, vol. 3, p. 35). It is only with the availability of names that we have a means of preserving sensory information beyond its half-life in one person's head. Names are vital aids to memory and help to make possible the communication of original knowledge. They also enable people to represent whole classes of things on the basis of similitudes between a few observed members of a class.

Propositions that link general names allow us to reason about things we could not even have thoughts about in the absence of language, namely relations between classes. Thanks to these “universal propositions”, we are able to conceive truths about relations between classes, e.g. all triangles and all figures whose interior angles add up to two right angles: thanks to demonstrations made out of universal propositions we are also able to register relations of dependence between such truths, which is what constitutes



scientific knowledge. Science, as Hobbes puts it in *Leviathan*, is the knowledge of the consequence of one affirmation to another (Molesworth, vol. 3, p. 71). It is knowledge of what follows from what or what is caused or necessitated by what (cf. Molesworth, vol. 1, p. 3). Hobbes's model of a work of science was Euclid's *Elements*, in which informative and even surprising conclusions are shown to be remote consequences of undeniable and obvious axioms and postulates (see GEOMETRY). He also admired writings in physics, such as Galileo's, which presented scientific demonstrations as exercises in applied geometry. Mere tabulations or registers of scientific results, such as were insisted upon by Bacon's method, Hobbes regarded as works of natural history only, and a cut below natural science or philosophy. The reason was that they only listed, and did not find principles or axioms from which to derive, the facts that were registered. The experimentalists who were responsible for the contents of the registers were in their turn not really scientists, for they lacked the demonstrative understanding that science required (Molesworth, vol. 4, pp. 436–7). The same went for the craftsmen, engineers and artificers who were skilled in such applied sciences as fortification-making: though they received the credit for their products, it was really mathematics that gave birth to them (Molesworth, vol. 3, p. 75).

Scientific knowledge could overcome more than limitations of original knowledge; it could act against some of the distorted valuations suggested by the passions. In other words, it could improve upon the pre-scientific practice of distinguishing good from bad or good from evil according to what gives pleasure or pain. That the practice was unreliable Hobbes thought needed little demonstration. He points out that something pleasant is good only in relation to an occasion and the constitution of the one who experiences the pleasure (cf. Molesworth, vol. 3, pp. 40–1). A person may find a thing pleasant on one occasion and call it "good" then, only to change his mind later (Molesworth, vol. 2, p. 196; vol. 3, p. 146). And the very thing that one person calls "good" on an occasion, another person may call "bad" on

the same occasion. Experiences of pleasure and pain, then, cannot be expected to generate consistent valuations over time or between people. Worse, such experiences can suggest defeasible valuations. Something that seems to be worth avoiding because it is unpleasant may turn out to be worth suffering all things considered, and something that seems to be worth pursuing because it is gratifying may not really be worth pursuing when all relevant considerations, including calculable consequences, are allowed to weigh. To give some illustrations drawn from politics, the felt unpleasantness of losing one's liberty under government does not by itself show that the loss of liberty is evil and to be avoided (Tonnie, pp. 138–40; Molesworth, vol. 2, pp. 126–9). Maybe the loss of liberty is acceptably unpleasant when one considers the unpleasant consequences of everyone's holding on to their liberty. Symmetrically, the fact that it is exhilarating to people to compete with one another, and highly gratifying to win, does not mean that competing and winning are good without qualification: on the contrary, a knowledge of the consequences of competing and winning may show that, on balance, they are not good after all. What Hobbes calls civil science or the science of politics is in large part a doctrine for correcting naive valuations in the light of a knowledge of the consequences.

One of the more striking features of Hobbes's account of scientific knowledge is his extravagant praise of its benefits. If it were not for natural philosophy and geometry, he claims repeatedly in his writings, human existence would largely be devoid of wellbeing. The fruits of navigation, architecture and agriculture – the main ingredients of "commodious living" as he and his contemporaries knew it – would have been enjoyed far less widely, if at all, and life would have been reduced to "the rude simpleness of antiquity" (Molesworth, vol. 2, p. iv). Not that natural philosophy and geometry on their own were able to supply the best life people were capable of: they did not provide for security. A distinct branch of science – a science of bodies politic or civil science – was required to ensure that material prosperity and life itself were not lost through

war – through reversion to a life that was “solitary, poor, nasty, brutish and short” (Molesworth, vol. 3, p. 113). Hobbes did not hold, then, that all of the good things in life were due to natural philosophy or geometry, but he did look to science of one kind or another to make human beings flourish or at least to keep human life from deteriorating.

Though Hobbes thinks that human beings need science in order to live well, he denies that science comes naturally to human beings. All of the abilities that science requires – from a knack for the apt imposition of names to an ability to develop or follow a long train of reasoning – can only be acquired with great effort, and probably only by a few. Perhaps surprisingly for a philosopher who thinks that human beings are not cut out by nature to discover natural causes or to live according to clear conceptions of the good and the just, Hobbes never confronts the sceptical arguments that suggest that human intellectual limitations forever put a true natural science or a true civil or moral science beyond our reach.

He was certainly aware of such arguments and of attempts in the 1600s to refute them. He was a close friend of Marin Mersenne and Pierre Gassendi, both of whom had published treatises concerned with scepticism as early as the 1620s, and he was one of the first readers of Descartes’ *Meditations* (see DESCARTES), being among the few handpicked to supply Objections (his was the Third Set) to the work in advance of its publication. However, his extremely terse comments on Meditation I show that he was unmoved by the sceptical hypotheses unfolded there, and probably that he did not appreciate their novelty. In his own published writings there is scarcely any mention of philosophical scepticism, let alone any investigation or refutation of it. Perhaps he did not need to vindicate his belief in the possibility of science against scepticism: he believed in a pretty modest sort of natural science and an unpretentious civil science. In common with many writers of the 1600s Hobbes held that the hypotheses proposed by a modern physics could not be conclusively demonstrated to be true, because there was no natural effect an omnipotent God could

not have produced in more than one way. A physics that attempted to show that a given effect couldn’t but have been produced in a certain way was thus out of the question. Yet this was the sort of physics that sceptical arguments were primarily directed against. Hobbes did believe that a demonstrative geometry and pure mechanics was possible, because the effects dealt with in those sciences were human artefacts rather than divine ones. But physics was less certain than these sciences. By aiming only to show what could have caused natural phenomena, it was suitably modest by sceptical standards. It had no pretensions to conclusiveness. Hobbesian civil philosophy did have such pretensions, but justifiably so in Hobbes’s opinion, since in telling us how to make and keep the peace it did not trace a natural phenomenon independent of our will to an equally independent cause; it traced something that depends on our will – namely peace – to precepts that can govern the will, and it motivated the will to accept the precepts. It thus did not attempt to transcend the order of things that were accessible to demonstrative science.

See also SCEPTICISM, MODERN.

#### WRITINGS

*English Works of Thomas Hobbes*, ed. Sir W. Molesworth, 11 vols (London: J. Bohn, 1839–45).

*Thomas Hobbes Malmesburiensis Opera Philosophica quae Latine scripsit omnia*, ed. Sir W. Molesworth, 5 vols (London: J. Bohn, 1839–45).

*The Elements of Law*, ed. F. Tonnies (London: Marshall and Simpkin, 1889).

New editions are being published by the Clarendon Press under the direction of N. Malcolm, and, in French, by Vrin, under the direction of Y.-C. Zarka.

#### BIBLIOGRAPHY

Jessop, T. E.: *Thomas Hobbes* (London: Longmans, 1960).

Laird, J.: *Hobbes* (London: Ernest Benn, 1934).

- Peters, R.: *Hobbes* (Harmondsworth: Penguin, 1956).
- Reik, M.: *The Golden Lands of Thomas Hobbes* (Detroit, MI: Wayne State University Press, 1977).
- Robertson, G. C.: *Hobbes* (Edinburgh: Blackwood, 1886).
- Sorell, T.: *Hobbes* (London: Routledge, 1986).
- Spragens, T.: *The Politics of Motion: The World of Thomas Hobbes* (Lexington, KY: University Press of Kentucky, 1973).
- Taylor, A. E.: *Hobbes* (London: Constable, 1908).
- Tuck, R.: *Hobbes* (Oxford: Oxford University Press, 1989).

TOM SORELL

**holism** Epistemological holism is the view that whole theories are the units of confirmation. Single hypotheses yield observational predictions only with the aid of a body of background theory. This means that a failed prediction does not conclusively refute the hypothesis from which it is derived: it is always possible to save the hypothesis by revising some of the background beliefs. Such holism has been defended by many twentieth-century philosophers, including Pierre Duhem, Rudolf Carnap and W. V. Quine. It is a consequence of epistemological holism that whether a belief is justified depends upon the support of the whole structure of beliefs to which it belongs. For empiricist philosophers who wish to explain meaning in terms of verification, epistemological holism may lead to a controversial semantic holism which asserts that more inclusive bodies of theory serve as the units of meaning.

## BIBLIOGRAPHY

- Carnap, R.: *The Logical Syntax of Language* (London: Routledge and Kegan Paul, 1937).
- Duhem, P.: *The Aim and Structure of Physical Theory* (New York: Atheneum, 1962).
- Quine, W. V.: *Pursuit of Truth* (Cambridge, MA: Harvard University Press, 1990), ch. 1.

CHRISTOPHER HOOKWAY

**Hume, David (1711–76)** Scottish philosopher, born in Edinburgh. Hume's theory

of knowledge starts from the distinction between perception and thought. When we see, hear, feel, etc. (in general, perceive) something we are aware of something immediately present to the mind through the senses. But we can also think and believe and reason about things which are not present to our senses at the time: e.g. objects and events in the past, the future or the present beyond our current perceptual experience. Such beliefs make it possible for us to deliberate and so to act on the basis of information we have acquired about the world.

For Hume all mental activity involves the presence before the mind of some mental entity. Perception is said to differ from thought only in that the kinds of things that are present to the mind in each case are different. In the case of perception it is an "impression"; in the case of thought, although what is thought about is absent, what is present to the mind is an "idea" of whatever is thought about. The only difference between an impression and its corresponding idea is the greater "force and liveliness" with which it "strikes upon the mind".

All the things that we can think or believe or reason about are either "relations of ideas" or "matters of fact". Each of the former (e.g. that three times five equals half of thirty) holds necessarily; its negation implies a contradiction. Such truths are "discoverable by the operation of pure thought, without dependence on what is anywhere existent in the universe". Hume has no systematic theory of this kind of knowledge; what is or is not included in a given idea, and how we know whether it is, is taken as largely unproblematic.

Each "matter of fact" is contingent; its negation is distinctly conceivable and represents a possibility. That the sun will not rise tomorrow is no less intelligible and no more implies a contradiction than the proposition that it will rise. Thought alone is therefore never sufficient to assure us of the truth of any matter of fact. Sense experience is needed. Only what is directly present to the senses at a given moment is known by perception alone. A belief in a matter of fact which is not present at the time must therefore be arrived at by a transition of some kind from present

impressions to a belief in the matter of fact in question. Hume's theory of knowledge is primarily an explanation of how that transition is in fact made. It takes the form of an empirical "science of human nature" which is to be based on careful observation of what human beings do and what happens to them.

Hume finds that all reasonings from a current impression to an absent matter of fact are founded on the relation of cause and effect. He therefore seeks the source of the idea of causality. That one thing is the cause of another cannot be discovered by thought alone. It must be derived from experience. It cannot be derived from any single observed instance; the most we can observe is that two things are contiguous and one happens before the other. But that is not sufficient for causation; "there is a *necessary connexion* to be taken into consideration".

We get that idea only after the repeated experience of things of one kind being followed by things of another kind. That sets up a connection in thought between things of the two kinds; the thought of one of them naturally leads the mind to a thought of the other. Then whenever we actually perceive (i.e. get an impression of) a thing of the first kind, we come, not just to think about, but to believe that a thing of the second kind will occur. What turns the mere thought of something into the belief that it is so is the presence of an impression, not just an idea, of something associated with that thing in the past. A belief, for Hume, is therefore "a lively idea related to or associated with a present impression". It is in following out connections established in our minds by past experience that we are led by a present impression to believe in an absent matter of fact.

In explaining in detail how this works, Hume first rejects the traditional idea that the "faculty of reason" or "the understanding", either when operating on its own or in combination with the results of sense experience, is the source of the inference to an absent matter of fact. Reason alone cannot be the source of the inference, since one matter of fact never by itself implies any other distinct matter of fact. We need the help of experience. But reason, even in combination with past

experience, cannot be what leads us from an observed correlation between things of two kinds to the belief that that correlation will continue in the future. If it did, it would have to "proceed upon the supposition" that future instances will resemble past instances, or that "the course of nature continues always uniformly the same".

All inferences from past or present experience to an unobserved matter of fact "proceed upon" that principle. But no assurance can be given to that principle from reason alone; it is not impossible, in the sense of implying a contradiction, for the future to be different from the past. Whether the future will resemble the past is a contingent matter of fact. Experience is therefore needed to assure us of that principle. It cannot do so alone, since the principle partly concerns the future, and past experience alone can tell us only how things have been in the past. Something more than past experience is needed.

But reason, even when combined with past experience, cannot be what leads us to believe that the future will resemble the past. If it did, it would be by means of an inference from past experience to the principle that the future will resemble the past. And, as before, any such inference would have to "proceed on the supposition" that the future will resemble the past. But that would "be evidently going in a circle, and taking that for granted which is the very point in question".

Reason, or the understanding, is therefore not in play.

There is nothing in any object, consider'd in itself, which can afford us a reason for drawing a conclusion beyond it; and . . . even after the observation of the frequent or constant conjunction of objects, we have no reason to draw any inference concerning any object beyond those of which we have had experience.

It is not reason but "the imagination" that is found to be responsible for our making the empirical inferences that we do. There are certain general "principles of the imagination" according to which ideas naturally come and go in the mind under certain conditions. It is the task of the "science of human nature" to discover such principles, but

without itself going beyond experience. For example, an observed correlation between things of two kinds can be seen to produce in everyone a propensity to expect a thing of the second sort given an experience of a thing of the first sort. We get a feeling, or an “impression”, when the mind makes such a transition, and that is what leads us to attribute *necessity* to the relation between things of the two kinds. There is no necessity in the relations between things that happen in the world, but, given our experience and the way our minds naturally work, we cannot help thinking that there is.

A similar appeal to certain “principles of the imagination” is what explains our belief in a world of enduring objects. Experience alone cannot produce that belief; everything we directly perceive is “momentary and fleeting”. And whatever our experience is like, no reasoning could assure us of the existence of something independent of our impressions which continues to exist when they cease. The series of our constantly changing sense impressions presents us with observable features which Hume calls “constancy” and “coherence”, and these naturally operate on the mind in such a way as eventually to produce “the opinion of a continu’d and distinct existence”. The explanation is complicated, but it is meant to appeal only to psychological mechanisms which can be discovered by “careful and exact experiments, and the observation of those particular effects, which result from [the mind’s] different circumstances and situations”.

We believe not only in bodies, but also in persons, or selves, which continue to exist through time, and this belief too can be explained only by the operation of certain “principles of the imagination”. We never directly perceive anything we can call ourselves; the most we can be aware of in ourselves are our constantly changing, momentary perceptions, not the mind or self which has them. For Hume, there is nothing that really binds the different perceptions together; we are led into the “fiction” that they form a unity only because of the way in which the thought of such series of perceptions works upon our minds.

The mind is a kind of theatre, where several perceptions successively make their appearance . . . there is properly no *simplicity* in it at one time, nor *identity* in different; whatever natural propensity we may have to imagine that simplicity and identity. The comparison of the theatre must not mislead us. They are the successive perceptions only, that constitute the mind. (See APPERCEPTION; KANT.)

Hume is often described as a sceptic in epistemology (see SCEPTICISM), largely because of his rejection of the role of reason, as traditionally understood, in the genesis of our fundamental beliefs. That rejection, although allied to the scepticism of antiquity, is only one part of an otherwise positive general theory of human nature which would explain how and why we think and believe and do all the things we do.

See also KANT; PROBLEMS OF INDUCTION; RATIONALISM; SCEPTICISM, MODERN; SELF-KNOWLEDGE AND SELF-IDENTITY.

#### WRITINGS

- A Treatise of Human Nature* (London: 1739–40); ed. L. A. Selby-Bigge, revised P. H. Nidditch (Oxford: Oxford University Press, 1978).  
*Enquiry Concerning the Human Understanding* (London: 1748); in *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, ed. L. A. Selby-Bigge, revised P. H. Nidditch (Oxford: Oxford University Press, 1975).

#### BIBLIOGRAPHY

- Fogelin, R.: *Hume's Scepticism in the Treatise of Human Nature* (London: Routledge and Kegan Paul, 1985).  
 Noonan, H. W.: *Hume on Knowledge* (London and New York: Routledge, 1999).  
 Owen, D.: *Hume's Reason* (Oxford: Oxford University Press, 2000).  
 Pears, D.: *Hume's System: An Examination of the First Book of His Treatise* (Oxford: Oxford University Press, 1991).  
 Penelhum, T.: *Hume* (New York: St Martin's Press, 1975).  
 Smith, N. Kemp: *The Philosophy of David Hume* (London: Macmillan, 1941).



Stove, D. C.: *Probability and Hume's Inductive Scepticism* (Oxford: Oxford University Press, 1973).  
 Stroud, B.: *Hume* (London: Routledge and Kegan Paul, 1977).

BARRY STROUD

**Husserl, Edmund (1859–1938)** German philosopher. Husserl was the creator of phenomenology (see PHENOMENOLOGY), in his *Logical Investigations* (1900–1), *Ideas* (1913) and later works. He was born in Moravia, received a PhD in mathematics, working with Weierstrass, and then turned to philosophy under the influence of Franz Brentano (see BRENTANO). He took over the latter's concern with intentionality and developed it further into what was to become phenomenology. After an attempt to found mathematics in psychology, in *Philosophie der Arithmetik* (vol. 1, 1891; vol. 2 never appeared), he gave up psychologism (see PSYCHOLOGISM) and started, around 1895, the work on *Logical Investigations*, his first phenomenological work. The first volume is a critique of psychologism; the second consists of six studies of basic logical notions.

In 1906 Husserl got the idea of the transcendental reduction, or epoche (see EPOCHE), and he developed phenomenology in an "idealist" direction. His *Ideas* is the first work which gives a full and systematic presentation of phenomenology. Husserl here distinguishes between the natural attitude and the transcendental, or phenomenological, attitude. The natural attitude we have when we are dealing with the objects in the world around us: physical things, events, actions, persons and the various features of these things (their shape, colour, etc.). These latter features, which can be shared by several objects, Husserl calls essences or *eidos* (see ESSENCE (HUSSERL)). Essences are studied in the eidetic sciences, of which mathematics is the most highly developed. Phenomenology is for Husserl a study of a different realm of objects, the *noemata*, the features that make one's consciousness be consciousness of objects. We arrive at these features through a special kind of reflection upon our own consciousness, the epoche (see EPOCHE).

The notion of a noema is a key to the understanding of Husserl's phenomenology. Brentano had characterized intentionality as "directedness upon an object". This characterization may work well when the act has an object, but leads to problems in the case of acts that fail to have an object, like hallucinations or people who are thinking about Pegasus. Instead of characterizing intentionality by appeal to an object towards which our consciousness is directed, Husserl concentrated on the features of our consciousness which make our acts be *as if of* an object. The *noema* is the collection of all these "as if of" features. The *noemata* are akin to Frege's "third world" objects (see FREGE), that is, the meanings of linguistic expressions. According to Husserl, "the *noema* is nothing but a generalization of the notion of meaning (*Bedeutung*) to the field of all acts" (*Ideas* III, 89, 2–4). Just as by distinguishing between an expression's meaning and its reference one can account for the meaningful use of expressions that fail to refer, so, according to Husserl, can the distinction between an act's *noema* and its object help us to overcome Brentano's problem of acts without an object.

The *noema* has two components: first, the "object meaning" that integrates the various components of our experience into experiences of the various features of *one* object, and, second, the "thetic" component that differentiates acts of different kinds, e.g. the act of perceiving an object from the act of remembering it or thinking about it. In acts of perception the *noema* that we can have is restricted by what goes on at our sensory surfaces, but the restriction does not narrow our possibilities down to just one. Thus in a given situation I may perceive a man, but later come to see that it was a doll, with a corresponding shift of *noema*. Such a shift of *noema* is always possible, corresponding to the fact that perception is always fallible. These boundary conditions, which constrain the *noema* that we can have, Husserl calls *hyle*. The *hyle* are not objects experienced by us, but are experiences of a kind which we typically have when our sensory organs are affected, but also can have in other cases, for example, under the influence of fever or drugs.

Unlike the meaning of linguistic expressions, as usually conceived, the *noemata* are rich objects, with an inexhaustible pattern of components. These are largely sedimentations of past experience and are to a high degree culture-dependent. They are influenced by our living together with other subjects where we mutually adapt to one another and come to conceive the world as a common world in which we all live, but which we experience from different perspectives. This adaptation (*Einfühlung*) was extensively studied by many of Husserl's students, notably Edith Stein. The common world, as experienced by us, Husserl calls our *lifeworld*. It was the main theme of his last big work *The Crisis of the European Sciences*, of which a part was published in 1930. Husserl's conception of the lifeworld has become important for the methodology of the humanities and the social sciences, largely because it provides a framework for discussing the subjective perspective and the many features of our way of structuring the world of which we are unaware, e.g. the many features of the culture we have grown up in.

From 1917 on, Husserl became increasingly interested in the role of the body and of human activities in the way we structure the world. These ideas recur in different versions in Heidegger's existentialism (see HEIDEGGER) and in Merleau-Ponty's phenomenology (see MERLEAU-PONTY). Sartre, too, was strongly influenced by Husserl, particularly by the idea that our material surroundings do not uniquely determine our *noema*. Sartre developed this idea into a philosophy of freedom.

Husserl died in Freiburg just in time to avoid being sent to a concentration camp (he was a Jew). His family, his library and all his 40,000 pages of manuscripts were rescued

from Germany by the Franciscan Van Breda, who established the Husserl archive in Louvain where the material is now accessible to researchers.

#### WRITINGS

- Logische Untersuchungen* I–II, in *Husserliana*, the standard edition of Husserl's works (The Hague: Martinus Nijhoff), vols 18–19; trans. J. N. Findlay as *Logical Investigations*, 2 vols (London: Routledge and Kegan Paul, 1970).
- Ideen* I–III, in *Husserliana*, vols 3–5; book 1 trans. F. Kersten as *Ideas* (The Hague: Martinus Nijhoff, 1982).
- Formale und transzendente Logik*, in *Husserliana*, vol. 17; trans. D. Cairns as *Formal and Transcendental Logic* (The Hague: Martinus Nijhoff, 1969).
- Kartesiansche Meditationen*, in *Husserliana*, vol. 1; trans. D. Cairns as *Cartesian Meditations* (The Hague: Martinus Nijhoff, 1960).
- Krisis der europäischen Wissenschaften und die transzendente Phänomenologie*, in *Husserliana*, vol. 6; trans. D. Carr as *The Crisis of European Sciences and Transcendental Phenomenology* (Evanston, IL: Northwestern University Press, 1970).

#### BIBLIOGRAPHY

- Bell, D.: *Husserl* (London: Routledge, 1990).
- Dreyfus, H., ed.: *Husserl, Intentionality and Cognitive Science* (Cambridge, MA: MIT Press, 1982).
- Miller, I.: *Husserl, Perception and Temporal Awareness* (Cambridge, MA: MIT Press, 1984).
- Mohanty, J.: *The Possibility of Transcendental Philosophy* (Dordrecht: Martinus Nijhoff, 1985).
- Smith, D. and McIntyre, R.: *Husserl and Intentionality: A Study of Mind, Meaning and Language* (Dordrecht: Reidel, 1982).

DAGFINN FOLLESDAL

# I

---

**idea** “Ideas” began with Plato, as eternal, mind-independent forms or archetypes of the things in the material world. Neoplatonism (see NEOPLATONISM) made them thoughts in the mind of God who created the world. The much criticized “new way of ideas”, so much a part of seventeenth- and eighteenth-century philosophy, began with Descartes’ (see DESCARTES) conscious extension of “idea” to cover whatever is in human minds too, an extension of which Locke made much use (see LOCKE). But are they like mental images, of things outside the mind; or non-representational, like sensations? If representational, are they mental objects, standing between the mind and what they represent; or are they acts and modifications of a mind perceiving the world directly? Finally, are they neither objects nor acts, but dispositions? Malebranche and Arnauld (and then Leibniz) famously disagreed about how “ideas” should be understood, and recent scholars disagree about how Arnauld, Descartes, Locke and Malebranche in fact understood them.

## BIBLIOGRAPHY

- Ashworth, F. J.: “Descartes’ Theory of Clear and Distinct Ideas,” in *Cartesian Studies*, ed. R. J. Butler (Oxford: Blackwell, 1972), 89–105.
- Gibson, J.: *Locke’s Theory of Knowledge, and Its Historical Relations* (Cambridge: Cambridge University Press, 1960), 13–28.
- Jolley, N.: *The Light of the Soul: Theories of Ideas in Leibniz, Malebranche, and Descartes* (Oxford: Clarendon Press, 1990).
- McRae, R.: “‘Idea’ as a Philosophical Term in the Seventeenth Century,” *Journal of the History of Ideas* 26 (1965), 175–90.
- Yolton, J. W.: “Ideas and Knowledge in Seventeenth-century Philosophy,” *Journal of the History of Philosophy* 13 (1975), 145–65.

R. S. WOOLHOUSE

**idealism** The philosophical doctrine that reality is somehow mind-correlative or mind-co-ordinated – that the real objects comprising the “external world” are not independent of cognizing minds, but only exist as in some way correlative to the mental operations. The doctrine centres on the conception that reality as we understand it reflects the workings of mind. And it construes this as meaning that the inquiring mind itself makes a formative contribution not merely to our understanding of the nature of the real but even to the resulting character we attribute to it.

For a long time, a dispute raged within the idealist camp over whether “the mind” at issue in such idealistic formulas was a mind emplaced outside of or behind nature (*absolute* idealism), or a nature-pervasive power of rationality of some sort (*cosmic* idealism), or the collective impersonal social mind of people-in-general (*social* idealism), or simply the distributive collection of individual minds (*personal* idealism). Over the years, the less grandiose versions of the theory came increasingly to the fore, and in recent times virtually all idealists have construed “the minds” at issue in their theory as a matter of separate individual minds equipped with socially engendered resources.

As the accompanying table (on p. 428) shows, idealist doctrine takes many forms. Perhaps the most radical of these is the ancient Oriental spiritualistic or panpsychistic idea – renewed in Christian Science – that minds and their thoughts are all there is; that reality is simply the sum-total of the visions (or dreams?) of one or more minds.

It is quite unjust to charge idealism with an antipathy to reality, with *ontophobia*, as Ortega y Gasset called it. For it is not the *existence* but the *nature* of reality that the idealist puts in question. It is not reality but

materialism that classical idealism rejects – and even here the idealists speak with divided voice. Berkeley's "immaterialism" does not so much reject the existence of material objects as their unperceivedness.

There are certainly versions of idealism short of the spiritualistic position of an ontological idealism that holds that (as Kant puts it at *Prolegomena*, s. 13, n. 2) "there are none but thinking beings". Idealism need certainly not go so far as to affirm that mind *makes* or *constitutes* matter; it is quite enough to maintain (for example) that all of the characterizing properties of physical existents resemble phenomenal sensory properties in representing dispositions to affect mind-endowed creatures in a certain sort of way, so that these properties have no standing at all without reference to minds. Weaker still is an explanatory idealism which merely holds that an adequate *explanation* of the real always requires some recourse to the operations of mind.

Historically, positions of the generally idealistic type have been espoused by numerous thinkers. For example, Berkeley (see BERKELEY) maintained that "to be (real) is to be perceived" (*esse est percipi*). This does not seem particularly plausible because of its inherent commitment to omniscience; it seems more sensible to claim "to be is to be perceivable" (*esse est percipi posse*). For Berkeley, of course, this was a distinction without a difference: if something is perceivable at all, then God perceives it. But if we forgo philosophical reliance on God, the issue looks different, and now comes to pivot on the question of what is perceivable for perceivers who are *physically realizable* in "the real world", so that *physical* existence could be seen – not so implausibly – as tantamount to observability-in-principle.

The three positions to the effect that real things just exactly are things as *philosophy* or as *science* or as "*commonsense*" takes them to be – positions generally designated as *scholastic*, *scientific* and *naïve* realism, respectively – are in fact versions of epistemic idealism exactly because they see reals as inherently knowable and do not contemplate mind-transcendence for the real. Thus, for example,

the thesis of naïve ("commonsense") realism that "External things exist exactly as we know them" sounds realistic or idealistic according as one stresses the first three words of the dictum or the last four.

There is also another sort of idealism at work in philosophical discussion, an *axiological* idealism that maintains both that values play an objective causal or constitutive role in nature and that value is not wholly reducible to something that lies in the minds of its beholders. Its exponents join the Socrates of Plato's *Phaedo* in seeing value as objective and as productively operative in the world.

Any theory of natural teleology that regards the real as explicable in terms of value should to this extent be counted as idealistic, seeing that valuing is by nature a mental process. To be sure, the good of a creature or species of creatures (e.g. their well-being or survival, for example) need not be actually mind-represented. But nevertheless, goods count as such precisely because if the creatures at issue *could* think about it, they *would* adopt them as purposes. It is this circumstance that renders any sort of teleological explanation at least conceptually idealistic in nature. Doctrines of this sort have been the stock in trade of philosophy from the days of Plato to those of Leibniz (see LEIBNIZ), with his insistence that the real world must be the best possible. And this line of thought has recently surfaced once more in the controversial "anthropic principle" espoused by some theoretical physicists.

Then too, it is possible to contemplate a position along the lines envisioned in Fichte's *Wissenschaftslehre*, which sees the ideal as providing the determining factor for the real. On such a view, the real is not characterized by the science we actually have but by the ideal science that is the *telos* of our scientific efforts. On this approach, which Wilhelm Wundt characterized as "ideal-realism" (*Idealrealismus*; see his *Logik*, vol. I, 2nd edn, 1895, pp. 86ff), the knowledge that achieves adequation to the real (*adaequatio ad rem*) by adequately characterizing the true facts in scientific matters is not the knowledge actually afforded by present-day science as we have it, but only that of an *ideal* or *perfected*

science. On such an approach – which has seen a lively revival in recent philosophy – a tenable version of “scientific realism” requires the step to idealization, and realism becomes predicated on assuming a fundamentally idealistic point of view.

Over the years, many objections to idealism have been advanced. Samuel Johnson thought to refute Berkeley’s phenomenalism by kicking a stone. He conveniently forgot that Berkeley’s theory goes to great lengths to provide for stones – even to the point of invoking the aid of God on their behalf. G. E. Moore (see MOORE) pointed to the human hand as an undeniably mind-external material object. He overlooked that, gesticulate as he would, he would do no more than *induce people to accept* the presence of a hand on the basis of the hand-orientation of their *experience*. C. S. Peirce’s (see PEIRCE) “Harvard experiment” of letting go of a stone held aloft was supposed to establish scholastic realism because his audience could not control their expectation of the stone’s falling to earth. But an uncontrollable expectation is still an expectation, and the realism at issue is no more than a realistic thought-posture.

Immanuel Kant’s (see KANT) famous “Refutation of Idealism” argues that our conception of ourselves as mind-endowed beings presupposes material objects because we view our mind-endowed selves as existing in an objective temporal order, and such an order requires the existence of periodic physical processes (clocks, pendula, planetary regularities) for its establishment. At most, however, this argumentation succeeds in showing that such physical processes *have to be assumed by minds*, the issue of their actual mind-independent existence remaining unaddressed. (Kantian realism is an intra-experiential “empirical” realism.)

It is sometimes said that idealism is predicated on a confusion of objects with our knowledge of them and conflates the real with our thought about it. But this charge misses the point. The only reality with which we inquirers can have any cognitive commerce is reality as we conceive it to be. Our only information about reality is via the operations of mind – our only cognitive access to

reality is through the mediation of mind-devised models of it.

Perhaps the most common objection to idealism turns on the supposed mind-independence of the real. “Surely”, so runs the objection, “things in nature would remain substantially unchanged if there were no minds.” This is perfectly plausible in one sense, namely the *causal* one – which is why causal idealism has its problems. But it is certainly not true *conceptually*. The objection’s exponent has to face the question of specifying *just exactly what* it is that would remain the same. “Surely roses would smell just as sweet in a mind-denuded world!” Well . . . yes and no. Agreed; the absence of minds would not *change* roses. But roses and rose-fragrance and sweetness – and even the *size* of roses – are all factors whose determination hinges on such mental operations as smelling, scanning, measuring, and the like. Mind-requiring processes are required for something in the world to be discriminated as being a rose and determined as being the bearer of certain features. Identification, classification, property attribution are all required and by their very nature are all mental operations. To be sure, the role of mind is here *hypothetical*. (“If certain interactions with duly constituted observers took place, *then* certain outcomes would be noted.”) But the fact remains that nothing could be discriminated or characterized as a rose in context where the prospect of performing suitable mental operations (measuring, smelling, etc.) is not presupposed.

The preceding inventory of versions of idealism at once suggests the variety of corresponding rivals or contraries to idealism. On the *ontological* side, there is *materialism*, which takes two major forms: (1) a *causal* materialism which asserts that mind arises from the causal operations of matter, and (2) a *supervenience* materialism which sees as an epiphenomenon to the machinations matter (albeit not a *causal* product thereof – presumably because it is somewhere between difficult and impossible to explain how physical processes could engender psychological results). On the *epistemic* side, the inventory of idealism-opposed positions includes: (1) A



factual realism that maintains linguistically inaccessible facts, holding that complexity and diversity of fact outruns the limits of the reach of the mind's actual or possible linguistic (or, generally, symbolic) resources; (2) A cognitive realism that maintains that there are unknowable truths – that the domain of *truth* runs beyond the limits of the mind's cognitive access; (3) A substantival realism that

maintains that there exist entities in the world which cannot possibly be known or identified: incognizables lying in principle beyond our cognitive reach; (4) A conceptual realism which holds that the real can be characterized and explained by us without the use of any such specifically mind-invoking conceptions as dispositions to affect minds in particular ways. This variety of different

**Table 1.** Versions of idealism

---

I. ONTOLOGICAL VERSIONS

(1) *Causal idealism*

Everything there is, apart from minds themselves, arises causally from the operations of minds.

(2) *Supervenience idealism*

Everything there is, apart from minds themselves, is supervenient (*see* EPISTEMIC SUPERVENIENCE) upon the operations of minds (i.e. somehow inheres in them in ways that are not necessarily causal but involve some other mode of existential dependency).

II. EPISTEMIC VERSIONS

(1) *Fact idealism*

To be as a fact is to be a language-formulable fact – that is, a truth. Every fact can be semantically captured in a language-formulated truth.

(2) *Cognitive idealism*

To be as a truth is to be knowable. Every truth can – potentially – be cognitively captured as an item of knowledge. Truth stands co-ordinate with the cognitive potential of mind.

(3) *Strong substantival idealism*

To be as a thing or entity is to be actually discerned (discriminated, identified, perceived) by some knower. (This is simply a restatement of Berkeley's idealistic thesis that "To be is to be perceived".)

(4) *Weak substantival idealism*

To be as a thing or entity is to be discernible (discriminable, identifiable, perceivable). Any real thing (entity, object) can, in principle, be discerned by some knower; it must, in principle, be of a nature that admits cognitive access.

(5) *Explanatory idealism*

An adequate explanation of the nature of physical ("material") reality requires some recourse to mental characteristics or operations within the substantive content of the explanation.

(6) *Conceptual idealism*

Reality is to be understood in terms of the category of mind: Our knowledge of the real is grasped not merely in mind-supplied but indeed to some extent even in mind-patterned terms of reference. Our knowledge of fact always reflects the circumstances of its being a human artifact. It is always formed through the use of mind-made and indeed mind-invoking conceptions, and its contents inevitably bear the traces of its man-made origins. Whatever we have any knowledge of we know in terms of mind-construed terms of reference *in whose conceptual content* there is some reflection of its origin in operations characteristic of mind.

---

versions of idealism/realism means that some versions of the one will be unproblematically combinable with some versions of the other. In particular, a conceptual idealism maintaining that we standardly understand the real in somehow mind-invoking terms of reference is perfectly compatible with a materialism which holds that the human mind and its operations ultimately root (be it causally or superveniently) in the machinations of physical process.

Perhaps the strongest argument favouring idealism is that any characterization of the real that we can devise is bound to be a mind-constructed one: *our* only access to information about what the real is through the mediation of mind. What seems right about idealism is inherent in the fact that in investigating the real we are clearly constrained to use our own concepts to address our own issues; we can only learn about the real in our own terms of reference. But what seems right about realism is that the answers to the questions we put to the real are provided by reality itself – whatever the answers may be, they are substantially what they are because it is reality itself that determines them to be that way. Mind proposes but reality disposes. But, of course, in so far as one can learn about this reality, it has to be done in terms accessible to minds. Accordingly, while philosophical idealism has a long and varied past and a lively present, it undoubtedly has a promising future as well.

See also BERKELEY; KANT; OBJECTIVITY; REALISM.

# BIBLIOGRAPHY

- For the history of idealism: Otto Willmann, *Geschichte des Idealismus*, 3 vols (Braunschweig: Vieweg, 1894–7).
- For the German tradition: M. Krönerberg, *Geschichte des deutschen Idealismus*, 2 vols (Munich: Beck, 1909 and 1912), and Nicolai Hartmann, *Die Philosophie des deutschen Idealismus*, 2 vols (Berlin: De Gruyter, 1923 and 1929).
- For British idealism: J. Pucelle, *L'Idéalisme en Angleterre de Coleridge à Bradley* (Paris: Presses Universitaires de France, 1955), and A. C. Ewing, *Idealism: A Critical Survey* (London: Macmillan,

1934), as well as his collection of texts, *The Idealist Tradition* (Glencoe, IL: Free Press, 1957). Contemporary defences of idealist doctrines are presented in N. Rescher, *Conceptual Idealism* (Oxford: Blackwell, 1973); J. Foster, *The Case for Idealism* (London: Routledge and Kegan Paul, 1982); and T. Sprigge, *The Vindication of Absolute Idealism* (Edinburgh: Edinburgh University Press, 1983).

NICHOLAS RESCHER

**ideology** The term “ideology” was coined by Destutt de Tracy in 1796 to refer to the “science of ideas”. This discipline, inspired by the empiricism of Bacon, Locke and Condillac, was to give a naturalistic explanation of the processes by which the mind forms thoughts (see EMPIRICISM; BACON; LOCKE; NATURALISM). The aims of *idéologie* were ultimately pedagogical and political. A “natural history of ideas” was to provide valuable knowledge of human nature and, by revealing the sources of erroneous belief, aid the perfection of scientific method. In turn, this knowledge was to facilitate the creation of a harmonious social order answering to the permanent needs of human beings.

After a brief period of eminence in the *Institut National*, the *idéologues* were dismissed by Napoleon in 1803, and ridiculed as Utopian visionaries under the name *idéologues*. In the 1840s, Marx and Engels appropriated the term “ideology” in their critique of the Young Hegelians, and perpetuated its pejorative use, employing it to mean, not a discipline devoted to the persecution of error, but a distorted system of beliefs or “false consciousness”. For Marx and Engels, an ideology is a set of ideas pertaining to social life (philosophical, religious, historical, economic, political ideas, etc.) that systematically misrepresents or “inverts” reality. Ideologies are born and sustained by certain socio-economic conditions and, in turn, serve to bolster and legitimize those very conditions. For instance, it is characteristic of ideological inversion to “objectify” products of human activity (e.g. religious phenomena and economic relations), representing them as autonomous forces and entities. As a result, contingent

social relations appear as immutable parts of the “natural order”. The capitalist market, for example, is seen as the sole mode of economic organization corresponding to “human nature”, political constitutions are represented as expressions of “natural law”, etc. At the same time, human behaviour rendered economically necessary by the prevailing social order is portrayed, and legitimized, as an outcome of “free choice” (e.g. the worker’s “decision” to sell his or her labour). Ideologies also typically contain self-fulfilling beliefs that are not recognized as such (e.g. the belief that members of some racial group are intellectually inferior causes them to be denied educational resources which, in turn, causes their poor academic performance).

On a classical Marxist position, ideological misconceptions cannot be dispelled simply by presenting those in their grip with “the truth”, since ideologies contain features (e.g. standards of evidence and argumentation) that inhibit recognition of reality as it is. Consequently, many Marxists have argued that ideologies can be defeated only through the transformation of those socio-economic relations that generate them. Moreover, human beings will be trapped in ideological delusion until history produces a class, the proletariat, the true interests of which coincide with the necessity of conceiving things as they are.

However, some thinkers inspired by Marx, such as members of the Frankfurt School, have been sceptical of this vision of the proletariat as the vehicle of Reason in history, and have argued that effective critique of dominant ideologies can occur without a wholesale transformation of social relations. Victims of an ideology may appreciate the falsehood of their conceptions if they can be brought to see the true source of their beliefs and the political function they fulfil. Developing parallels between Marx and Freud, the Frankfurt School present *ideologiekritik* as akin to psychoanalysis: the victim’s emancipation depends on an exercise of self-understanding, in which the real nature and origin of his or her fundamental beliefs is gradually unmasked. On this view, an adequate

conception of society must be fundamentally reflexive or self-aware; hence “critical theories”, in contrast to ideologies, understand their own origin and function.

In twentieth-century discussions of ideology, the term has often been employed in a purely descriptive and non-pejorative sense, being used simply to refer to the set of beliefs, attitudes, standards of rationality, etc. that embody the basic values of some social group and that group’s conception of the political order appropriate to those values. Thus, in this sense, we may speak of “liberal”, “conservative” or “socialist” ideologies. This change of meaning was motivated principally by two developments, the first being the growth of the sociology of knowledge (see SOCIOLOGY OF KNOWLEDGE), pioneered by the work of Karl Mannheim. While, like Marx, sociologists of knowledge have been concerned with the way in which ideas and theories are created and mediated by economic relations, social institutions, class interests etc., they have frequently rejected the idea that ideologies are forms of “false” consciousness, either because there is no possible non-ideological perspective on the world, or because the normative content of ideologies renders it inappropriate to describe them as true or false, or because ideologies, as symbolic systems with which individuals interpret social life, are functionally indispensable.

The second move towards a non-pejorative use of “ideology” occurred within the Marxist tradition itself (see MARXISM). Lenin, Lukács and Gramsci each in different ways use the term to refer to forms of “class consciousness”, including the progressive vision of the proletariat. And in orthodox Marxism-Leninism, “ideology” refers to any form of social consciousness, including “scientific” accounts of the world supposedly free of mystification. Hence, many twentieth-century Marxists write of “proletarian”, “Marxist” and “scientific” ideologies in a way that Marx himself would have found nonsensical.

Another common twentieth-century use of the term “ideology” is found in American social science of the 1960s, especially in the work of Daniel Bell and the “end of ideology”

theorists. For these thinkers, an ideology is a programme for the radical transformation of social life based on an explicit theory of society and embodying certain clearly articulated values. Ideologies function as forms of "secular religion"; their proponents are obsessed with doctrinal purity, fidelity to the pronouncements of canonized authorities, and unquestioning obedience to charismatic leaders. After the defeat of Nazism, and with the apparent "thaw" in post-Stalinist Russia, Bell and others suggested that the preeminence of such "total ideologies" on the world political stage was finally over, to be replaced by the politics of liberal compromise. (Bell's critics were quick to point out the ideological role his own theories played as a contribution to the American liberalism of the Cold War.)

Whatever the merits of the original "end of ideology" thesis, the recent demise of Marxism-Leninism in Eastern Europe indeed represents the passing of one of the most powerful of this century's ideologies (in Bell's sense). However, while the age of the "total ideology" may indeed be waning, the significance of the concept of ideology has not diminished in the social sciences. Indeed, recent years have seen a renewed interest, particularly in the areas of feminism, post-modernism and cultural studies, in the project of unmasking relations of power and domination implicit in culturally dominant forms of theoretical and social discourse. While much of this work does not explicitly invoke the notion of ideology as such, there is no question of its debt to the critical project Marx began.

See also SOCIAL SCIENCES.

#### BIBLIOGRAPHY

- Bell, D.: *The End of Ideology*, 2nd edn (New York: Collier, 1962).  
 Geuss, R.: *The Idea of a Critical Theory* (Cambridge: Cambridge University Press, 1981), esp. ch. 1.  
 Gramsci, A.: *Selections from the Prison Notebooks*, ed. Q. Hoare and G. Nowell Smith (London: Lawrence and Wishart, 1971).

- Habermas, J.: *Legitimation Crisis* (London: Heinemann, 1976).  
 Kolakowski, L.: *Main Currents of Marxism* (Oxford: Oxford University Press, 1978), vol. 1, ch. 8.  
 Lenin, V. I.: "What Is to Be Done?," in his *Collected Works* (Moscow: Foreign Language Publishing House, 1960–3; Progress, 1964–70), vol. 5.  
 Lichtheim, G.: "The Concept of Ideology," in his *The Concept of Ideology and Other Essays* (New York: Random House, 1967).  
 Lukács, G.: *History and Class Consciousness* (London: Merlin, 1971).  
 Mannheim, K.: *Ideology and Utopia* (New York: Harcourt, Brace and World, 1936).  
 Marx, K. and Engels, F.: *The German Ideology*, in K. Marx and F. Engels, *Collected Works* (London: Lawrence and Wishart, 1976), vol. 5.  
 Shils, E.: "The Concept and Function of Ideology," in *International Encyclopedia of the Social and Sciences*, ed. D. L. Sills (London: Macmillan and the Free Press, 1968), 66–76.

DAVID BAKHURST

**illusion** see ARGUMENT FROM ILLUSION.

**immediacy, presence** The concepts of immediacy and presence refer to an alleged contrast between our knowledge of our own mental states on the one hand, and, on the other hand, our knowledge of physical objects and the mental states of others. Knowledge of the mental states of others is inferred from their behavior. Knowledge of physical objects may not be inferred, but it is acquired via complex causal chains involving spatially intermediate causes, and if challenged, it is justified by appeal to appearances. Knowledge of our own mental states is said to be known without such causal chains, and beliefs about these states are not justified by appeal to properties of a different sort. As Alston (1971) points out, the claim about causal immediacy is problematic, since we are ignorant of the causal antecedents involved in knowing our own mental states. However, the justification of beliefs about certain of our mental states does seem to appeal only to those states themselves.

See also BRENTANO; EXPERIENCE; the GIVEN; KNOWLEDGE BY ACQUAINTANCE/BY DESCRIPTION.

## BIBLIOGRAPHY

- Alston, W.: "Varieties of Privileged Access," *American Philosophical Quarterly* 8 (1971), 223–41.
- Brentano, F.: *Psychology from an Empirical Standpoint*, trans. A. C. Rancurello, D. B. Terrell and L. L. McAlister (New York: Humanities Press, 1973), bk 2, ch. 1.
- Lewis, C. I.: "Some Logical Considerations Concerning the Mental," in *Body and Mind*, ed. G. Vesey (London: Allen and Unwin, 1964), 330–7.

ALAN H. GOLDMAN

**in itself/for itself** The distinction between the "in itself" and the "for itself" originated in the Kantian logical and epistemological distinction between a thing as it is in itself, and that thing as an appearance, or as it is *for us*. For Kant (see KANT), the thing in itself is the thing as it is intrinsically, that is, the character of the thing apart from any relations in which it happens to stand. The thing for us, or as an appearance, on the other hand, is the thing in so far as it stands in relation to our cognitive faculties and other objects. "Now a thing in itself cannot be known through mere relations; and we may therefore conclude that since outer sense gives us nothing but mere relations, this sense can contain in its representation only the relation of an object to the subject, and not the inner properties of the object in itself" (CPR, B. 67). Kant applies this same distinction to the subject's cognition of itself. Since the subject can know itself only in so far as it can intuit itself, and it can intuit itself only in terms of temporal relations, and thus as it is related to its own self, it represents itself "as it appears to itself, not as it is" (CPR, B. 69). Thus, the distinction between what the *subject* is in itself and what it is for itself arises in Kant in so far as the distinction between what an object is in itself and what it is for a knower is applied to the subject's own knowledge of itself.

Hegel (see HEGEL) begins the transition of the epistemological distinction between what the subject is in itself and what it is for itself into an ontological distinction. Since, for Hegel, what is, as it is in fact or in itself, necessarily

involves relation, the Kantian distinction must be transformed. Taking his cue from the fact that, even for Kant, what the subject is in fact or in itself involves a relation to itself, or self-consciousness. Hegel suggests that the cognition of an entity in terms of such relations or self-relations does not preclude knowledge of the thing itself. Rather, what an entity is intrinsically, or in itself, is best understood in terms of the potentiality of that thing to enter into specific explicit relations with itself. And, just as for consciousness to be explicitly itself is for it to be for itself by being in relation to itself (i.e. to be explicitly self-conscious), the for itself of any entity is that entity in so far as it is actually related to itself. The distinction between the entity in itself and the entity for itself is thus taken to apply to every entity, and not only to the subject. For example, the seed of a plant is that plant in itself or implicitly, while mature plant which involves actual relations among the plant's various organs is the plant "for itself". In Hegel, then, the in itself/for itself distinction becomes universalized, in that it is applied to all entities, and not merely to conscious entities. In addition, the distinction takes on an ontological dimension. While the seed and the mature plant are one and same entity, the being in itself of the plant, or the plant as potential adult, is ontologically distinct from the being for itself of the plant, or the actually existing mature organism. At the same time, the distinction retains an epistemological dimension in Hegel, although its import is quite different from that of the Kantian distinction. To know a thing it is necessary to know both the actual, explicit self-relations which mark the thing (the being for itself of the thing) and the inherent simple principle of these relations, or the being in itself of the thing. Real knowledge, for Hegel, thus consists in a knowledge of the thing as it is in and for itself.

Sartre's (see SARTRE) distinction between being in itself and being for itself, which is an entirely ontological distinction with minimal epistemological import, is descended from the Hegelian distinction. Sartre distinguishes between what it is for consciousness to be, (i.e. being for itself), and the being of the



transcendent being which is intended by consciousness (i.e. being in itself). Being in itself is marked by the total absence of relation, either with itself or with another. On the other hand, what it is for consciousness to be, being for itself, is marked by self relation. Sartre posits a "pre-reflective Cogito", such that every consciousness of *x* necessarily involves a "non-positional" (in Husserlian terms, "horizontal") consciousness of the consciousness of *x*. While in Kant every subject is both in itself, i.e. as it is apart from its relations, and for itself in so far as it is related to itself by appearing to itself, and in Hegel every entity can be considered as it is both in itself and for itself, in Sartre, to be self related or for itself is the distinctive ontological mark of consciousness, while to lack relations or to be in itself is the distinctive ontological mark of non-conscious entities.

*See also* NOUMENAL/PHENOMENAL.

#### BIBLIOGRAPHY

- Hegel, G. W. F.: *The Phenomenology of Spirit* (1807), trans. A. V. Miller (Oxford: Clarendon Press, 1977).  
 Hegel, G. W. F.: *Hegel's Science of Logic* (1812–16, 1831), trans. A. V. Miller (London: Allen and Unwin, 1969).  
 Kant, I.: *Critique of Pure Reason* (1781), trans. N. Kemp Smith (London: Macmillan, 1964).  
 Sartre, J.-P.: *Being and Nothingness*, trans. H. Barnes (New York: Philosophical Library, 1956).  
 Taylor, C.: *Hegel* (Cambridge: Cambridge University Press, 1975).

M. OKRENT

**incorrigibility** Etymologically, "incorrigibility" is most aptly used for the impossibility of one's being refuted, "corrected", shown to be mistaken, in a belief, though there is a persistent misuse of the term for the impossibility of being mistaken (*see* INFALLIBILITY). As with infallibility, incorrigibility is most often asserted of one's beliefs about one's own current states of consciousness. It is plausible to think that a person is the final authority on what s/he is thinking, feeling, or sensing at the

moment, and that his/her sincere report will always prevail against any counter-evidence (Ayer, 1963, ch. 3). On the other hand, it seems plausible to suppose that physiological psychology could develop to the point that we would be justified in preferring neural readings to the subject's report where they conflicted (Armstrong, 1963).

*See also* bibliography of INFALLIBILITY.

#### BIBLIOGRAPHY

- Armstrong, D. M.: "Is Introspective Knowledge Incorrigible?," *Philosophical Review* 72 (1963), 417–32.  
 Ayer, A. J.: *The Concept of a Person* (New York: St Martin's Press, 1963).

WILLIAM P. ALSTON

**indeterminacy of reference** The "inscrutability" or "indeterminacy" of reference (Quine, 1990, p. 50) is a thesis of Quine's (*see* QUINE) related to the indeterminacy of translation (*see* INDETERMINACY OF TRANSLATION): there is no fact of the matter what kinds of objects the terms of a language refer to. By making compensating adjustments in the translations of other words in a sentence, a predicate might be interpreted as applying to rabbits, to stages in the history of rabbits, to volumes of space one mile north of rabbits, and so on. Such indeterminacy is compatible with determinacy in the translation of sentences as a whole (Quine, 1990, p. 50).

*See also* ONTOLOGICAL RELATIVITY.

#### BIBLIOGRAPHY

- Davidson, D.: *Inquiries into Truth and Interpretation* (Oxford: Oxford University Press, 1984), ch. 16.  
 Quine, W. V.: *Word and Object* (Cambridge, MA: MIT Press, 1960), ch. 2.  
 Quine, W. V.: *Pursuit of Truth* (Cambridge, MA: Harvard University Press, 1990), chs 2–3.

CHRISTOPHER HOOKWAY

**indeterminacy of translation** Radical translation is the attempt to understand a hitherto unencountered language without the aid of dictionaries or knowledge of related tongues. Quine (*see* QUINE) argues that such translation must exploit information about the stimulus conditions under which utterances are accepted, and he holds that alternative incompatible translations will be compatible with all such evidence. He concludes that there is no fact of the matter what any expression of the alien language means, and thus that a naturalistic philosophy cannot make sense of objective semantic facts. This reinforces his opposition to the analytic/synthetic distinction and to philosophical use of intensional notions. If effective, the argument also challenges the objectivity of intentional notions, propositional attitudes like belief and desire. Critics have claimed that he ignores much of the evidence relevant to translation although a version of the doctrine has been defended by some, like Davidson, who dispute Quine's empiricism and naturalism.

*See also* DAVIDSON.

#### BIBLIOGRAPHY

- Quine, W. V.: *Word and Object* (Cambridge, MA: MIT Press, 1960), ch. 2.  
 Quine, W. V.: *Pursuit of Truth* (Cambridge, MA: Harvard University Press, 1990), ch. 3.

CHRISTOPHER HOOKWAY

**Indian epistemology** The Sanskrit synonym for "cognition" is "*jñāna*", and for "knowledge" is "*pramā*". As is then to be expected, "*jñāna*" applies to all cognitive states including true and false cognitions as well as doubt (*saṃśaya*) and mere conceptual thinking (*kalpanā*). *Pramā* is restricted to true cognition. Theories of knowledge, in the Indian tradition, are concerned, first, with (1) cognitions in general, (2) specifically, with true cognitions, and also (3) with false cognitions. The abstract noun "truth" translates into "*pramātva*", also into "*prāmānya*".

The instrumental cause of a true cognition is called "*pramāna*". A *pramāna* theory is one which theorizes about the (causal) modes by which true cognitions are acquired and their truth claims validated.

In India, the *pramāna* theories seem to have come into being in order to answer burgeoning sceptical tendencies. Sanjaya (pre-Buddhist) raised sceptical questions about religious, ethical and eschatological beliefs. Nāgārjuna (second century AD) questioned the very foundations of the distinction between means of true cognition and objects of such cognition. These sceptical arguments were later – by which time, the *pramāna* theories had been considerably developed – strengthened and made more pointed by Jayarasi (eighth century AD) and Sri Harsa (eleventh century AD). In order to show that definitions of knowledge in terms of correspondence or in terms of appropriate causal chains won't do, Śrī Harsa cites cases like a lucky guess of a gambler, a true conclusion derived from false premises (mistaking a column of dust for smoke, one infers the presence of fire which happens to be there), or inferring rightly that the animal over there is a bull from observing his horns where the real horns of the animal had been sawed off and replaced by imitation ones.

The Indian epistemologists asked certain, standard questions. These questions may be divided into two groups:

- A: questions concerning any cognitive state or episode; and
- B: questions concerning those cognitions which have a truth value.

A. With regard to any cognitive state or episode (irrespective of its truth-value), two major questions were asked, leaving out of consideration ontological questions regarding the mode of being of cognitions (e.g. are they substances, qualities or actions?): (1) do cognitions have a form ("*ākāra*") intrinsic to them or are they intrinsically formless, deriving their seeming form ("blue", "yellow", etc.) from their objects? and (2) are cognitions themselves known at the very moment

of their occurrence, or are they only subsequently known by another cognition? As regards (1), the Buddhists defended the thesis that the cognition “This is blue” really has “blue” as its form (or content), while the Nyāya philosophers argued that the cognition “This is blue” is itself formless, it is only *of* a blue thing which is out there in the world. However, according to the Nyāya, although the cognition of blue is not itself blue and does not have “blue” as a *real* form, its logical analysis requires bringing out its intentional contents, called its qualifiers (*prakāratā*) which indeed constitute its total logical content. Thus the cognition expressed in the sentence “This is blue” has a qualificandum-content “this”, which is determined by a qualifier-content “thisness”, another qualificandum-content “blue” which again is determined by a qualifier-content “blueness”. In general, on the Nyāya view, a cognition must have a content that is not linguistically expressed (in the above example, “thisness” and “blueness” are not), and it must have a content which is only a qualificandum and not also a qualifier (in the above example, “this” is such a content).

Over issue (2), likewise, the Buddhists, the Prābhākara Mīmāṃsakas and the Vedāntins insisted – overlooking the differences between them – that a cognition is always self-revealing (*svaprakāśa*); it is known by its mere presence, so that it is never the case that I know something but do not know that I know (see KK-THESIS). The Nyāya defended the opposite thesis according to which if  $K_1$  is knowledge of the object  $o$  at time  $t_1$ ,  $K_1$  itself is not known at  $t_1$ ,  $K_1$  can be known, and is usually known, by another cognition  $K_2$  occurring at the succeeding moment  $t_2$ .

B. Keeping these two issues in mind, let us focus upon the concepts “true cognition”, “means of true cognition” and “truth”. The specific questions to be discussed are:

- (1) What are the different, not further reducible, types of true cognition? and, what are the instrumental causes of each?
- (2) How is “truth” to be defined?
- (3) How is truth known?

(1) The list of irreducible types of true cognition ranged, in the history of Indian thought on these matters, from perception alone (the Cārvākas), to the most liberal Vedānta list which includes perception, inference, word, comparison, postulation and non-perception. In between these are the Vaiśeṣikas who admitted only perception and inference, and the Nyāya which would have the first four alone from the Vedānta list. The uniqueness of a type of cognition (and so its admissibility on to the list) required that it had a unique set of causal conditions, and that it be not reducible to any other. In this essay we shall consider only perception, inference and what is called ‘word’ (*śabda*).

(a) Perception (*pratyakṣa*) is defined (*Nyāyasūtra* 1.1.4) as the cognition which is caused by contact of the (appropriate) sense organs with (their) objects. There were two extreme views about what perceptual cognition is like. At the one end, there was the Buddhist view that if it is caused by the object, the perceptual cognition must be non-linguistic and non-conceptual, it can only be an instantaneous and ineffable awareness of the mere *this*. At the other extreme, is the view of the Grammarians (chiefly, Bhartrhari, fifth century AD) that all cognition, perception included, is linguistic. The Nyāya prefers an intermediate position: perceptual cognition is initially non-linguistic and non-conceptual, but this initial phase is soon replaced by a linguistic and conceptual cognition which is still perceptual, though now a perceptual judgement. A distinctive feature of the Nyāya view is that we perceive not only physical objects and their sensuous qualities, but also universals, relations and even absences. Looking at my blue pencil, I perceive the *thing* (*dravya*) pencil, the blue-particular inhering in that pencil, the universal blueness inhering in the blue-particular, (under certain conditions) the absence of red colour in the pencil, and the relation of inherence obtaining between the pencil and its colour – in each case, the visual sense-organ is in contact with the object in an appropriate relation (the Nyāya has here a ramified theory of relations at its disposal). I have also a

perceptual cognition of my own self, which is expressed in judgements such as “I am happy”. Some other schools of philosophy do not accept the Nyāya account of perception. The Jains define perception by its “clarity and distinctness” (*vaiśadya*). The Advaita Vedāntins regard perception as being of the nature of consciousness, meaning that the only thing that is directly evident is one’s consciousness, everything else is given only as its content. The Nyāya view, however, was widely accepted and was most developed in detail.

(b) Inferential cognition (*anumiti*) is introspectively recognized by the cognizer as such, i.e. she says “I am inferring”. It is caused by a succession of cognitive episodes: first, I see a column of smoke rising from the top of the yonder mountain. On seeing it, I remember the law-like relation (*vyāpti*) “wherever there is smoke, there is fire” which I had previously learnt, and which is instantiated in the familiar and uncontroversial case of the stove in the kitchen. This remembrance makes me now see the smoke on the mountain top as a mark of fire. This last perceptual cognition (called by the Nyāya “*parāmarsa*”) gives rise to the inferential cognition “Therefore, there is fire on the mountain”. Let us for the present purpose overlook the large discussions in the literature regarding the precise number and nature of these cognitive episodes which culminate in inferential cognition. What is noteworthy is that here we have a psychology of inference which is made to serve the purposes of a logic of inference. It is interesting and important to ask if this does not entail the sort of psychologism (see PSYCHOLOGISM) which Frege (see FREGE) and Husserl (see HUSSERL) wanted to overcome: brief remarks will be made about this question at the end of this essay.

(c) *śabda*: What is most distinctive of Indian epistemologies is the recognition of a type of cognition which the auditor (or reader) acquires upon hearing (or reading) the sentence or sentences uttered (or written) by a speaker (or writer) who is intellectually competent (i.e. knows what he is talking about) and morally honest (i.e. is not lying, does not want to mislead, etc.). Those who did not want to recognize such an irreducible class of

knowledge generally sought to reduce the cases of such cognition to cases of inference, but it was easy to show that the accepted theories of inference stipulated requirements which the alleged inference (to which the word-generated cognition was to be reduced) could not satisfy. Examples of word-generated cognition are our knowledge of the past derived from reading history books, of contemporary events from reading newspapers, of moral rules and supersensible truths from reading the scriptures. This sort of cognition presupposes that the auditor/reader knows the meaning of the words that make up the sentence. Over and above that, each of the words needs to arouse a semantic expectancy (*ākāṅkṣā*) pointing to what follows; the words must have semantic fitness (*yoqyatā*) (which rules out “virtue is green” from producing a cognition); the words must succeed each other contiguously (in space and/or in time); and there must be an understanding of what the speaker/writer intends to convey. Granted that one understands a sentence *p* uttered by a competent speaker who is also known to be competent – so the theory goes – this *understanding* of the meaning of *p* itself amounts to *knowing* that *p*. The theory then rules out the possibility of understanding a false sentence, and has to deny any theory of “sense” as distinguished from “reference”.

(2) “Truth” (*prāmāṇya*): There are three kinds of definition of “truth” in Indian theories of knowledge. One group of definitions define “truth” in terms entirely of epistemic concepts, of features intrinsic to a cognition. For example, according to one such definition, a cognition is true if it has a qualifier which is possessed by its qualificandum. Since “qualifier” and “qualificandum” are epistemic features, on this and kindred definitions (advanced by the Prābhākara School of *Mīmāṃsā*) every cognition is true, there is no (unitary) false cognition, the seemingly false cognition being a complex of cognitions each of which is true, but which are not distinguished from each other. Another group of definitions take a cognition to be true as long as it is not known to be false, or is not contradicted: on this view then “truth” cannot be “established”, “falsity” however can, and a

cognition is *eo ipso* taken to be true as long as it is not determined to be false. In this sense again, truth is intrinsic to knowledge. Neither of these two sets of definitions can make room for the possibility of determining the truth or the falsity of a cognition. For a more satisfactory account, then, we turn to the Nyāya theory which unlike the first group makes truth consist of both epistemic and ontological features (so that merely inspecting the internal epistemic structure of a cognition leaves it undecided if the cognition is true or false), and unlike the second group, makes it possible to *ascertain* if a cognition is true or if it is false by advancing, beside the definition of “truth”, also a set of marks or tests of truth. First the definition:

Let the cognition under consideration be “S is F”. This cognition has among others two qualifiers, “S-ness” and “F-ness”. The cognition is true if these qualifiers belong to the *thing* (not merely the epistemic object) that is allegedly known by that cognition. This celebrated definition of “*prāmānya*” as *tadvati tatprakāraakatva*, was first advanced by Gangeśa (fourteenth century) and refined and elaborated by a whole series of illustrious commentators (see Mohanty, 1967).

The Nyāya distinguishes the definition of “truth” from the test of truth. Truth is ascertained either through the ability of a cognition to lead to successful practice or through consensus. Most Indian philosophers, except the Vedāntins, have accepted some notion of pragmatic workability as a *test* of truth. The Vedāntins who reject this test – for, on their view, there is no test of truth, one can only test falsity – contend that within the context of a dream dream-water may satisfy dream-thirst, and so may indeed work.

With regard to the last of the issues stated at the beginning, we have two major positions:

On the Nyāya view, a cognition is intrinsically neither true nor false. Both truth and falsity are “extrinsic” possibilities, depending upon (1) how it relates to the ontological reality that is allegedly being cognized, and (2) if in addition to the generic causal factors for the sort of cognition, there is a special excellence (*guna*) or a special defect (*dosa*).

On the Vedāntin’s view, a cognition is intrinsically true (= taken to be true), unless and until it is proved to be otherwise, thus only falsity is extrinsic. Gangeśa argues that if this latter view were correct, immediately after I have a cognition, I could not have the *doubt* “Is my cognition true or false?” The Vedāntins argue that if a cognition were not taken to be true, it could not stimulate appropriate practical response, to which Gangeśa replies that in most cases what is needed for such response is not truth-determination but absence of doubt about falsity.

*General features:* The Indian epistemological theories are characterised by the following general features:

1. The analysis of a cognition is made under the “guidance” of both its linguistic expression and one’s phenomenological experience: the two constraints supplement and correct each other.
2. A cognition is construed as an episode in the inner life of the cognizer. Thus it is individuated by an ego and a temporal location. However, a cognition has a logical structure which allows for being exemplified in another numerically distinct episode belonging to another ego and/or another temporal location. This avoids the ruinous consequences of psychologism.
3. The theories are predominantly causal (see CAUSAL THEORIES IN EPISTEMOLOGY). The *pramānas* are construed both as (specific) causes of (specific sorts of) true cognitions and also as providing evidence and justificatory ground for them.
4. There is an overall fallibilism (see FALLIBILISM). There is no cognition for which error is ruled out on logical grounds (unless, it is by definition, as with the Mīmāṃsā theories referred to in B (2)).
5. The distinction between analytic (see ANALYTICITY) and synthetic truths either is not to be found, or at most, remains on the distant horizon.
6. The opposition between rationalism (see RATIONALISM) and empiricism (see EMPIRICISM) never made its appearance. The question whether perception is the only



*pramāna* or inference and *śabda* are also to be counted amongst *pramānas* is not to be construed as the question whether experience is the only source of knowledge or not. Neither is perception (*pratyaksa*) the same as "experience" (one also perceives, according to Nyāya, universals) nor is "inference" (*anumana*) the same as "reason". Strictly speaking no concept of a priori knowledge is available.

# BIBLIOGRAPHY

- Chatterjee, S. C.: *Nyaya Theory of Knowledge* (Calcutta: Calcutta University Press, 1978).
- Datta, D. M.: *Six Ways of Knowing* (Calcutta: Calcutta University Press, 1972).
- Hattori Masaaki, D.: *On Perception* (Cambridge, MA: Harvard University Press, 1968).
- Jayatilke, K. N.: *Early Buddhist Theory of Knowledge* (London: Allen and Unwin, 1963).
- Matilal, B. K.: *Perception: An Essay on Classical Indian Theories of Knowledge* (Oxford: Clarendon Press, 1986).
- Mohanty, J. N.: *Gangesa's Theory of Truth*, 2nd edn revised (Delhi: Motilal Banarasidass, 1989).
- Potter, K. H., ed.: *Encyclopedia of Indian Philosophies*, vol. 2, *Indian Metaphysics and Epistemology: The Tradition of Nyaya-Vaisesika up to Gangesa* (Princeton, NJ: Princeton University Press, 1977).

J. N. MOHANTY

**indubitability** "Indubitability" means, of course, not being subject to doubt, or, modally expressed, the impossibility of doubt. Think of the term as applied to a particular bit of knowledge or a particular belief, for example, my knowledge that the wind is blowing hard here and now. We may distinguish at least three importantly different understandings of indubitability: (1) a psychological impossibility of entertaining a doubt: (2) a logical impossibility of entertaining a doubt (it is sometimes held that no sense can be attached to a person's being in doubt as to whether s/he is having a certain sensation); (3) the impossibility of there being any (real) grounds for doubt (Alston, 1989, ch. 10). Each of these has been supposed by various

philosophers to attach to knowledge of self-evident truths or of one's current conscious states, though it is (3) that is of the most obvious and indubitable epistemological significance. Descartes (see DESCARTES) is a philosopher who notoriously put great stock in indubitability (of the third sort) as necessary for knowledge.

# BIBLIOGRAPHY

- Alston, W. P.: *Epistemic Justification* (Ithaca, NY: Cornell University Press, 1989).

WILLIAM P. ALSTON

**induction, problem of** see PROBLEMS OF INDUCTION.

**induction: enumerative and hypothetical**

# KINDS OF INDUCTION

In *enumerative induction*, a generalization is inferred from evidence about instances of the generalization. Suppose, for example, that Sam has examined several peaches and found pits in them and has not found any peaches without pits. Making an enumerative induction, Sam concludes that all peaches have pits.

Sam's inference is non-demonstrative; its conclusion does not follow necessarily from his evidence. In contrast, a valid deductive argument is demonstrative. The conclusion of a valid deductive argument is guaranteed to be true if the premises are true. The premises, (P1) "All peaches have pits", and (P2) "X is a peach", logically entail the conclusion (C) "X has a pit". (C) cannot be false if both (P1) and (P2) are true. But Sam's evidence, consisting of instances of peaches with pits, does not guarantee the truth of his conclusion that all peaches have pits. Other peaches, not yet examined, could conceivably fail to have pits.

This familiar contrast between non-demonstrative induction and demonstrative deduction is useful here, but it can also be misleading, as we shall see.

In *hypothetical induction*, a hypothesis is inferred as the best explanation of the evidence. Seeing the tips of a man's shoes peeping out from under the curtain, Sam concludes that a man is hiding behind the curtain. Again Sam's inference is non-demonstrative. The shoes might be empty shoes that just happen to have ended up pointing out from under the curtains (*see INFERENCE TO THE BEST EXPLANATION*).

Induction in the widest sense includes any non-demonstrative inference, including non-demonstrative predictive inferences and default assumptions. Given that Albert now plans to be in Pittsburgh tomorrow, Sam infers that Albert will be in Pittsburgh tomorrow: this is a non-demonstrative predictive inference because Sam realizes that his evidence concerning Albert's present plan does not guarantee that Albert will carry out his plan.

This is not enumerative induction, since Sam's conclusion is not a generalization of his evidence. It is not hypothetical induction, because Sam's conclusion is not a hypothesis that might explain his evidence – Sam does not suppose that Albert's being in Pittsburgh tomorrow explains Albert's current plan to be there tomorrow. Instead he makes the converse supposition, namely, that Albert's having this plan explains why Albert will be in Pittsburgh tomorrow.

Given that Albert is a person, Sam assumes that Albert can walk up some stairs, so Sam invites Albert to visit his third-floor apartment; here Sam relies on a default assumption that could be incorrect, since it is possible that Albert cannot walk up stairs even though he is a person. Induction in the widest sense includes such default assumptions (Ginsberg, 1987).

Induction in this sense is, of course, to be sharply distinguished from so-called mathematical induction, which is a form of demonstrative proof. (Mathematical induction is the following principle. Suppose you can show that a claim about natural numbers (0, 1, 2, . . . ) holds for zero, and you can show that, for any natural number  $n$ , if the claim holds for all natural numbers less than or equal to  $n$  the claim also holds for  $n$ . Then you can

conclude that the claim holds for all natural numbers.)

In all cases of non-demonstrative induction, the evidence is compatible with several hypotheses. We can only draw a conclusion if it is more reasonable to accept one of these hypotheses rather than another.

Consider Sam's evidence that several peaches have been found to have pits and no peaches have been found not to have pits. That evidence is compatible with Sam's hypothesis that all peaches have pits and also with the hypothesis that only some peaches have pits but Albert has been removing all the peaches without pits before Sam sees them.

#### THE BASIC RIDDLE

An important question about induction is how people are to decide among the hypotheses that are compatible with the evidence. Perhaps there are two questions here: (1) How do people decide among such hypotheses? and (2) How should people decide among such hypotheses? A sceptical answer to (2) is that people should not decide; people can have no reason to believe that one rather than another hypothesis is true, if both hypotheses are compatible with the evidence. A non-sceptical answer to (2) is that people are justified in deciding among hypotheses in whatever way it is they do decide, at least until they are presented with special reasons not to decide among hypotheses in that particular way. Further pursuit of this non-sceptical approach requires trying to answer the first question about how people actually do decide among those hypotheses that are compatible with the evidence.

After examining a few peaches, Sam concludes that all peaches have pits. When asked why he prefers that conclusion to the hypothesis that, although there are peaches without pits, Albert has been removing those peaches before Sam sees them, Sam responds that he hasn't even considered the other hypothesis and has no reason to consider it since there is no reason to suspect Albert of removing peaches without pits.

Consider the reply that there is reason to suspect Albert of removing such peaches, namely, the very evidence that Sam is relying on! Given that the evidence is compatible with both possibilities, is there any more reason for Sam to take the evidence to support his conclusion, that all peaches have pits, than there is for him to take the evidence to support the hypothesis that Albert has been removing the peaches that have no pits?

Sam might answer that there is more reason to accept his conclusion than to accept the suggested competitor: namely, his conclusion is simpler than the suggested alternative, and it leaves him with fewer questions to answer. Although his conclusion leaves him with the question why all peaches have pits, the alternative hypothesis not only leaves him with a similar question, namely, why some peaches do and some peaches do not have pits, but also raises additional questions such as why Albert should be removing the peaches without pits and how Albert would be able to tell when peaches don't have pits so that he could screen them out.

People prefer simpler hypotheses that raise fewer questions to more complex hypotheses that raise more questions. But it is not easy to say what makes one hypothesis simpler than another and what it is for a hypothesis to raise further questions.

So far, we have been considering competition between an enumerative induction and a hypothetical induction – an enumerative induction to the generalization, “All peaches have pits”, and a hypothetical induction to the conclusion “Albert is removing peaches with pits before Sam can see them”. Goodman (1965) has stressed that there are also competing enumerative inductions. Let  $t$  be the present time and consider the following generalization, “All peaches either have been examined by Sam at or before  $t$  and have pits or have not been examined by Sam at or before  $t$  and do not have pits”. Let us abbreviate this generalization as “All peaches are pitst”, where by definition something is pitst if and only if, either Sam examines it at or before  $t$  and it has a pit or Sam does not examine it at or before  $t$  and it does not have a pit. The evidence for this generalization

parallels the evidence for Sam's generalization. Just as Sam has found several peaches to have pits and has not found any to be without pits, Sam has found several peaches to be pitst and has not found any not to be pitst. In a case like this, what distinguishes a reasonable generalization, like “All peaches have pits”, from an unreasonable generalization, like “All peaches are pitst”?

Goodman calls this the “new riddle” of induction. The problem is not to justify our practice of preferring the first sort of generalization to the second. The problem is to characterize our practice by specifying those features of these generalizations that can lead us to infer one conclusion rather than any of its competitors (*see* GOODMAN; PROBLEMS OF INDUCTION; PROJECTION, PROJECTIBILITY).

Goodman's own answer is that, other things being equal, we prefer generalizations using “entrenched” predicates to generalizations using less entrenched predicates, where a predicate is more entrenched the more we have used it as a principal predicate in previous inductions. Sam has used “has a pit” in previous inductive conclusions and has never used “is pitst”, i.e. “is such that, either Sam examines it at or before  $t$  and it has a pit or Sam does not examine it at or before  $t$  and it does not have a pit”. So the former predicate is better entrenched than the second and Sam is led to accept the generalization “All peaches have pits”, rather than its competitor “All peaches are pitst”.

So far so good. But what if Sam is a young child who has never before made inferences about things having pits? In that case, Goodman's answer is not available, although Sam will still be more inclined to infer that all peaches have pits rather than that all peaches are pitst. What distinguishes these hypotheses for Sam at this stage?

#### SIMPLICITY

Simplicity (*see* SIMPLICITY) may play a role, but not just the simplicity with which a hypothesis is stated, since competing hypotheses can always be expressed equally simply. Any hypothesis can be given an arbitrarily

simple formulation by using a single symbol, e.g. *H*, to stand for that hypothesis.

It is possibly relevant that the favoured hypothesis, "All peaches have pits", is much easier for Sam to comprehend than the other hypothesis, "All peaches are pitst", which Sam has to understand as "All peaches are such that either Sam examines them at or before *t* and they have pits or Sam does not examine them at or before *t* and they do not have pits".

More significantly, the psychological relation for Sam between Sam's first hypothesis and his evidence is much simpler than the relation between this second hypothesis and his evidence. It is easier for Sam to understand how his favoured hypothesis can account for his finding some peaches with pits and none without pits than it is for Sam to understand how the second hypothesis would account for this evidence. The argument from his favoured hypothesis to the evidence is briefer and more direct than the argument from the other hypothesis to his evidence.

This is so, at any rate, if Sam's evidence is expressed as we have just expressed it. The matter would be otherwise if the evidence were described as follows: "Sam has found some peaches to be pitst and has not found any peaches not to be pitst." If the evidence were described that way, then it would be simpler to account for the evidence using the hypothesis that all peaches are pitst than it would be to account for the evidence using the hypothesis that all peaches have pits. But Sam is not interested in describing his evidence in this second way. He is interested in the fact that he has found peaches with pits and has not found peaches without pits. That is what he wants to account for. He is not directly interested in accounting for the fact that he has found peaches to be pitst, etc. (If Sam were unusually precocious, he might be interested in accounting for the fact that all peaches are pitst to the extent that such an account would help to explain why he has found peaches to have pits.)

So one thing that may lead Sam to distinguish these hypotheses in the way that he does is a concern to have the simplest account for aspects of evidence in which he is interested.

Notice that, in this view, the simplest, most inferable hypothesis is not to be identified with the hypothesis with the simplest natural formulation. It is the simplicity of the connection between that hypothesis and the evidence that is important, not just the simplicity of the hypothesis taken by itself (if that even makes sense).

A contrary view would hold that the only relevant question about the connection between evidence and hypothesis is whether the hypothesis is compatible with the evidence. In this contrary view, Sam should be concerned to find the simplest hypothesis compatible with the evidence. But that could not be right. The evidence might be completely irrelevant to the simplest hypothesis compatible with it. An ideally simple true hypothesis, such as "Everything exists", would on this view pre-empt and prevent any other inductive inference, since it would always be the simplest hypothesis compatible with the evidence!

We would need at the very least to restrict attention to competing hypotheses: find the simplest of competing hypotheses compatible with the evidence. But that would still be unacceptable. Given several competing hypotheses, there can be evidence for one that does not rule out the others.

Sam's hypothesis that all peaches have pits competes with the hypothesis that Albert is removing the peaches with pits before Sam sees them and there may come to be evidence for this second hypothesis that does not absolutely rule out the first hypothesis. Sam may notice that Albert always does something to the peaches before Sam gets to examine them. He may discover that Albert has recently bet Alice that Sam will never find a peach without a pit, something that gives Albert a motive to remove peaches without pits. He may learn that Sam has a pit-detecting machine that can be used to determine whether or not a given piece of fruit has a pit in it. Enough evidence of this sort may eventually lead Sam to conclude that the second hypothesis is correct – Albert is removing peaches without pits before he, Sam, sees them. Sam may reach this conclusion even though the evidence remains compatible

with both hypotheses so that there is no change in which is the simpler hypothesis compatible with the evidence.

#### COHERENCE

For evidence that is compatible with several competing hypotheses to support one hypothesis over the others, there has to be some further special connection between the evidence and that particular hypothesis that does not hold between the evidence and the others. The hypothesis must cohere better with that evidence than the competing hypotheses do.

A potential conclusion can relevantly cohere with evidence by helping to account for the evidence, as for example "Albert is removing peaches without pits" would help to account for Sam's complex evidence. But there are other cases too, as in predictive induction. When Sam concludes from Albert's intention to be in Pittsburgh tomorrow that Albert will be in Pittsburgh tomorrow, Sam's conclusion does not help to account for his evidence. Things are the other way round: Sam infers that Albert's having that intention will help to account for his being in Pittsburgh.

Coherence is not an isolated relation between Sam's hypothesis and the evidence for it. Other things that Sam accepts are relevant also. Sam's conclusion that Albert will be in Pittsburgh tomorrow coheres with his evidence only given Sam's beliefs about how intentions can lead to action. What is really relevant, we might say, is how the acceptance of this conclusion will affect the overall coherence of Sam's view.

It can be difficult to achieve total overall coherence even in science. Sometimes there are two or more large internally coherent disciplines that do not cohere with each other. For example, in nineteenth-century science, thermodynamics implied that the earth was younger than the biological theory of evolution held it to be.

A further complication is that evidence is sometimes left unaccounted for or even abandoned as "observational error" in the interest of a larger coherence. Favoured

background assumptions may be given up for similar reasons. Inference is really a process of change in view. One's initial views are modified in order to make them more coherent. This can involve adding new beliefs but it can also involve removing old beliefs.

This is why it can be (and usually is) misleading to contrast induction with deduction. Induction and deduction are of very different categories. Deduction is concerned with implication and consistency. It is not directly concerned with inference. For a conclusion to be deducible from certain premisses is not for that conclusion to be inferable from those premisses. If you believe those premisses and then deduce that conclusion, you learn that your beliefs imply that conclusion. That by itself is not enough for you to be able to be justified in inferring that conclusion. It may instead provide a reason for you to question your belief in the premisses.

Coherence is not the only factor in inductive inference. Otherwise, Sam would make his view as coherent as possible, without any other constraint, perhaps by adopting as his only belief the principle "All is one". But that would be to throw away all of his evidence and all of his prior beliefs without sufficient reason.

#### CONSERVATISM

Sam balances coherence against conservatism. He tries to make a minimal change in his overall view that will improve its coherence. To give up everything else in favour of the single belief that "All is one" would be to put too much emphasis on coherence and not enough on conservatism. Never to abandon prior beliefs and evidence would be to put too much emphasis on conservatism.

The conflicting demands of conservatism versus coherence can lead to the acceptance of two largish views that are each internally coherent although they conflict with each other on certain points, as in the example of nineteenth-century thermodynamics and evolutionary theory.

Historically, some philosophers seem to have supposed that we should ignore conservatism. We should build up a secure view via



self-evident steps from a self-evident foundation. In this view, anything that cannot be given such a justification is to be rejected. Other philosophers have assumed, more reasonably, that what requires justification is changing what you believe; you are justified in believing as you do in the absence of any specific challenge to that belief.

Sam's goals, desires, needs and interests also play a role in his inductive reasoning. As we have seen, his tendency to choose the simpler of two competing hypotheses is a tendency to choose the hypothesis that gives simpler answers to questions in which he is interested. Furthermore, Sam's goals limit the conclusions he is interested in drawing. When Sam is trying to answer a particular question, he will try to restrict his inferences to conclusions that might bear on that question. He will try not to waste his time on trivial conclusions that do not advance the matter at hand.

Sam may be affected by a certain amount of wishful thinking. He may end up accepting one hypothesis (all peaches have pits) rather than another (Albert is removing peaches with pits before I can test them) because he would prefer the first hypothesis to be true (Sam doesn't want to think that he is being manipulated by Albert).

#### UNRESOLVED ISSUES

There are many unsolved problems about induction. One is to characterize coherence in a relatively precise way. Another is to provide a rule for balancing coherence against conservatism. There are questions about wishful thinking. When does wishful thinking occur? Is wishful thinking ever justified?

Also, there is the following unsolved, difficult and generally unrecognized problem about induction: to provide a non-circular account of what constitutes competition among hypotheses.

Although hypotheses compete if they cannot both be true, there are often competing explanations that could all be true. It is consistent with Sam's evidence to suppose both (1) that Albert is screening peaches and is not

going to allow any peaches without pits to be seen by Sam and (2) that all peaches have pits (so that, in fact, Albert will never have to remove any peaches before Sam sees them). However, if Sam comes to accept hypothesis (2), he cannot continue to suppose that his evidence – that all the peaches he has examined have pits – supports (1).

On the other hand, we cannot say that all explanations of certain data are in conflict. Suppose Sam has certain evidence concerning how Max died. The hypothesis that Max was strangled might conflict with the hypothesis that Max was poisoned, but neither of these hypotheses automatically conflicts with the hypothesis that Max died because oxygen wasn't getting to his brain (Harman, 1989).

Finally, it should be observed that some writers believe that probability theory offers the key to problems like these (van Fraassen, 1980, 1989; Jeffrey, 1983). However, probability theory yields results only given a distribution of prior probabilities and the issues discussed here will reemerge as issues about how these prior probabilities are to be determined.

*See also* BAYESIAN EPISTEMOLOGY in Part I; INFERENCE TO THE BEST EXPLANATION; PROBLEMS OF INDUCTION.

#### BIBLIOGRAPHY

- Ginsberg, M. L.: *Readings in Nonmonotonic Reasoning* (Los Altos, CA: Morgan Kaufmann, 1987).
- Glymour, C.: *Theory and Evidence* (Princeton, NJ: Princeton University Press, 1980).
- Goodman, N.: *Fact, Fiction, and Forecast*, 2nd edn (Indianapolis, IN: Bobbs-Merrill, 1965).
- Harman, G.: *Change in View: Principles of Reasoning* (Cambridge, MA: MIT/Bradford Books, 1986).
- Harman, G.: "Competition for Evidential Support," in *Proceedings of the Eleventh Annual Meeting of the Cognitive Science Society* (Mahwah, NJ: Lawrence Erlbaum, 1989), 220–6.
- Hempel, C. G.: *Aspects of Scientific Explanation* (New York: Free Press, 1965).
- Hempel, C. G.: "Provisos: A Problem Concerning the Inferential Function of Scientific Theories," in *The Limits of Deductivism*, ed. A. Grunbaum and

- W. C. Salmon (Berkeley, CA: University of California Press, 1988), 19–36.
- Holland, J., Holyoak, K., Nisbett, R. and Thagard, P.: *Induction: Processes of Inference, Learning, and Discovery* (Cambridge, MA: MIT/Bradford Books, 1986).
- Jeffrey, R. C.: *The Logic of Decision*, 2nd edn (Chicago, IL: University of Chicago Press, 1983).
- Quine, W. V. and Ullian, J. S.: *The Web of Belief* (New York: Random House, 1978).
- Van Fraassen, B.: *The Scientific Image* (Oxford: Oxford University Press, 1980).
- Van Fraassen, B.: *Laws and Symmetry* (Oxford: Oxford University Press, 1989).

GILBERT HARMAN

**infallibility** In epistemology this term is most appropriately used for the epistemic position of a subject vis-à-vis a given subject-matter, more specifically for the impossibility of the subject's being mistaken about that subject matter. Thus it is often held, and as often denied, that a normal human being cannot be mistaken as to whether s/he is currently in a certain kind of conscious state – a thought, feeling or sensation. Those who reject infallibility here do not, of course, deny that people have a high degree of cognitive access to such matters; they only deny that it is *impossible* for mistake to occur. It is claimed, for example, that there is always the possibility of applying the wrong concept to an item. Infallibility will be differently conceived, as will incorrigibility and indubitability (see INCORRIGIBILITY; INDUBITABILITY), depending on the kind of impossibility involved (Alston, 1989, ch. 10), whether, for example, this is logical or causal impossibility.

See also FALLIBILISM.

## BIBLIOGRAPHY

- Alston, W. P.: *Epistemic Justification* (Ithaca, NY: Cornell University Press, 1989).
- Chisholm, R. M. and Swartz, R. J., eds: *Empirical Knowledge* (Englewood Cliffs, NJ: Prentice-Hall, 1973), pt 4.

WILLIAM P. ALSTON

**inference** It is not unusual to find it said that, an inference is a (perhaps very complex) act of thought by virtue of which act (1) I pass from a set of one or more propositions or statements to a proposition or statement and (2) it appears that the latter is true if the former is or are. This psychological characterization has occurred widely in the literature under more or less inessential variations.

It is natural to desire a better characterization of inference. But attempts to do so by constructing a fuller psychological explanation fail to comprehend the grounds on which inferences will be objectively valid – a point elaborately made by Gottlob Frege (see FREGE). And attempts to better understand the nature of inference through the device of the representation of inferences by formal-logical calculations or derivations (1) leaves us puzzled about the relation of formal-logical derivations to the informal inferences they are supposed to represent or reconstruct, and (2) leaves us worried about the sense of such formal derivations. Are these derivations inferences? And aren't informal inferences needed in order to apply the rules governing the constructions of formal derivations (inferring that this operation is an application of that formal rule)? These are concerns cultivated by, for example, Wittgenstein (see WITTGENSTEIN).

Coming up with a good and adequate characterization of inference – and even working out what would count as a good and adequate characterization here – is a hard and by no means nearly solved philosophical problem.

See also INTUITION AND DEDUCTION; PROBLEM OF RULE-FOLLOWING; PROOF.

## BIBLIOGRAPHY

- Baldwin, J. M., ed.: *Dictionary of Philosophy and Psychology* (Gloucester: Peter Smith, 1901, 1960).
- Frege, G.: *The Basic Laws of Arithmetic*, trans. and ed. M. Furth (Berkeley, CA: University of California Press, 1964).
- Harman, G.: *Thought* (Princeton, NJ: Princeton University Press, 1973).

Wittgenstein, L.: *Remarks on the Foundations of Mathematics*, trans. G. E. M. Anscombe (Oxford: Blackwell, 1956).

ROBERT S. TRAGESSER

**inference to the best explanation** This is claimed by many to be a legitimate form of non-deductive reasoning, which provides an important alternative to both deduction and enumerative induction. Indeed, some would claim that it is only through reasoning to the best explanation that one can justify beliefs about the external world, the past, theoretical entities in science, and even the future. Consider belief about the external world and assume that we know what we do about the external world through our knowledge of our subjective and fleeting sensations. It seems obvious that we cannot *deduce* any truths about the existence of physical objects from truths describing the character of our sensations. But neither can we *observe* a correlation between sensations and something other than sensations since by hypothesis all we ever have to rely on ultimately is knowledge of our sensations. Nevertheless, we may be able to posit physical objects as the best explanation for the character and order of our sensations. In the same way, various hypotheses about the past might best explain present memory; theoretical postulates in physics might best explain phenomena in the macro-world; and it is even possible that our access to the future is through universal laws that are formulated to explain past observations. But what exactly is the form of an Inference to the best explanation?

When one presents such an inference in ordinary discourse it often seems to have the following form:

1. O is the case.
2. If E had been the case O is what we would expect.

Therefore there is a high probability that

3. E was the case.

This is the argument form that Peirce called hypothesis or abduction (see PEIRCE;

ABDUCTION). To consider a very simple example, we might upon coming across some footprints on a beach, reason to the conclusion that a person walked along the beach recently by noting that if a person had walked along the beach one would expect to find just such footprints.

But is abduction a legitimate form of reasoning? Obviously, if the conditional in 2 above is read as a material conditional such arguments would be hopelessly bad. Since the proposition that E materially implies O is entailed by O, there would always be an infinite number of competing inferences to the best explanation and none of them would seem to lend even *prima facie* support to its conclusion. The conditionals we employ in ordinary discourse, however, are seldom, if ever, material conditionals. Indeed, the vast majority of "if . . . , then . . ." statements do not seem to be truth-functionally complex. Rather, they seem to assert a *connection* of some sort between the states of affairs referred to in the antecedent (after the "if") and in the consequent (after the "then"). Perhaps the argument form has more plausibility if the conditional is read in this more natural way. But consider an alternative footprints explanation:

1. There are footprints on the beach.
2. If cows wearing boots had walked along the beach recently one would expect to find such footprints.

Therefore, there is a high probability that

3. Cows wearing boots walked along the beach recently.

This inference has precisely the same form as the earlier inference to the conclusion that people walked along the beach recently and its premisses are just as true, but we would no doubt regard both the conclusion and the inference as simply silly. If we are to distinguish between legitimate and illegitimate reasoning to the best explanation it would seem that we need a more sophisticated model of the argument form. It would seem that in reasoning to an explanation we need *criteria* for

choosing between alternative explanations. If reasoning to the best explanation is to constitute a genuine alternative to inductive reasoning, it is important that these criteria not be implicit premisses which will convert our argument into an inductive argument. Thus, for example, if the reason we conclude that people rather than cows walked along the beach is *only* that we are implicitly relying on the premiss that footprints of this sort are usually produced by people, then it is certainly tempting to suppose that our inference to the best explanation was really a disguised inductive inference of the form:

1. Most footprints are produced by people.
2. Here are footprints.

Therefore in all probability

3. These footprints were produced by people.

If we follow the suggestion made above, we might construe the form of reasoning to the best explanation as follows:

1. O (a description of some phenomenon).
2. Of the set of available and competing explanations E1, E2 . . . , En capable of explaining O, E1 is the best according to the correct criteria for choosing among potential explanations.

Therefore in all probability,

3. E1.

Notice that there is a crucial ambiguity in the concept of the best explanation. It might be true of an explanation E1 that it has the best chance of being correct without it being probable that E1 is correct. If I have two tickets in the lottery and one hundred other people each have one ticket, I am the person who has the best chance of winning, but it would be completely irrational to conclude on that basis that I am likely to win. It is much more likely that one of the other people will win than that I will win. To conclude that a given explanation is actually likely to

be correct one must hold that it is more likely that it is true than that the disjunction of all other possible explanations is correct. And since on many models of explanation the number of potential explanations satisfying the formal requirements of adequate explanation is unlimited this will be no small feat.

The model of explanation sketched above must be filled out, of course. Specifically, we need to know what the relevant criteria are for choosing among alternative explanations. Perhaps the single most common virtue of explanation cited by philosophers is *simplicity*. Sometimes simplicity seems to be understood in terms of the number of things or events the explanation commits one to. Sometimes the crucial question concerns the number of *kinds* of things the theory commits one to.

Explanations are also sometimes taken to be more plausible the more explanatory "power" they have. This power is usually defined in terms of the number of things or more likely, the number of kinds of things, the theory can explain. Thus Newtonian mechanics was so attractive, the argument goes, partly because of the range of phenomena the theory could explain.

The familiarity of an explanation in terms of its resemblance to already accepted kinds of explanations is also sometimes cited as a reason for preferring that explanation to less familiar kinds of explanation. So if one provides a kind of evolutionary explanation for the disappearance of one organ in a creature, one should look more favourably on a similar sort of explanation for the disappearance of another organ.

The above are just three examples of criteria one might use in choosing among alternative explanations. There are many other candidates. But in evaluating the claim that inference to the best explanation constitutes a legitimate and *independent* argument form, one must explore question of whether it is a *contingent* fact at least most phenomena have explanations and that explanations that satisfy a given criterion, simplicity, for example, are more likely to be correct. While it might be nice (for scientists and writers of textbooks) if the universe were structured in such a way that simple, powerful, familiar

explanations were usually the correct explanation, it is difficult to avoid the conclusion that if this is true it would be an empirical fact about our universe discovered only a posteriori. If the reasoning to the explanation relies on such criteria, it seems that one cannot without circularity use reasoning to the best explanation to discover that the reliance on such criteria is safe. But if one has some independent way of discovering that simple, powerful, familiar explanations are more often correct, then why should we think that reasoning to the best explanation is an independent source of information about the world? Indeed, why should we not conclude that it would be more perspicuous to represent the reasoning this way?

1. Most phenomena have the simplest, most powerful, familiar explanations available.
2. Here is an observed phenomenon, and E1 is the simplest, most powerful, familiar explanation available.

Therefore, in all probability,

3. This is to be explained by E1.

But the above is simply an instance of familiar *inductive* reasoning.

See also EXPLANATION; the GIVEN; INDUCTION; PROBLEM OF THE EXTERNAL WORLD.

#### BIBLIOGRAPHY

- Harman, G.: "The Inference to the Best Explanation," *Philosophical Review* 74 (1965), 88–95.
- Lipton, P.: *Inference to the Best Explanation* (London: Routledge, 1991).
- Peirce, C. S.: *Collected Papers*, ed. C. Hartshorne and P. Weiss (Cambridge, MA: Harvard University Press, 1931–5).
- Russell, B.: *The Problems of Philosophy* (Oxford: Oxford University Press, 1959).
- Thagard, P.: "The Best Explanation: Criterion for Theory Choice," *Journal of Philosophy* 75 (1978), 76–92.

RICHARD FUMERTON

**infinite regress argument** According to the infinite regress argument for foundationalism (see FOUNDATIONALISM), if every justified belief could be justified only by inferring it from some further justified belief, there would have to be an infinite regress of justifications; because there can be no such regress, there must be justified beliefs that are not justified by appeal to some further justified belief. Instead, they are noninferentially or immediately justified; they are basic or foundational, the ground on which all our other justified beliefs are to rest.

Variants of this ancient argument have persuaded and continue to persuade many philosophers that the structure of epistemic justification must be foundational. Aristotle (see ARISTOTLE) recognized that if we are to have knowledge of the conclusion of an argument on the basis of its premisses, we must know the premisses. But if knowledge of a premiss always required knowledge of some further proposition, he argued, then in order to know the premise we would have to know each proposition in an infinite regress of propositions. Since this is impossible, there must be some propositions that are known, but not by demonstration from further propositions; there must be basic, non-demonstrable knowledge, which grounds the rest of our knowledge.

Foundationalist enthusiasm for regress arguments often overlooks the fact that they have also been advanced on behalf of scepticism, relativism, fideism, contextualism (Annis, 1978) and coherentism (see SCEPTICISM; RELATIVISM; COHERENTISM). Sceptics agree with foundationalists both that there can be no infinite regress of justifications and that nevertheless there must be one if every justified belief can be justified only inferentially, by appeal to some further justified belief. But sceptics think all true justification must be inferential in this way – the foundationalist's talk of immediate justification merely obscures the lack of any rational justification properly so-called. Sceptics conclude that none of our beliefs is justified. Relativists follow essentially the same pattern of sceptical argument, concluding that our beliefs can only be justified relative to the arbitrary



starting assumptions or presuppositions either of an individual or of a form of life.

Fideists also agree with foundationalists that there can be no infinite regress and that nevertheless there must be one if every justified belief can be justified only inferentially. And, again like sceptics and relativists, fideists reject foundationalist talk of rational but immediate justification. Instead, there are beliefs (the fideist's core religious beliefs) that are certified-hence justified, but non-rationally by faith, where faith is usually construed as some divinely inspired act, state or faculty that yields warranted trust in the otherwise unjustified beliefs. What stops the fatal regress of justifications is not belief justified by some immediate foundationalist rationalist intuition, but belief certified by a non-inferential affair beyond the pale of rationality.

Sceptics and relativists see little to choose between such fideism and foundationalism. They are not alone in doing so. Contextualists and coherentists are likely to agree that whether one appeals to faith or to immediacy, the effect is the same: arbitrariness in one's starting point, which would lie beyond responsible canons of justification and criticism (Annis, 1978; BonJour, 1978).

Regress arguments are not limited to epistemology. In ethics there is Aristotle's regress argument (in *Nichomachean Ethics*) for the existence of a single final end of rational action. In metaphysics there is Aquinas's (see AQUINAS) regress argument for an unmoved mover: if everything in motion were moved only by a mover that itself is in motion, there would have to be an infinite sequence of movers each moved by a further mover; since there can be no such sequence, there is an unmoved mover. A related argument has recently been given to show that not every state of affairs can have an explanation or cause of the sort posited by principles of sufficient reason; such principles are false, for a priori reasons having to do with their own concepts of explanation (Post, 1980; Post, 1987, pp. 84–98).

How can the same argument serve so many masters, from epistemology to ethics to metaphysics, from foundationalism to coherentism to scepticism? One reason is that the argument

has the form of a reduction to absurdity of conjoined assumptions. Like all such arguments, it cannot tell us, by itself, which assumption we should reject in order to escape the absurdity. Foundationalists reject one, coherentists another, sceptics a third, and so on. Furthermore, the same argument form can be instantiated by different subject matters, of which epistemology is but one.

What exactly is the form of the argument? Black (1988) suggests the following. The first assumption or premiss has the form

- (1)  $(\forall x)(Ax \rightarrow (\exists y)(Ay \ \& \ xRy))$ .

That is, for every  $x$  that has property  $A$ , there is a  $y$  such that  $y$  has  $A$  and  $x$  bears relation  $R$  to  $y$ . Compare: for every belief  $x$  that is justified, there is a belief  $y$  such that  $y$  is justified and  $x$  is justified by  $y$  (or  $x$  is based on  $y$ , or  $x$  is inferable from  $y$ , or  $y$  is a reason for  $x$ ). Compare also: for everything  $x$  that is in motion, there is a  $y$  in motion that moves  $x$ . The next assumption is

- (2)  $(\exists x)Ax$ .

That is, there are  $A$ 's – there are justified beliefs, there are things in motion. Additionally, one must assume

- (3)  $R$  is irreflexive, and  
(4)  $R$  is transitive.

That is, (3) nothing bears  $R$  to itself; and (4) if  $x$  bears  $R$  to  $y$  and  $y$  bears  $R$  to  $z$ ,  $x$  bears to  $z$ . For instance, if  $x$  justifies  $y$  and  $y$  justifies  $z$ ,  $x$  justifies  $z$ ; if  $x$  moves  $y$  and  $y$  moves  $z$ ,  $x$  moves  $z$ . Finally, the argument assumes

- (5) There is no infinite sequence each of whose elements both has  $A$  and bears  $R$  to its predecessor.

These five assumptions entail a contradiction. In particular, it follows from (1)–(4) that, contrary to (5),

- (6) There is an infinite sequence each of whose elements both has  $A$  and bears  $R$  to its predecessor.

It can be shown rigorously not only that (1)–(4) entail (6), but that each of (1)–(4) is necessary for the entailment (Black, 1988). For example, (6) is not entailed by (1)–(3); R must also be transitive. Thus the regress argument for foundationalism works only if all inferential justification is transitive (Post, 1980).

Since (1)–(5) entail a contradiction, one or more of (1)–(5) must be rejected. Foundation-alists reject (1), or rather the relevant instantiation of (1); there are beliefs that are justified but not by appeal to some further justified belief. (A few foundationalists may also reject (3), allowing some beliefs to be self-justifying.) Fideists likewise reject the relevant instantiation of (1), but disagree with foundationalists about the nature of the justification of the otherwise unjustified beliefs (faith versus rational intuition). Sceptics and relativists, on the other hand, hold to (1) but reject (2); there are no justified beliefs. Coherentists hold to (1)–(3) but reject (4); inferential justification is often a holistic affair that is non-transitive. Contextualists may also reject (4), but mainly reject (1) in favour of *contextually* justified beliefs (Annis, 1978) – those which are unchallenged by the relevant objectors in a given context of justification.

Few philosophers if any seem to have rejected the relevant instantiation of (5), thus opting for what we might call justificational infinitism (as Peirce may do in *Collected Papers* 5.259–5.263). Nevertheless, foundationalists and others have often argued at length against the infinitist option. The usual attempts to do so prove to beg the question against infinitists, typically in favour of foundationalism. For example, it is often said that a regress of justifications would at best provide only conditional justification for its elements, and that we must appeal to some affair outside the regress (hence to something non-inferentially justified, so far as the resources of the regress are concerned). This is to assume just what the infinitist denies. But it now appears that a non-question-begging argument can be given, in the form of a reduction to absurdity of infinitism (Post, 1980; 1987, p. 91: for objections see Sosa, 1980; Moser, 1985). Other instantiations of (5), for example

in metaphysics, have often been rejected, as when philosophers argue that there can be an infinite sequence of movers or causes each moved or caused by its predecessor.

Regress arguments evidently are not the knock-down affairs their advocates have so often supposed them to be. Only if one's favoured way out of the contradiction is the only way, or at least the best way, need such arguments persuade. But showing this has proved surprisingly difficult, requiring forms of argument and evidence that go well beyond the resources of the regress argument itself.

For example, consider a regress argument for foundationalism. Suppose we grant the foundationalist that there are justified beliefs and that justification is irreflexive; this is to grant the relevant instantiations of (2) and (3). What about (4)? Is justification transitive? Some varieties clearly are, including deductive inferential justification, according to which *x* justifies *y* if *x* is justified and *y* is deductively inferable from *x*. Suppose further that *y* justifies *z* in the same sense. It follows that *z* is justified and deducible from *x*, hence that *x* justifies *z*; deductive inferential justification is transitive. Indeed the model or ideal of deductive justification, from Aristotle's theory of demonstration through Euclid nearly to the present, helps explain why so many have supposed that inferential justification must be transitive.

But not all justification is deductive. For example, the justified belief *b*, that Sam is a bartender, inductively justifies belief *c*, that Sam can make a whisky-sour. Now consider the justified belief *a*, that Sam is a bartender who has forgotten how to make a whisky-sour. Belief *a* justifies *b* which inductively justifies *c*, yet obviously *a* does not justify *c* and indeed defeats it; transitivity apparently fails (Klein, 1976, pp. 806–7; Post, 1980, p. 39; Black, 1988, p. 431). Related problems affect varieties of justification according to which *x* justifies *y* only if *x* confers a sufficiently high degree of probability on *y* (Lehrer, 1970, pp. 122–3; Klein, 1976, p. 806; Black, 1988, pp. 431–2).

Another variety of inferential justification is inference to the best explanation (see INFERENCE TO THE BEST EXPLANATION), roughly what

Pierce called abduction (see ABDUCTION). Here  $x$  justifies  $y$  if  $y$  is the best explanation of (the phenomena described by)  $x$ ; if evolutionary theory best explains the fossil record, the record justifies the theory. But explanation relations may not all be transitive (Lehrer, 1970, pp. 112–13). Furthermore, as regards inference to the best explanation, suppose  $y$  is the best explanation of  $x$  (so that  $x$  justifies  $y$ ) and  $z$  is the best explanation of  $y$  (so that  $y$  justifies  $z$ ). If transitivity held,  $z$  would be the best explanation of  $x$ . Yet this contradicts the supposition that  $y$  is the best explanation of  $x$ : presumably there can be only one best explanation of  $x$  (Post, 1980, p. 40).

Foundationalists are not the only ones affected by these troubles with transitivity. So are those fideists, sceptics and relativists who advance regress arguments for their distinctive views. Like foundationalists, they must assume that justification is transitive; otherwise we are not forced to reject (1) or (2), as they contend we are, in order to escape vicious regress. It therefore seems that coherentists, who reject transitivity, are in the best position of all to advance a regress argument for their view – a situation of some irony, in light of long tradition to the contrary, from Aristotle on. But the regress argument is slippery footing even for coherentists. If all beliefs are to be justified by inferring them from other beliefs, as (1) requires, how do we break out of the circle of beliefs to make contact with the world beyond? There are good coherentist answers to this question, some having the possibly welcome effect of denying (1), but they all require support from kinds of argument and evidence that exceed anything to be found in the regress argument itself.

## BIBLIOGRAPHY

- Alston, W. P.: "Two Types of Foundationalism," *Journal of Philosophy* 73 (1976), 165–85.  
 Annis, D. B.: "A Contextualist Theory of Epistemic Justification," *American Philosophical Quarterly* 15 (1978), 213–19.  
 Black, O.: "Infinite Regresses of Justification," *International Philosophical Quarterly* 28 (1988), 421–37.

- BonJour, L.: "Can Empirical Knowledge Have a Foundation?," *American Philosophical Quarterly* 15 (1978), 1–13.  
 Klein, P. D.: "Knowledge, Causality, and Defeasibility," *Journal of Philosophy* 73 (1976), 792–812.  
 Lehrer, K.: "Justification, Explanation, and Induction," in *Induction, Acceptance, and Rational Belief*, ed. M. Swain (Dordrecht: Reidel, 1970), 100–33.  
 Moser, P. K.: "Whither Infinite Regresses of Justification?," *Southern Journal of Philosophy* 23 (1985), 65–74.  
 Post, J. F.: "Infinite Regresses of Justification and of Explanation," *Philosophical Studies* 38 (1980), 31–52.  
 Post, J. F.: *The Faces of Existence: An Essay in Nonreductive Metaphysics* (Ithaca, NY: Cornell University Press, 1987).  
 Sosa, E.: "The Raft and the Pyramid: Coherence versus Foundations in the Theory of Knowledge," *Midwest Studies in Philosophy* 5 (1980), 3–25.

JOHN F. POST

**informal fallacies** There are two basic kinds of informal fallacy: sophistical tactics and erroneous inferences. A sophistical tactic is a technique of argumentation used to get an unfair advantage over one's opponent. This sort of fallacy need not involve an actual intention to cheat in every case, but it does have to be the kind of tactic that would characteristically be used for this purpose. Erroneous inferences, however, can be used in contexts other than those where two people are arguing with each other; one person is enough. This type of fallacy is an inference that falls short of some standard of correct inference (deductive, inductive, or whatever). All that is required, as far as context is concerned, is an appropriate standard of correct inference.

With both kinds of fallacy, it is important to distinguish between a fallacy as an error in a particular case and a fallacy as a general type of inference or sophistical tactic that can often go wrong, or be used wrongly in argumentation, and which we need to watch out for and guard against. It is also important to distinguish between fallacies and other less serious kinds of errors in argumentation, for example, blunders and weak arguments. For

these reasons, there is always a burden of proof on a critic who makes a charge of fallacy in a particular case; he must bring forward evidence to substantiate his claim (see BURDEN OF PROOF). All fallacies (but especially the sophistical tactic type) presuppose a "context of dialogue" from which the relevant evidence should be drawn.

Such "contexts of dialogue" differ according to the purpose of the argument at hand. In a critical discussion, the purpose is to resolve a conflict of opinion. In a negotiation, the purpose is to come to an agreement on how to divide goods or interests that are in short supply. In an inquiry, the purpose is to prove some proposition, based on premisses that can be agreed on as established knowledge. In an eristic dialogue, the purpose is to hit out verbally at the other party, defeating and humiliating the other party, even by violating rules of polite conversation. Many instances of fallacies involve a dialectical shift from one context of dialogue to another. An argument that is correct or appropriate in one context of dialogue may be incorrect or fallacious in another context.

To call a fallacy "informal" is to say that in deciding whether it has been committed we need to take into account both the actual discourse (as with any fallacy) and the context of dialogue for the case. An argument may have a fallacious form, but not every instance of an invalid form of inference is an invalid inference. The reason is that a given argument can have many forms, some valid, and some invalid. If an argument is an instance of an invalid form of inference, it does not follow that the argument must be invalid. Whether the argument is fallacious or not depends on whether and how that form of inference has been used in the context of dialogue. Erroneous inference fallacies are generally less sensitive to context, and are therefore sometimes classified as "formal" fallacies. But it is best to see them as informal, because the context of dialogue can still make some difference.

Only the most prominent fallacies are mentioned below. For more extensive treatments, see Hamblin (1970) or Walton (1987).

## ERRONEOUS INFERENCES

### *Improper Contraposition*

The form of inference "If A then B; therefore if not A then not B" is deductively invalid. An argument having this form may well be fallacious. But it may not. For example, "If the nations disarm, there will be peace; therefore if the nations do not disarm, there will not be peace" is an instance of *improper contraposition*. But that does not necessarily mean that it is a fallacious argument in every context of dialogue. It depends on how it was used, and on what went before in the dialogue. If the premiss were to have been supported, for example, by an earlier argument that war would break out if disarmament did not take place soon enough, it could be that no fallacy was committed.

*Composition* Inference from a property of the part to a property of the whole is warranted only in some cases. An example of a case where it is an erroneous inference: "All the players on this hockey team are good, therefore this is a good hockey team." If the players lack team skills, even though each player is individually good, it could be a poor team.

*Division* Inference from a property of the whole to a property of the part is warranted only in some cases. An example of a case where it is an erroneous inference: "This machine is heavy, therefore all the parts of this machine are heavy." In this case, the property of heaviness does not necessarily transfer from the whole to the parts.

### *Denying the antecedent*

An example: "If capital punishment deters offenders, it is justified, but it does not deter offenders; therefore it is not justified." An inference of this type may seem correct, perhaps because it resembles the valid form of inference on the left. But actually its specific form is that of the invalid inference form on the right. It would be an error to take it as a valid inference.

If A then B	If A then B
Not B	Not A
<hr/>	<hr/>
Not A ( <i>valid</i> )	Not B ( <i>invalid</i> )

The valid type of inference on the left is called *modus tollens*.

### *Affirming the consequent*

An example: "If abortion is acceptable, then it is a woman's right; but abortion is a woman's right; therefore it is acceptable." It would be an error to take this reasoning as a valid inference. It may seem correct, perhaps because it resembles the valid form of inference on the left. But its specific form is that of the invalid inference form on the right.

If A then B A	If A then B B
<hr/>	<hr/>
B (valid)	A (invalid)

The valid type of inference on the left is called *modus ponens*.

### *Post hoc arguments*

A *post hoc* inference starts from a premiss postulating a positive correlation between two events or states, and concludes with postulating a causal relationship between them. This kind of inference is, in general, reasonable as defeasible, presumptive reasoning, because positive correlation is one good, though not infallible sign of the existence of a causal relationship. However, this kind of inference can be erroneous or even fallacious if other factors are overlooked or suppressed. You need to ask several questions. How large is the correlation? Could the causal relationship go the other way? Could it be accounted for by a third factor, a common cause? Could there be an intervening variable, a chain of causality? The causal hypothesis gains in presumptive weight as these factors can be ruled out.

### *Insufficient statistics*

The error of insufficient statistics may be committed where a sample selected is so small that a statistical generalization to a larger population may be worthless or highly unreliable.

### *Biased statistics*

The error of biased statistics arises where the distribution of a property in a sample may not match the distribution in the population as a whole. In such a case, the sample is said to be biased. For example, a civic election poll taken in one neighbourhood may not match voter preferences in the whole city, e.g. it may be a wealthy suburb, while the majority live in poorer urban areas and have different voting preferences.

## SOPHISTICAL TACTICS

### *Argumentum ad verecundiam*

To support your argument in a critical discussion by citing the opinion of an expert is, in general, a reasonable and legitimate move. But it is a kind of tactic that can go wrong in many ways. The expert opinion could be misquoted, or interpreted incorrectly, for example. A secondary context of dialogue, a type of information-seeking expert consultation dialogue, is always involved. The fallacious type of *ad verecundiam* argument occurs in cases where the purported expert opinion is used too aggressively or uncritically by a proponent who, for example, tries to browbeat a respondent by making him appear to lack sufficient respect for the word of an authority.

### *Argumentum ad hominem*

Personal attack in argumentation occurs in two basic forms. In the *personal* or *abusive ad hominem* argument, the respondent's argument is attacked on the ground that the respondent has a bad moral character, and in particular, bad character for veracity is often emphasized. In the circumstantial *ad hominem* argument, the respondent's personal circumstances are claimed to be at variance with his own argument, e.g. "You don't practise what you preach." The circumstantial attack imputes a pragmatic inconsistency, whereas the abusive argument is a direct personal attack. *Ad hominem* argumentation can be reasonable in many cases, e.g. in law, it is legitimate, within



limits, to question the character for veracity of a witness. But it is a kind of argumentation that can go badly wrong if pressed too hard, or used in an inappropriate context of dialogue. For example, in a scientific inquiry, attacking the character of a fellow scientist would normally have no legitimate place as a part of the inquiry.

#### *Arguing in a Circle*

Arguing in a circle is only fallacious where the context of dialogue contains a requirement of evidential priority, i.e. where the premisses must each be better established than the conclusion in order for an argument to be successful. This requirement generally holds in a scientific inquiry, but not always in a critical discussion, where circular argumentation can sometimes be a blunder, due to confusion or inefficiency, rather than a fallacy. Thus circular arguments are not necessarily fallacious, and there is a burden of proof on a critic to show that a circular argument is fallacious in a given case, by referring to evidence from the context of dialogue. The fallacy of arguing in a circle is sometimes called *petitio principii* or begging the question (see CIRCULAR REASONING).

#### *Argumentum ad ignorantiam*

Most arguments are based on a tacit agreement about how much evidence or support one is required to give for one's view in order to have won. This is because hard knowledge is generally insufficient to resolve conflicts of opinion on controversial issues, especially those relating to values and practical conduct. Arguments from ignorance are attempts to change the amount of support required in a given case. They come in two forms: (1) this proposition is not known to be true, therefore it is false, or (2) this proposition is not known to be false, therefore it is true. Whether a particular instance of one of these inferences is warranted or fallacious depends on how it is used in a context of dialogue. For example, "It has not been established that Mr X is a spy, therefore he is not a spy" could be a warranted inference to draw at the conclusion of a thorough and exhaustive inquiry, but

it could be fallacious in another context, e.g. in a critical discussion which aims to establish its conclusion "beyond doubt".

#### *The fallacy of many questions*

The question, 'Have you stopped beating your spouse?' could be reasonable in some contexts, e.g. if put to a confessed spouse-beater by a cross-examining barrister in a court of law. However, a complex question of this type which contains presuppositions highly detrimental to the respondent can be an unfairly aggressive tactic aimed at trapping him into making concessions that will be harmful to his side. When this occurs, the respondent should question the question (if possible), requesting that it be broken down into a series of smaller questions asked in a reasonable order, allowing for better possibilities of choice (Walton, 1989). In some cases, such questions involve unfair dichotomies, needing the reply "None of the above".

#### *Argumentum ad baculum*

Appeal to threat of force or sanctions can be quite legitimate in negotiation dialogue, especially where the threat is expressed (as an indirect speech act) in the form of a warning, e.g. "If you don't meet these terms, our union will go out on strike". But in the context of a critical discussion, an attempt to block or close off the discussion by using a threat is often an illicit dialectical shift. This move is a common tactic called the *ad baculum* fallacy.

#### *Argumentum ad misericordiam*

Appeal to the emotion of pity can be quite reasonable as a guide to action in some cases, but it becomes fallacious when used to distract attention from relevant evidence that should be taken into account in arriving at a conclusion.

#### *Argumentum ad populum*

Critical discussion is an opinion-based type of dialogue which has as its goal the resolution of a conflict of opinions, often on controversial issues of public policy. Therefore, appealing to a weight of common presumption or

popular opinion is by no means out of place. However, when such an appeal is made in an uncritical manner, by suggesting that popular opinion can never be wrong, for example, the fallacy of *ad populum* is said to be committed. It should be noted that appeal to popular opinion is a kind of defeasible argumentation that is inherently open to critical questioning.

### *Equivocation*

One way of getting in trouble with vague or ambiguous terms in reasoning arises through a shift in context, where the same word or phrase has been used twice, making it plausible to seem that there is a correct inference, erroneously. A transparent case: "Smith went to the bank to get some cash; the bank is receding from the river; therefore, Smith went to an area that is receding from the river to get some cash." It could be, in such a case that each premiss is individually true, but the conclusion is false, despite the (apparent) validity of the inference. Equivocation tends to be an especially serious problem in longer sequences of reasoning, where such shifts can easily pass unnoticed (Walton, 1987, ch. 10).

### *Slippery slope*

A slippery slope argument is a kind of argument that warns you if you take a first step, you will find yourself involved in an irreversible sequence of consequences, speeding faster and faster towards some disastrous outcome. A good example was the argument used to support the majority opinion in the recent US Supreme Court decision not to ban burning of the American flag as a criminal act. Justice William J. Brennan, Jr argued that any ruling to ban physical desecration of the flag would lead to further cases that would "enter territory having no discernible or defensible boundaries". Wouldn't the court then have to consider prohibiting the burning of state flags, or the Constitution? Justice Brennan worried that in order to evaluate these choices, the court would end up imposing its own political preferences to suppress all kinds of unpopular protests. This kind of outcome is obviously dangerous in a democratic

country where freedom of speech is important. In many cases, the slippery slope is a legitimate technique of practical reasoning which a proponent can use in action-directed dialogue to warn a respondent about the consequences of a path of action the respondent is embarking upon. However, in some cases, the technique can take the form of a scare tactic used fallaciously, without sufficient justification, to try to forestall action without fairly considering the consequences or alternatives.

### BIBLIOGRAPHY

- Fearnside, W. W. and Holther, W. B.: *Fallacy: The Counterfeit of Argument* (Englewood Cliffs, NJ: Prentice-Hall, 1959).
- Hamblin, C. L.: *Fallacies* (London: Methuen, 1970).
- Mackie, J. L.: "Fallacies," in *The Encyclopedia of Philosophy*, ed. P. Edwards (New York and London: Collier-Macmillan, 1967), vol. 3, pp. 169–79.
- Pirie, M.: *The Book of the Fallacy* (London: Routledge, 1985).
- Sidgwick, A.: *Fallacies* (London: Kegan Paul, Trench, 1886).
- Van Eemeren, F. H. and Grootendorst, R.: "Fallacies in Dialectical Perspective," *Argumentation* 1 (1987), 283–302.
- Van Eemeren, F. H. and Grootendorst, R.: *Argumentation, Communication and Fallacies* (Hillsdale, NJ: Lawrence Erlbaum, 1992).
- Walton, D. N.: *Informal Fallacies* (Amsterdam: Benjamins, 1987).
- Walton, D. N.: *Question–Reply Argumentation* (New York: Greenwood Press, 1989).
- Woods J. and Walton, D. N., eds: *Fallacies: Selected Papers, 1972–1982* (Dordrecht: Foris, 1990).

DOUGLAS N. WALTON

**innate ideas** These have been variously defined by philosophers either as ideas consciously present to the mind prior to sense experience (the non-dispositional sense), or as ideas which we have an innate disposition to form (though we need not be actually aware of them at any particular time, e.g. as babies) – the dispositional sense.

Understood in either way they were invoked to account for our recognition of certain truths without recourse to experiential

verification, such as those of mathematics, or to justify certain moral and religious claims which were held to be capable of being known by introspection of our innate ideas. Examples of such supposed truths might include “murder is wrong” or “God exists”.

One difficulty with the doctrine is that it is sometimes formulated as one about concepts or ideas which are held to be innate and at other times as one about a source of propositional knowledge. In so far as concepts are taken to be innate the doctrine relates primarily to claims about meaning: our idea of God, for example, is taken as a source for the meaning of the word God. When innate ideas are understood propositionally their supposed innateness is taken as evidence for their truth. This latter thesis clearly rests on the assumption that innate propositions have an unimpeachable source, usually taken to be God, but then any appeal to innate ideas to justify the existence of God is circular. Despite such difficulties the doctrine of innate ideas had a long and influential history until the eighteenth century and the concept has in recent decades been revitalized through its employment in Noam Chomsky's influential account of the mind's linguistic capacities.

The attraction of the theory has been felt strongly by those philosophers who have been unable to give an alternative account of our capacity to recognise that some propositions are certainly true where that recognition cannot be justified solely on the basis of an appeal to sense experience. Thus Plato (*see* PLATO) argued that, for example, recognition of mathematical truths could only be explained on the assumption of some form of recollection (*see* ANAMNESIS). Since there was no plausible post-natal source the recollection must refer back to a pre-natal acquisition of knowledge. Thus understood, the doctrine of innate ideas supported the view that there were important truths innate in human beings and it was the senses which hindered their proper apprehension.

The ascetic implications of the doctrine were important in Christian philosophy throughout the Middle Ages and the doctrine featured powerfully in scholastic teaching until its displacement by Locke's philosophy (*see*

LOCKE) in the eighteenth century. It had in the meantime acquired modern expression in the philosophy of Descartes (*see* DESCARTES), who argued that we can come to know certain important truths before we have any empirical knowledge at all. Our idea of God, for example, and our coming to recognize that God must necessarily exist, are, Descartes held, logically independent of sense experience. In England the Cambridge Platonists such as Henry More and Ralph Cudworth added considerable support.

Locke's rejection of innate ideas and his alternative empiricist account was powerful enough to displace the doctrine from philosophy almost totally. Leibniz (*see* LEIBNIZ), in his critique of Locke, attempted to defend it with a sophisticated dispositional version of the theory, but it attracted few followers.

The empiricist alternative to innate ideas as an explanation of the certainty of propositions was in the direction of construing all necessary truths as analytic. Kant's refinement (*see* KANT) of the classification of propositions with the fourfold distinctions analytic/synthetic and a priori/a posteriori did nothing to encourage a return to the innate ideas doctrine, which slipped from view (*see* ANALYTICITY). The doctrine may fruitfully be understood as the production of confusion between explaining the genesis of ideas or concepts and the basis for regarding some propositions as necessarily true.

Chomsky's revival of the term in connection with his account of human speech acquisition has once more made the issue topical. He claims that the principles of language and “natural logic” are known unconsciously and are a precondition for language acquisition. But for his purposes innate ideas must be taken in a strongly dispositional sense – so strong that it is far from clear that Chomsky's claims are as in conflict with empiricists' accounts as some (including Chomsky) have supposed. Quine, for example, sees no clash with his own version of empirical behaviourism, in which old talk of ideas is eschewed in favour of dispositions to observable behaviour.

*See also* ANALYTICITY; A PRIORI KNOWLEDGE in Part I; EMPIRICISM.

## BIBLIOGRAPHY

- Chomsky, N.: *Cartesian Linguistics* (New York: Harper and Row, 1966).
- Leibniz, G. W.: *New Essays on Human Understanding* (1704), trans. P. Remnant and J. Bennett (Cambridge: Cambridge University Press, 1981).
- Locke, J.: *An Essay Concerning Human Understanding* (1690), ed. P. H. Nidditch (Oxford: Oxford University Press, 1975).
- Plato: *Meno*.
- Quine, W. V.: "Linguistics and Philosophy," in Stich (ed.), 1975.
- Stich, S. P., ed.: *Innate Ideas* (Berkeley, CA: University of California Press, 1975).

G. A. J. ROGERS

**introspection** Derived from the Latin *intro* (within) + *specere* (to look), introspection is the attention the mind gives to itself or to its own operations and occurrences. I can know there is a fat hairy spider in my bath by looking there and seeing it. But how do I know that I am seeing it rather than smelling it, or that my attitude to it is one of disgust rather than delight? One answer is: by a subsequent introspective act of "looking within" and attending to the psychological state – *my seeing the spider*. Introspection, therefore, is a mental occurrence, which has as its object some other psychological state like perceiving, desiring, willing, feeling, etc. In being a distinct awareness-episode it is different from a more general "self-consciousness" which characterizes all or some of our mental history.

The awareness generated by an introspective act can have varying degrees of complexity. It might be a simple *knowledge of (mental) things* – such as a particular perception-episode; or it might be the more complex *knowledge of truths* about one's own mind. In this latter full-blown judgemental form, introspection is usually the self-ascription of psychological properties and, when linguistically expressed, results in statements like "I am watching the spider" or "I am repulsed".

In psychology this deliberate inward look becomes a *scientific method* when it is "directed towards answering questions of theoretical importance for the advancement of our systematic knowledge of the laws and

conditions of mental processes" (Stout, 1938). In philosophy, introspection (sometimes also called "reflection") remains simply "that notice which the Mind takes of its own Operations" (Locke, 1690, 2.1.4) and has been used to serve the following important functions:

1. *Methodological*: Thought experiments are a powerful tool in philosophical investigation. The Ontological Argument, for example, asks us to try to think of the most Perfect Being as lacking existence and Berkeley's Master Argument challenges us to conceive of an unseen tree. Conceptual results are then drawn from our failure or success. For such experiments to work, we must not only have (or fail to have) the relevant conceptions but also *know* that we have (or fail to have) them – presumably by introspection.
2. *Metaphysical*: A metaphysics of mind needs to take cognisance of introspection. One can argue for "ghostly" mental entities, for "qualia", for "sense-data" by claiming introspective awareness of them. First-person psychological reports can have special consequences for the nature of persons and personal identity: for example, Hume (*see* HUME) was content to reject the notion of a soul-substance because he failed to find such a thing by "looking within". Moreover, some philosophers argue for the existence of additional *perspectival facts* – the fact of "what it is like" to be the person I am or to have an experience of such-and-such-a-kind (Nagel, 1974). Introspection as our access to such facts becomes important when we construct a complete metaphysics of the world.
3. *Epistemological*: Surprisingly, the most important use made of introspection has been in accounting for our knowledge of the *outside* world. According to a foundationalist (*see* FOUNDATIONALISM) theory of justification an empirical belief is either *basic* and "self-justifying" or is justified in relation to basic beliefs. Basic beliefs therefore, constitute the rock-bottom of all justification and knowledge. Now introspective awareness is said to have a unique epistemological

status: in it, we are said to achieve the best possible epistemological position and consequently, introspective beliefs become prime candidates for “basic” beliefs and thereby constitute the foundation of all justification (see GIVEN).

#### THE TRADITIONAL THEORY

The traditional theory of introspection, as I call it, is an explanation of this capacity of our “looking within” constructed from a Descartes–Locke–Kant perspective. It develops as an epistemological corollary to a metaphysical dualism. The world of Matter is known through external/outer sense-perception. So cognitive access to Mind must be based on a *parallel* process of introspection which “though . . . not Sense, as having nothing to do with external Objects; yet (it) is very like it, and might properly enough be call’d *internal Sense*” (Locke, 1690, 2.1.4). However, “having mind as object” is not sufficient to make a way of knowing “inner” in the relevant sense because mental facts can be grasped through sources other than introspection. The point is rather that an “inner perception” provides a *kind* of access to the mental not obtained otherwise – it is a “look within *from within*”. Stripped of metaphor this indicates the following epistemological features:

1. *Only* I can introspect my mind.
2. I can introspect *only my* mind.
3. Introspective awareness is *superior* to any other knowledge of contingent facts that I or others might have.

(1) and (2) are grounded in the Cartesian idea of “privacy” of the mental. Normally, a single object can be perceptually or inferentially grasped by many subjects, just as the same subject can perceive and infer different things. The epistemic peculiarity of introspection is that it is exclusive – it gives knowledge only of the *mental history of the subject introspecting*.

Tenet (3) of the traditional theory is grounded in the Cartesian idea of “privileged access”. The epistemic superiority of introspection lies in its being an infallible source of

knowledge. First-person psychological statements which are its typical results cannot be mistaken. This claim is sometimes supported by an “imaginability test”, i.e. the impossibility of imagining that I believe that I am in pain while at the same time imagining evidence that I am not in pain. An apparent counter-example to this infallibility claim would be the introspective judgement “I am perceiving a dead friend” when I am really hallucinating. This is taken care of by reformulating such introspective reports as “I *seem* to be perceiving a dead friend”. The importance of such privileged access is that introspection becomes a way of knowing immune from the pitfalls of other sources of cognition. The basic asymmetry between first and third-person psychological statements can be traced to their being generated (respectively) by introspective and non-introspective methods.

The traditional theory of introspection, therefore, can be encapsulated in the following four theses: (1) Perceptual Model Thesis, (2) Distinct Act Thesis, (3) Privacy Thesis and (4) Privileged Access Thesis.

Before looking at the criticisms of this theory, an important qualification needs to be made regarding tenets (1) and (2) stated above. Introspection, so far, has been defined as yielding the knowledge of the subject’s own mind or mental history. The umbrella terms “mental history” or “my mind”, however, tend to gloss over an important controversy centring on the actual mental *items* revealed in introspection. The debate here has greater significance than just generating a list: if we find uncontroversial psychological entities not amenable to introspection or dubiously “mental” items that are uncontroversially introspected, then it would be clear that introspectibility is either not a necessary or not a sufficient criterion of the mental. Some of the philosophically interesting putative objects of introspection are:

1. Psychological/mental states: Even if many psychological states are introspected it is doubtful if *all* such states can be known in this manner. There are many types of mental states and it is not clear that all of them are introspectible or introspectible



in the same way. A *dispositional* psychological state is a possible exception.

2. Self or I: Introspection is generally supposed to reveal not only psychological states but also the *subject* or *seat* of these states. Some, à la Hume (*see* HUME), however, confess to a failure to discover a Self over and above its states by “looking within”. The issue here hinges on whether, in becoming aware of my experiences, I am also not aware of them as-*my*-experiences and whether the latter awareness is possible without an *introspective* awareness of the Self.
3. Bodily sensations like aches, itches, etc: Reports like “I am dizzy”, “I have a sinking feeling in my stomach” are sometimes said to be known introspectively. To hold them to be bona fide introspections we would need either to construe bodily sensations as mental or to allow an “introspective awareness” of some physical states.
4. Time and temporal determination: This is part of Kant’s (*see* KANT) idiosyncratic theory of “inner sense”. Our faculty of Sensibility is exercised either as “outer sense” or as “inner sense”. The “intuitional aspect” or Form of outer sense is Space and that of inner sense is Time. This means that while all objects of outer sense are represented as spatial, all inner perceptions are processed as temporal. But more interestingly, even our ascription of temporal succession to events in the world is dependent on and derived from the (introspected?) successiveness of our inner perceptions.

#### OBJECTIONS TO THE TRADITIONAL THEORY

These either question the *plausibility* of the four theses constituting the theory or expose their *incompatibility*.

##### *Against the perceptual model thesis*

The motivation for construing introspection along the lines of perception is a desire for theoretical neatness. Though Mind and Matter are metaphysically different, we become aware of them in fundamentally parallel ways. But the

difficulty is that we cannot find introspectional analogues to many crucial elements in ordinary perceptual processes. Many *de facto* disanalogies have been pointed out (for example, the absence of an introspective sense-organ and an absence of a distinct phenomenological character of our “experience” of the inner) but the difficulty is really logical. Any theory of perception must leave room for *misperceptions*, for perception is an “achievement” or “success” word. Introspective awareness allegedly cannot be mistaken. Consequently, to account for it in terms of the same theoretical structure as perception is misconceived. At best, the Perceptual Model Thesis is at odds with the Privileged Access Thesis.

##### *Against the distinct act thesis*

S’s awareness of O, written as (i) S-a-O, if grasped by another awareness-episode (ii) S-a-(S-a-O) immediately suggests a regress – for would not we need yet another episode to know (ii)? However, this is not a problem. Mental states need only be introspectible and not introspected, (i) as a conscious mental state reveals its object and this illumination (of O) is not borrowed from the subsequent act (ii); similarly, if all we want is to know (i), it is sufficient to move to (ii) which, as a conscious mental episode, can adequately reveal its object (S-a-O) without requiring a move higher. Of course, with nothing better to do we could introspect our introspections ad infinitum and ad nauseam – but this is not necessary if our purpose was to grasp an initial awareness.

Ryle (1949) (*see* RYLE) has a stronger objection to the Distinct Act Thesis. Introspection is logically self-defeating because it destroys its very object. In (i) S knows O because S attends to O. In (ii) S knows (i) because S attends to (S-a-O). But S can attend to only one object at a time. Thus, while attending to (S-a-O) in (ii), S must withdraw attention from O in which case there will be *no* (S-a-O). But without its object (S-a-O), there can be no introspective awareness at all of the form (ii). Ryle’s suggested way out of this in terms of “retrospection” virtually abandons the Perceptual Model for a

“Memory (or at best a Very-Fresh-Memory) Model” of introspection.

Again, we cannot in the same breath say that introspection is a distinct mental operation *and* that it is a *logically* infallible way knowing. If pain and the awareness of are “distinct existences” then the logical possibility of awareness of pain without pain is still present (see Armstrong, 1968) and the doctrine of infallibility falls. There is thus a tension between the Distinct Act Thesis and Privileged Access.

#### *Against the privacy thesis*

A broadly Wittgensteinian approach (see WITTGENSTEIN) questions the idea of an inward look picking out mental phenomena not accessible from a third-person perspective. The argument has many versions. On one version, there would be a tension between our Privacy Thesis and Privileged Access. According to the latter, introspective awareness cannot be mistaken. But if such awareness is of what is “private”, then there can be no way of *checking* our experience. And without the notion of a check or test, no sense can be made of being *right*. Furthermore, even if the possibility of such an inward look is granted, what it would yield could not be *reported* in a learnable language.

#### *Against the privileged access thesis*

We can think of instances of introspection yielding mistaken belief. We have been known to misidentify our mental states and we can think of cases where a physiologist says that the brain state responsible for a particular mental state has not occurred even though my introspective report is that I am in that state. And so it seems better to weaken the claim that introspective reports are infallible. But any substantial weakening of this idea that introspection is a different kind of knowing.

#### ALTERNATIVES TO THE TRADITIONAL THEORY

These reject one or more of its constitutive tenets. By denying dualism, physicalists about

the mind abolish the metaphysical foundations of the standard view; but even *dualists* can account for introspective awareness in different ways. I sketch a few features of some of these options.

1. Non-perceptual models: Self-scrutiny need not be perceptual. My awareness of an object O changes the status of O. It now acquires the property of “being an object of awareness”. On the basis of this or the fact that the object is seen by me, I *infer* that I am aware of O. Such an “Inferential Model” of awareness is suggested by the Bhatta Mimamsa school of Indian epistemology (Matilal, 1986) (see INDIAN EPISTEMOLOGY). This view of introspection does not construe it as a *direct* awareness of mental operations but, interestingly, we will have occasion to refer to theories where the emphasis on directness itself leads to a non-perceptual or at least a non-observational account of introspection.
2. Reflexive models: Epistemic access to our minds need not involve a separate attentive act. Part of the meaning of a *conscious* state is that I *know* that I am in that state when I am in that state. Consciousness here is conceived as a “phosphorescence” attached to some mental occurrences and in no need of a subsequent illumination to reveal itself. Of course, if introspection is defined as a distinct act then reflexive models are really accounts of first-person access that make no appeal to introspection.
3. Public-mind theories and fallibility/infallibility models: The physicalists’ denial of metaphysically private mental facts naturally suggests that “looking within” is not merely *like* perception but is perception. For Ryle, mental states are “iffy” behavioural facts which, in principle, are equally accessible to everyone in *the same way*. One’s *own* self-awareness therefore is, in effect, no different in type from anyone else’s observations about one’s minds.

A more interesting move is for the physicalist to retain the truism that I grasp that I am sad in a very different way from that in which I know *you* to be sad. This *directness* or

non-inferential nature of self-knowledge can be preserved in some physicalist theories of introspection. For instance, Armstrong's identification of mental states with *causes* of bodily behaviour and of the latter with brain states, makes introspection the process of acquiring information about such inner physical causes. But since introspection is itself a *mental state*, it is a process in the brain as well; and since its grasp of the relevant causal information is *direct*, it becomes a process in which the brain scans itself.

Alternatively, a broadly "functionalist" view of mental states suggests the following machine-analogue of the introspective situation: a machine-table with the instruction "Print: 'I am in state A' when in state A" results in the output "I am in state A" when state A occurs. Similarly, if we define mental states and events functionally, we can say that introspection occurs when an occurrence of a mental state M directly results in an awareness of M. Note that this way of emphasizing *directness* yields a non-perceptual and *non-observational* model of introspection. The machine in printing "I am in state A" does so (when it is not making a "Verbal mistake") just *because* it is in state A. There is no computation of information or process of *ascertaining* involved. The latter, at best, consists simply in passing through a sequence of states.

This casts new light on the discussion. The legitimate question: How do I know that I am seeing a spider? was interpreted as a demand for the *faculty* or information-processing mechanism whereby I come to acquire this knowledge. Peculiarities of first-person psychological awareness and reports were carried over as peculiarities of this mechanism. However, the question *need* not demand the search for a *method* of knowing but rather for an *explanation* of the special epistemic features of first-person psychological statements. On this reading, the problem of introspection (as a way of knowing) dissolves but the problem of explaining "introspective" or first-person authority remains.

See also APPERCEPTION; CARTESIANISM; the GIVEN; PRIVATE LANGUAGE ARGUMENT; RYLE;

SELF-CONSCIOUSNESS; SELF-KNOWLEDGE AND SELF-IDENTITY; SUBJECTIVITY.

# BIBLIOGRAPHY

- Armstrong, D. M.: *A Materialist Theory of Mind* (London: Routledge and Kegan Paul, 1968).  
 Descartes, R.: *Meditations on First Philosophy* (1641), in *Philosophical Works of Descartes*, ed. J. Cottingham, R. Stoothoff and D. Murdoch (Cambridge: Cambridge University Press, 1975).  
 Kant, I.: *Critique of Pure Reason* (1781), trans. N. Kemp Smith (London: Macmillan, 1964).  
 Locke, J.: *An Essay Concerning Human Understanding* (1690), ed. P. H. Nidditch (Oxford: Oxford University Press, 1975).  
 Matilal, B. K.: *Perception: An Essay on Classical Indian Theories of Knowledge* (Oxford: Clarendon Press, 1986).  
 Nagel, T.: "What Is It Like to Be a Bat?," *Philosophical Review* 83 (1974), 435–50.  
 Ryle, G.: *The Concept of Mind* (Harmondsworth: Penguin, 1949).  
 Shoemaker, S.: "First-person access," in *Philosophical Perspectives* 4 (1990).  
 Stout, G. F.: *A Manual of Psychology*, 5th edn (London: University Tutorial Press, 1938).  
 Wittgenstein, L.: *Philosophical Investigations*, trans. G. E. M. Anscombe (Oxford: Blackwell, 1953).

VRINDA DALMIYA

**intuition and deduction** Most generally, one has *intuitive* "knowledge" that *p* when,

1. one knows that *p*,
2. one's knowledge that *p* is immediate, and
3. one's knowledge that *p* is not an instance of the operation of any of the five senses (so that knowledge of the nature of one's own experience is not intuitive).

On this account neither mediated nor sensory knowledge is intuitive knowledge. Some philosophers, however, want to allow sensory knowledge to count as intuitive; to do this, omit clause (3) above.

The two principal families of examples of *mediated* (i.e. not immediate) knowledge that have interested philosophers are, knowledge via *representation* and knowledge via *inference*. Knowledge by representation occurs

when the thing known is not what one appeals to as a basis for claiming to know it, as when one appeals to sensory phenomena as a basis for knowledge of the world (and the world is not taken to be a sense-phenomenal construct) or as when one appeals to words as a source of knowledge of the world (as when one claims that a proposition is true of the world solely by virtue of the meanings of the words expressing it).

(There are other idioms that are used to mark out the differences between non-intuitional and intuitional ways of knowing, such as knowing indirectly and knowing directly, or knowing in the absence of the thing known and knowing by virtue of the presence of the thing known. It is sometimes useful to speak of the object of knowledge being *intuitively given*, meaning that we can know things about it without mediation. The justification of a claim to knowledge by appeal to its object being intuitively given is surely as good as could be. What could be a better basis for a claim to knowledge than the object of knowledge itself given just as it is?)

We might say that *deductive inference* is a mode of achieving *conditional* knowledge. One *infers* a proposition  $p$  from one or more propositions  $p_1, \dots, p_n$ , called *premisses* of the inference,  $p$  being called the *conclusion* of the inference. Most generally, to *validly infer*  $p$  from premisses  $p_1, \dots, p_n$  is to think or reason one's way to  $p$  from those premisses in such a way that one can see that, if the premisses are known (and so true), then the conclusion is thereby known (and so true).

One of the fundamental problems of philosophy, overlapping epistemology and the philosophy of logic, is that of giving criteria for when a deductive inference is valid, criteria for when an inference (*see* INFERENCE) does or can continue knowledge or truth. There are in fact two very different proposals for solutions to this problem, one that had slowly come into fashion during the early part of this century, and another that has been much out of fashion, but is gaining in admirers. The former, which develops out of the tradition of Aristotelian syllogistic, holds that all valid deductive inferences can be analysed and paraphrased as follows:

- The sentences occurring in the deduction are aptly paraphrased by sentences with an explicit, *interpreted* logical syntax, which in the main consists of expressions for logical operations, e.g. predication, negation, conjunction, disjunction, quantification, abstraction . . . ; and
- The validity of the inferences made from sentences in that syntax to sentences in that syntax is entirely a function of the meaning of the signs for logical operations expressed in the syntax.

In particular, it is principally the meaning of the signs for logical operations that justify taking considered rules of inference as valid. (For a thorough discussion of which operations are to count as logical, and of what can count as a logical constant, see Koslow, 1991.) Here, for example, is such a justification as given by Frege (*see* FREGE), one of the great developers of this view of the nature of the proper criteria for valid deductive inference, someone who in fact, in the late nineteenth century, gave us an interpreted logical syntax (and so a formal deductive logic) far, far greater and more powerful than had been available through the tradition of Aristotelian syllogistic:

$A \rightarrow B$  is meant to be a proposition that is false when  $A$  is true and  $B$  is false; otherwise it is true (Frege, 1964, p. 51; paraphrased; variables restricted to the True, the False).

The following is a valid rule of inference, From  $A$  and  $A \rightarrow B$ , infer  $B$ , for if  $B$  were false, since  $A$  is true,  $A \rightarrow B$  would be false; but it is supposed to be true (Frege, 1964 p. 57; paraphrased).

Frege believed that the principal virtue of such formal-syntactical reconstructions of inferences – as validly moving on the basis of the meanings of the signs for the logical operations alone – was that it eliminated dependence on intuition and let one see exactly on what our inferences depended, e.g.:

we divide all truths that require justification into two kinds, those for which the proof can be carried out purely by means of logic and those for which it must be supported by facts of experience.

... Now, when I came to consider the question to which of these two kinds the judgments of arithmetic belong, I first had to ascertain how far one could proceed in arithmetic by means of inferences alone, with the sole support of those laws of thought that transcend all particulars. ... To prevent anything intuitive [*Anschauliches*] from penetrating here unnoticed, I had to bend every effort to keep the chain of inferences free from gaps.

(Frege, 1967, p. 5)

In the literature most ready to hand, the alternative view was supported by Descartes (see DESCARTES) and elaborated by John Locke (see LOCKE), who maintained that inferences move best and most soundly when based on *intuition* (their word):

Syllogism serves our Reason [in that it shows] the connexion of the Proofs [i.e. the connexion between premisses and conclusion] in any one instance and no more; but in this, it is of no great use, since the Mind can perceive such connexion where it really is, as easily, nay, perhaps better without [Syllogism].

If we observe the Actings of our own Minds, we shall find, that we reason best and clearest, when we only observe the connexion of the [ideas], without reducing our Thoughts to any Rule of Syllogism.

(Locke, 1975, p. 670)

What is it that one is intuiting? Ideas, or meanings, and relationships among them. Ideas or meaning are taken to be directly given. The difference being marked by Locke is between (a) inferring Socrates is mortal from the premisses All men are mortal and Socrates is a man by appealing to the formal-logical rule, AH A are B, C is an A, therefore C is B, which is supposed to be done without any appeal to the intuitive meanings of "All" and "is", and (b) seeing that Socrates is mortal follows from All men are mortal and Socrates is a man by virtue of understanding (the meanings of) those informal sentences without any appeal to the formal-logical rule. Locke is also making the point that inferences made on the basis of such an understanding of meanings are better, and more fundamental, than inferences made on the basis of an appeal to a formal-logical schema. Indeed,

Locke would certainly maintain that such informal, intuitive inferences made on the basis of understanding the meanings of sentences serve better as a check on the correctness of formal inferences than formal-logical inference serve as a check on intuitive inferences.

Such distrust of formal logical inference or greater trust in intuitive inference has been promoted in recent times by Henri Poincaré and L. E. J. Brouwer (Detlefsen, 1991).

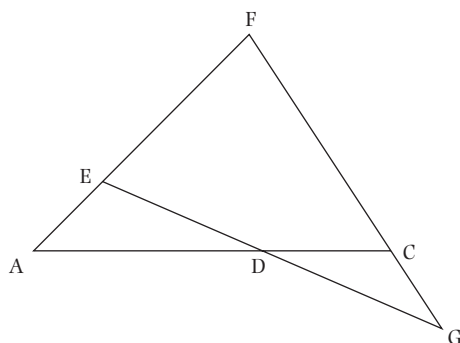
We might say that for Frege, too, logical inferences moved by virtue of intuition of meaning, the meaning of the signs for logical inference, for we have seen how Frege appealed to such meanings in order to justify formal-logical rules of inference. Of course, once the formal-logical rules are so justified, Frege is quite content to appeal to them in the construction of deductions, not returning each time to the intuited meanings of the logical signs. What is new in Frege is the conviction that inferences that proceed wholly on the basis of the logical signs, signs for logical operations, are complete with respect to logical implication – that if B logically follows from A, then we should in principle be able to deduce B from A by rules which mention only logical operations and not, e.g., the concrete meanings of predicate-expressions in the relevant propositions. There is a deep issue here which is destined to become the principal issue in the philosophy and epistemology of logical theory, viz. *To what extent, in what measure, does intuition of the non-logical content of propositions (i.e. contents other than the meanings of the signs for logical operations) rightly sustain inference?*

This is the issue that really concerned Brouwer and Poincaré (Detlefsen, 1991). But consider: Katz (1988) argued that Descartes' cogito (see COGITIO) is a sound inference made on the basis of intuitions of meanings and is *incapable of being articulated or paraphrased as formal-syntactic reasoning* after the now ubiquitously deployed method of Frege depending – as described above – on logical operations alone (see ANALYTICITY). But one not really need to reach for such examples. Virtually all inferences set out in mathematical proofs most obviously proceed on basis of intuitively given meaning content rather than appeal to



formal-logical rules, and it is easy to find examples of such proofs that clearly do not depend on the meanings of signs for logical operators, but rather on the non-logical content of the mathematical propositions. There is a good example in Hilbert (1971, p. 6, paraphrased):

*Theorem 3.* For two points A and C there always exists at least one point D on the line AC that lies between A and C.



**Figure 1**

Proof: By virtue of the axiom which says that there exists three points not on a line, there exists a point E not on the line AC.

By virtue of the axiom which says the for any two points A, E, there is a third point F on the line AE, E between A and F, there exists on AE a point F such that E is a point of the segment AF.

By the same axiom and the axiom that says of any three points on a line there exists no more than one that lies between the other two, there exists on FC a point G, and G does not lie on the segment FC.

Given the axiom which says, if three points A, B, C do not lie on a line and if a is a line which passes through a point of the segment AB, the line a also passes through either a point of the segment AC or a point of the segment BC – it follows that the line EG must then intersect the segment AC at a point D.

The proof actually consists of four lemmas put together via the picture Figure I to yield the construction of the point required by the theorem. Here the picture, which is to say, something expressive of the intuitively given content of the theorem, stands in for a formal-

logical construction that would bring the four lemmas together to yield the theorem. The intuitive understanding of the pictured content of the theorem stands in for the formal-logical construction. Of course, we could in fact formalize Hilbert, and give a formal-logical proof of the formal counterpart of the above theorem from the formal-logical counterparts of the relevant ones among Hilbert's axioms. It would not of course be anything like "the same" proof, for the formal-logical proof would trade on the signs for logical operations only, whereas the proof above trades on an understanding of what one means to be proving theorems about. And the formal-logical proof would consist of perhaps hundreds of steps. (Recall Locke's remark, more apt than he could have realized, that intuitive proofs are shorter.)

Most usually, then, instead of deductive inferences being something achieved independently of our intuition, they are by and large founded on intuition, and by no means just on the intuition of the maning of signs for logical operations. The great task before us is to understand better the nature of such intuition and to sharply distinguish its varieties.

See also ANALYTICITY; LOGICAL POSITIVISM; LOGICISM.

## BIBLIOGRAPHY

- Detlefsen, M.: "Brouwer's Philosophy of Mathematics," in *Essays on Proof*, ed. M. Detlefsen (London: Routledge, 1991).
- Frege, G.: *The Basic Laws of Arithmetic*, trans. and ed. M. Furth (Berkeley, CA: University of California Press, 1964).
- Frege, G.: "Begriffsschrift, a formula language, modelled upon that of arithmetic, for pure thought," in *From Frege to Gödel: A Source Book in Mathematical Logic*, ed. J. van Heijenoort (Cambridge, MA: Harvard University Press, 1967).
- Hilbert, D.: *Foundations of Geometry*, 2nd English edn, trans. L. Unger (LaSalle, IL: Open Court, 1971).
- Hilbert, D. and Ackermann, W.: *Principles of Mathematical Logic*, trans. L. M. Hammond, G. G. Leckie, F. Steinhardt with R. E. Luce (New York: Chelsea, 1950).
- Katz, J. J.: *Cogitations* (Oxford: Oxford University Press, 1988).

- Koslow, A.: *A Structuralist Account of Logic* (Cambridge: Cambridge University Press, 1991).
- Locke, J.: *An Essay Concerning Human Understanding* (1690), ed. P. H. Nidditch (Oxford: Oxford University Press, 1975).
- Parsons, C.: "Intuition in Constructive Mathematics," in *Language, Mind and Logic*, ed. J. Butterfield (Cambridge: Cambridge University Press, 1986), 211–29.
- Rosen, S.: *The Limits of Analysis* (New Haven, CT: Yale University Press, 1980).
- Tragesser, R. S.: *Phenomenology and Logic* (Ithaca, NY: Cornell University Press, 1977).
- Tragesser, R. S.: "Three Little Noticed Aspects of Mathematical Proof," in Detlefsen (ed.), 1991.

ROBERT S. TRAGESSER

**intuition in epistemology** Epistemologists appeal to intuitions when they say it is obvious that  $2 + 3 = 5$ , that bachelors are unmarried males, that knowledge requires true belief, or that Gettier cases refute the JTB account of knowledge. It is widely agreed that such intuitions are specifically intellectual seemings that are analogous to, but different from, perceptual seemings. The Mueller–Lyer illusion is an example of the latter. In this illusion, two lines – one ending in arrowheads pointing outward, the other ending in arrowheads pointing inward – appear to be of different length even though they actually are equally long. Interestingly, the seeming or intuition that the line with the arrowheads pointing inward is longer than the other line remains even when one knows that the two lines are equally long. (See Bealer, 1998, p. 208, for the above.)

In contrast, the Monte Hall Paradox is an example of an intellectual intuition. Suppose you are a participant on "Let's Make a Deal", a game show whose host is Monte Hall. You may open one of three doors: A, B, C. Behind one of them is a big prize, behind the other two are, say, goats. Suppose you indicate your choice of door A. In response, the host opens another door – say, door B – behind which is a goat, and then asks you whether you want to choose door C instead. Is it to your advantage to switch your choice of doors? If you are like most people, it will intuitively seem to you that the likelihood of the big prize being

behind door A is the same as the likelihood of it being behind door C, even after you learn, say, through reliable testimony, that switching doors increases the probability of winning the prize.

In both examples, the way things seem to one is misleading; what seems to be the case is actually not the case. We must conclude, therefore, that intuitions are fallible. The two examples also show that an intuition that  $p$  is not necessarily accompanied by a belief that  $p$ . After one learns that the two lines are equally long, they still seem to be of different length, but one does not believe they are. Likewise, after one learns that switching doors increases one's chances at winning the prize, one no longer believes that sticking to one's original choice bears an equal chance even though it may still seem that way. Conversely, a belief that  $p$  is not necessarily accompanied by an intuition that  $p$ . Beliefs are stable mental states that persist through time. Intuitions, on the other hand, are episodic: they last only as long as something or other seems to one to be the case. For example, one's belief that no object can be wholly in two different places at the same time is not at any given time one holds that belief accompanied by an intuition with that content (see Bealer, 1998, p. 208).

So intuitions, conceived of as seemings, are occurrent mental states and have propositions as their objects. But what sort of propositions are the objects of intuitions? According to one view, the objects of intuitions are modal propositions, propositions of the form *necessarily p* or *possibly p* (Bealer, 1998, pp. 207–8; BonJour, 1998). According to an alternative view, it is possible to intuit  $p$  without intuiting *necessarily p* (Huston, 2003, pp. 48, 108). For instance, one might intuit that  $2 + 3 = 5$  without intuiting that this equation holds by necessity.

Bealer distinguishes between physical and philosophical intuitions (1998, p. 207). For example, one might have the physical intuition that a house whose foundation is undermined will fall. According to Bealer, this is not a *rational* intuition because the proposition does not present itself as necessary. What is intuited is not the proposition that a house

when undermined *must* fall. According to an alternative classification, we might want to distinguish between intuitions that are based solely on the understanding, and those that are not (BonJour, 1998, pp. 101–3, suggests this view even though he does not himself hold it). According to this approach, the intuition that  $2 + 3 = 5$  would be a *rational* intuition if it arises *solely* from understanding what the equation asserts, whereas the intuition that a house undermined will fall would not be a rational intuition since it is based partly on observation or testimony. (See Audi, 1997, p. 41, for the requirement that intuitions be based on the understanding.)

Although BonJour holds that rational insights take as their objects propositions of the form *necessarily p* (1998, pp. 114, 127), elsewhere he allows for the possibility that an unsophisticated person who lacks the concept of necessity might accept a proposition because it seems overwhelmingly obvious, not because it seems necessarily true (1998, p. 114, n. 23). According to BonJour, such an intuition could provide a significant degree of justification if it is a response to what is in fact “the apparent necessity of the proposition”. However, it would not be a full-fledged rational insight.

What makes intuitions a source of evidence? Some argue that intuitions are a basic source of evidence on the basis of concept possession (Bealer, 1998; Goldman, 2007). For example, anyone who possesses the concept of knowledge will for the most part judge correctly whether, in certain hypothetical cases, a person has knowledge. The reliability of such intuitions, qualifying them as a source of justification, is a consequence of possessing and understanding the concept of knowledge.

However, intuitions involving non-analytic statements, such as that happiness is an intrinsic good for the person who is happy, also seem to provide evidence. If so, the justification the intuitions provide must come from an understanding of the *relation between the concepts*, not merely from possession of the concepts themselves.

Of course, the plausibility of the defense of the view that intuitions are a basic source of

evidence that is based on the idea of concept possession depends on what concepts are and what is involved in their possession. The worry is that each person’s intuitions about, say, knowledge are evidence only for his or her conception of knowledge, not for some objective or intersubjective concept of knowledge (see Goldman, 2007).

Another argument says that intuitions provide evidence because it is impossible to argue for anything without appealing to intuition at some point. Suppose you are an empiricist and claim that we are justified in believing propositions about our physical environment on the basis of perceptual observations. But what justifies you in thinking that observation is a source of justification? Or suppose you are a coherentist and hold that if you have an appropriately coherent system of beliefs, then you are justified in holding each of the beliefs in that system. But what justifies you in holding that coherence can generate justification? Philosophers who hold that intuitions have evidential force will say that any answer to questions like these will appeal to intuition in some way or other. Even if you defend your substantive philosophical views through elaborate chains of deductive reasoning, each argument you employ will justify your views only if you *see* that the conclusion follows from the premises. However, seeing this is, of course, a form of intuition.

Some philosophers deny that intuitions are a source of justification. According to Hilary Kornblith, intuitions merely provide us with information about our folk concepts of, say, knowledge, personal identity, or happiness (Kornblith, 2006). But what is of genuine philosophical interest is the nature of the things to which the concepts refer, not our folk understanding of the concepts themselves. Just as our concepts of an acid, of gold, and of rubies may be based on error and ignorance, so may our concepts of knowledge, happiness, and justice. Critics who employ this objection claim that in epistemology we should be concerned with the *nature* of knowledge, not with the *concept* of knowledge. We can appeal to our rational intuitions to fix the subject matter of our

philosophical explorations, but beyond that intuitions have no further use. For example, our intuition that knowledge requires belief directs us to beliefs and away from, say, rocks when we begin our empirical inquiry into the nature of knowledge. Our rational intuition that acids are certain sorts of natural substances keeps us from starting with household furniture as part of the sample with which we will begin our empirical investigation into the essence of acid (see Kornblith, 2006, pp. 12–13, for the examples involving rocks and furniture). However, once it is settled what we are going to investigate, all the justification we can hope to find will arise not from any intuitions but from empirical observation.

An obvious reply to such an approach is to claim concepts such as “knowledge”, “justice”, and “happiness” are not natural kind terms such as “acid”, “water”, and “aluminium”, which are. The concepts that philosophers are interested in are socially constructed kinds; they are like concepts such as “SUV”, “bachelor”, or “marriage”. In response to this objection, Kornblith (2006, pp. 19–23) argues that, even if the concepts that philosophers are interested in are not natural kind concepts, people often are ignorant of the concept in question and make errors about its extension. My intuitions only reflect *my* conception of the kind in question, and your intuitions reflect *your* conception. Given, then, that intuitions reveal merely one’s own subjective understanding of a concept, why should philosophers be interested in them? For intuitions about a given concept to be of philosophical value, we would have to investigate empirically how that concept is used in a given society. Carrying out such an investigation, we would be engaged in an empirical inquiry – in social research – which is not really philosophy according to the defenders of the traditional appeal to intuition in philosophy (see Goldman, 2007).

To this line of criticism, defenders of intuitions might reply that, when philosophers appeal to intuitions, this appeal is part of an attempt to give an analysis of some non-natural kind of concept that they believe others will come to share *if they think and*

*reason well*. Philosophers are not trying to say what people in fact mean by some term. That *would* require an empirical investigation. Rather, when epistemologists appeal to intuitions about that nature of knowledge, they aim at making a proposal about what people *should* mean by “knowledge” once they have considered all the relevant examples and accompanying arguments that are employed in the conceptual investigation. Thus understood, conceptual analysis aims not at reporting what a single person means by some concept, nor at predicting what understanding of a concept other people in fact have. Rather, conceptual analysis and the appeal to intuitions that comes with it is essentially *normative*: its aim is to identify how we ought to understand a concept once we have properly examined it.

A normative understanding of conceptual analysis and the use of intuitions renders empirical research into conceptual diversity irrelevant. There are recent studies that allegedly show that students from different backgrounds have different intuitions about various philosophical issues having to do with Gettier cases and reliability as a necessary or sufficient condition of justification and knowledge. Bealer (1998) doubts that these studies involve intuitions *properly understood*; and, if they do not, the disagreements observed in the “intuitions” of different subjects is not relevant to the question of the reliability of intuitions in the strict sense. But, even if such real intuitional diversity exists, it remains unclear what its significance is supposed to be. The real question is whether convergence on intuitions and analyses will result after a large number of examples and accompanying arguments have been considered. If no convergence emerges, the question will be what the best explanation for the recalcitrant disagreement is. Perhaps it is that some people do not understand the concept as well as others, just as some people see better, and are better eyewitness reporters, than others (cp. Williamson, 2004, p. 150). Perhaps some people equivocate or make other errors in their arguments. Hence, an epistemologist who is confronted with diverging intuitions in others need not

conclude that her own intuitions are without evidential force. Rather, bearing in mind that appeal to intuitions is fallible, she should conclude that the lack of convergence is evidence of an incomplete understanding of the investigated concept, either *her own* or on the part of others involved in the dispute.

Critics of the view that appeal to intuitions constitutes a legitimate philosophical method have advanced an alternative view as to how philosophy should proceed. Engineers try to design cars that will meet certain desiderata such as ease of handling, good gas mileage, speed, comfort, safety, and good looks. Jonathan Weinberg advocates using a similar, pragmatic approach when determining what epistemic norms, and what epistemic methodology, we should adopt. Epistemic norms, according to Weinberg, should be “engineered” (2006, p. 35). As the relevant desiderata for epistemic norms, he identifies accuracy across time – diachronic reliability – and the fostering and resolution of evidence-based conversations – dialectical robustness (2006, pp. 36–9). He offers additional goals for meta-epistemology, which concerns the adoption of a method for selecting the appropriate epistemic norms.

Interestingly, Weinberg rejects other pragmatic ends to determine the proper meta-epistemology, such as fostering agreement among our peers (Rorty, 1989) and attaining what we find valuable (Stich, 1990; see Weinberg, 2006, p. 31, for these citations). One wonders what the rejection is based on. It would seem that what justifies Weinberg in rejecting these ends is that, *intuitively*, they do not strike him as *epistemic* ends even though they might be worthwhile non-epistemic ends.

Suppose we accept Weinberg’s list of ends as the relevant ones to use for assessing an epistemic method and selecting epistemic norms. In what way are we to judge how well a given epistemic norm furthers the relevant epistemic ends? Suppose, for instance, one end is to maximize truth in one’s belief system. What justifies us in believing a first norm, N1, better achieves this goal than a second norm, N2? We shall have to employ some meta-norm of justification to determine

whether we are *justified* in believing that N1 better achieves the relevant end than N2. On a pragmatic approach, this means we shall have to assess how well the meta-norm achieves some relevant epistemic end. But how shall we do this? It would appear that we shall need some meta-meta-norm. Which one would be the correct one to use? Shan’t we have to assess the meta-meta-norm on the basis of how well it promotes some epistemic goal? But assessing how well it promotes that goal will require appeal to a meta-meta-meta-norm. And so it goes, seemingly on to infinity. Some might appeal to epistemic intuitions to justify an epistemic norm at some level of ascent, but this appeal is not open to the pragmatist who is trying to offer an alternative approach to epistemology that relies crucially on the idea of epistemic goals and minimally, if at all, on rational intuitions.

To sum up, it would appear it is difficult to do epistemology without appealing to intuitions. First, for pragmatists, intuitions are needed to choose among various, competing epistemic ends and norms. Second, epistemic intuitions are needed to determine which norms serve our epistemic ends, and which beliefs meet our epistemic norms. Finally, there seems no way to avoid skepticism if intuitions are not a basic source of evidence.

#### BIBLIOGRAPHY

- Audi, R.: *Moral Knowledge and Ethical Character* (Oxford: Oxford University Press, 1997).
- Audi, R.: *The Architecture of Reason: The Structure and Substance of Rationality* (Oxford: Oxford University Press, 2001).
- Bealer, G.: “The Philosophical Limits of Scientific Essentialism,” in *Philosophical Perspectives I: Metaphysics*, ed. James E. Tomberlin (Atascadero, CA: Ridgeview, 1987), 289–365.
- Bealer, G.: “The Incoherence of Empiricism,” in *The Aristotelian Society Supplementary* 66 (1992), 99–137.
- Bealer, G.: “A Priori Knowledge and the Scope of Philosophy” and “A Priori Knowledge: Replies to William Lycan and Ernest Sosa,” in *Philosophical Studies* 81 (1996a), 121–42 and 163–74.
- Bealer, G.: “On the Possibility of Philosophical Knowledge,” in *Philosophical Perspectives* 10:



- Metaphysics*, ed. James E. Tomberlin (Atascadero, CA: Ridgeview, 1996b), 1–34.
- Bealer, G.: “Intuition and the Autonomy of Philosophy,” in *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, ed. Michael R. DePaul and William Ramsey (Lanham, MD: Rowman and Littlefield Publishers, 1998), 201–39.
- BonJour, L.: *In Defense of Pure Reason* (Cambridge: Cambridge University Press, 1998).
- DePaul, M. and Ramsey, W., eds: *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry* (Lanham, MD: Rowman and Littlefield Publishers, 1998).
- Goldman, A.: “Philosophical Intuitions: Their Target, Their Source, and Their Epistemic Status,” in *Grazer Philosophische Studien* 74 (2007), 1–25.
- Goldman, A. and Joel, P.: “Philosophical Theory and Intuitional Evidence,” in DePaul and Ramsey (eds), 1998, 179–97.
- Harman, G.: “Doubts about Conceptual Analysis,” in *Reasoning, Meaning, and Mind* (Oxford: Oxford University Press, 1999), 138–43.
- Hauser, M.: *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong* (New York: Harper Collins, 2006).
- Houston, M.: “Intuition: A Discussion of Recent Philosophical Views,” PhD dissertation, Wayne State University, 2003.
- Jackson, F.: *From Ethics to Metaphysics: A Defence of Conceptual Analysis* (Oxford: Clarendon Press, 1998).
- Kornblith, H.: “Precis” and “Replies to Alvin Goldman, Martin Kusch and William Talbott,” in *Philosophy and Phenomenological Research*, vol. 71 (2005), 399–402 and 427–41.
- Kornblith, H.: “Intuitions and Epistemology,” in *Epistemology Futures*, ed. Stephen Hetherington (Oxford: Oxford University Press, 2006), 10–25.
- Lycan, W.: “Bealer on the Possibility of Philosophical Knowledge,” in *Philosophical Studies* 81 (1996), 143–50.
- Sosa, E.: “Rational Intuition: Bealer on Its Nature and Epistemic Status,” in *Philosophical Studies* 81 (1996), 151–61.
- Sosa, E.: “Minimal Intuition,” in DePaul and Ramsey (eds), 1998, 257–69.
- Stratton-Lake, P. (ed.): *Ethical Intuitionism* (Oxford: Clarendon Press, 2002).
- Thomson, J. J.: “Introduction” to *The Realm of Rights* (Cambridge, MA: Harvard University Press, 1990), 1–33.
- Weatherston, B.: “What Good Are Counterexamples?” in *Philosophical Studies* 115, 1–31.
- Weinberg, J. M., Nichols, S., and Stich, S.: “Normativity and Epistemic Intuitions,” in *Philosophical Topics* 29 (2001), 429–60.
- Weinberg, J. M.: “Metaskepticism in Ethno-Epistemology,” in *The Sceptics*, ed. S. Luper (Aldershot: Ashgate Publishing, 2003), 227–47.
- Weinberg, J. M.: “Semantics, Cross-cultural Style,” in *Cognition* 92 (2004), B1–B12.
- Weinberg, J. M.: “What’s Epistemology For?,” in *Epistemology Futures*, ed. S. Hetherington (Oxford: Oxford University Press, 2006), 26–47.
- Weinberg, J. M.: “How to Challenge Intuitions without Risking Skepticism”, unpublished manuscript.
- Williamson, T.: “Philosophical ‘Intuitions’ and Scepticism about Judgement,” in *Dialectica* 58 (2004), 109–53.
- Williamson, T.: “Armchair Philosophy, Metaphysical Modality and Counterfactual Thinking,” in *Proceedings of the Aristotelian Society* 105 (2005), 1–23.
- Williamson, T.: “Philosophical Knowledge and Knowledge of Counterfactuals,” in *Grazer Philosophische Studien*, (2007).

BRUCE RUSSELL

**isostheneia** This is a term used by ancient Greek sceptics and is usually rendered into English by “equipollence”. It refers to the equal weight and credibility of opposing points of view discovered upon examining the evidence and arguments on both sides of a question. Greek sceptics made a practice of setting up oppositions between appearances and arguments to show that the theories of dogmatic philosophers are equally balanced and therefore equally unconvincing in their various claims to truth.

See also SEXTUS EMPIRICUS.

CHARLOTTE STOUGH

**James, William (1842–1910)** James, whose formal education was in chemistry, comparative anatomy, physiology and medicine, taught physiology and psychology before teaching philosophy. His entire career as student and teacher was spent at Harvard University.

Although with characteristic generosity he exaggerated his debt to Peirce (*see* PEIRCE), James cofounded pragmatism (*see* PRAGMATISM). His “Remarks on Spencer’s Definition of Mind as Correspondence” was published in the same month of 1878 that Peirce’s “How to Make Our Ideas Clear” appeared. Peirce introduced to the world the “pragmatic maxim”, while James outlined the interconnected set of doctrines that came to be associated with the label Peirce first used. James’s epistemology is found in both his pragmatism and his “radical empiricism”.

#### PRAGMATISM

From his earliest writing James understood cognitive processes in teleological terms; thought, he held, assists us in the satisfaction of our interests. His Will to Believe doctrine, the view that we are sometimes justified in believing beyond the evidence, relies on the notion that a belief’s benefits are relevant to its justification. His pragmatic method of analysing philosophical problems, which requires that we find the meaning of terms by examining their applications to objects in experimental situations, similarly reflects a teleological approach in its attention to consequences (James, 1975a, pp. 27–44).

Such an approach sets James’s theory of meaning apart from verificationism’s dismissive of metaphysics. Unlike the verificationist (*see* VERIFICATIONISM), who takes cognitive

meaning to be a matter only of consequences in sensory experience, James took pragmatic meaning to include emotional and motor responses. Moreover, his method was a way of clarifying the meanings of metaphysical propositions, not a way of dismissing them as meaningless (James, 1975a, pp. 45–62). It should also be noted that in his more circumspect moments James did not hold that even his broad set of consequences was exhaustive of a term’s meaning. “Theism”, for example, he took to have antecedent, definitional meaning in addition to its more important pragmatic meaning (Giuffrida and Madden, 1975, pp. 18–35).

James’s theory of truth reflects his teleological conception of cognition by considering a true belief to be one which is compatible with our existing system of beliefs and leads us to satisfactory interaction with the world (James, 1975b, *passim*). Although many commentators have supposed that James was thus committed to subjectivism (*see* SUBJECTIVISM), it has been persuasively argued by H. S. Thayer that James did not intend his speaking of truth in these terms to replace the concept of truth as a relation to reality (Thayer, 1981, pp. 527–56). Instead, James was insisting that objective reference was only one of three conditions of truth. Unfortunately, in both technical polemics and popular lectures James often neglected to mention objective reference.

#### RADICAL EMPIRICISM

Although James always considered himself an empiricist, from his earliest writing he was intent on modifying the concept of experience traditional in empiricism. Central to this modification was the claim that we are

directly acquainted with relations – temporal relations, causal relations, etc. In some of the most vivid philosophy ever written, he described how we directly experience continuities in the world (1976, pp. 21–44, 79–95). Of equal importance was his functional theory of consciousness in which the existence of mental entities was rejected in favour of functional processes (1976, pp. 3–19).

Like his theory of truth, James's radical empiricism has led many commentators to suppose that James rejected epistemological realism (i.e. the view that objects exist independently of the perceiver and can be known as they are) in favour of phenomenalism (see PHENOMENALISM). Regrettably, James's speaking sometimes of the objects of knowledge as "pure experience" was misleading. Despite his insistence that pure experience was neither mental nor physical, readers have found it difficult to avoid the conclusion that he is speaking of mental entities. Fortunately, more discerning commentators have understood that James was attempting to use his radical empiricism to support the "natural realism" he repeatedly avowed throughout his career (Madden and Chakrabarti, 1976, pp. 3–17). He wished to use a theory of experience in which relational continuities are fundamental to substitute an "ambulatory" conception of the knower–known relation for a "saltatory" one (James, 1975b, pp. 13–32, 78–89). In the saltatory view, it is supposed that there is a gap between knower and known requiring a self-transcending leap, while in the view James proposed, "ambulation" mediates in a continuous process between knower and known, a process different from the messenger mediation found in representative realism (see REPRESENTATIVE REALISM).

Undeniably, James in his earlier work was unwilling in print to declare himself a realist (see REALISM). This reluctance is understandable if one is sensitive to James's intellectual biography (Hare and Chakrabarti, 1980, pp. 231–45). Although James throughout his life believed in natural realism, for many years he felt intensely his inability to marshal convincing arguments for such an epistemology. Only late in life, when he had fully developed

both his radical empiricism and his theory of truth, did he feel that he possessed adequate justification for the realism to which he had always been committed.

See also DEWEY; KNOWLEDGE BY ACQUAINTANCE/BY DESCRIPTION; PRAGMATISM; PEIRCE.

#### WRITINGS

*Pragmatism: A New Name for Some Old Ways of Thinking* (Cambridge, MA: Harvard University Press, 1975).

*The Meaning of Truth: A Sequel to Pragmatism* (Cambridge, MA: Harvard University Press, 1975).

*Essays in Radical Empiricism* (Cambridge, MA: Harvard University Press, 1976).

#### BIBLIOGRAPHY

Giuffrida, R. and Madden, E. H.: "James on Meaning and Significance," *Transactions of the C.S. Peirce Society* 11 (1975), 18–35.

Hare, P. H. and Chakrabarti, C.: "The Development of William James's Epistemological Realism," in *History, Religion and Spiritual Democracy*, ed. M. Wohlgeleirter (New York: Columbia University Press, 1980), 231–45.

Madden, E. H. and Chakrabarti, C.: "James' 'Pure Experience' versus Ayer's 'Weak Phenomenalism,'" *Transactions of the C. S. Peirce Society* 12 (1976), 3–17.

Myers, G. E.: *William James: His Life and Thought* (New Haven, CT: Yale University Press, 1986), ch. 10.

Thayer, H. S.: *Meaning and Action: A Critical History of Pragmatism*, 2nd edn (Indianapolis, IN: Hackett, 1981), 527–56.

PETER H. HARE

**judgement** This term may refer to a faculty (of judgement), an act (of judging), or the product of an act (what is judged). My judging that this is warm might be thought to consist in my coming to be in a state with the content, "This is warm", as a result of the operation of a particular mental faculty of judgement (distinguishable from other faculties: will, imagination, memory). Kant equates

judging with the application of concepts: In judging this to be warm, I judge this to satisfy the concept *warm*.

Aristotle identified four logical forms of judgement – All S are P, Some S are P, No S are P, and Some S are not-P – noting that some combinations of these basic forms yield valid arguments, while others do not, an idea that survives in modern logic.

Whatever their source, acts of judgement may be spontaneous or result from theoretical or practical deliberation. Thus, you may judge that this is warm simply on the basis of touching it; I judge the same on the basis of evidential considerations. Practical judgement invokes values and may be partial (this is politically best) or all-out (this is best *tout court*).

#### BIBLIOGRAPHY

- Aristotle: *Prior Analytics* trans. W. D. Ross (Oxford: Clarendon Press, 1949).  
 Kant, I.: *Critique of Pure Reason* (1781), trans. N. Kemp Smith (London: Macmillan, 1964).

JOHN HEIL

**justification** According to the justified true belief (JTB) account of knowledge, justification is needed to turn true belief into knowledge. True belief by itself is not

sufficient for knowledge, for a belief can be true due to luck, and a belief's being true because of luck is commonly considered incompatible with the belief's being an instance of knowledge. However, as Gettier cases show, justified true belief is not sufficient for knowledge, either. Hence, for a JTB-type analysis of knowledge to be successful, it would have to be amended with a fourth condition, the purpose of which would be to rule out Gettier cases.

Advocates of the JTB account tend to consider justification an internal matter, that is, a matter of the subject's having adequate evidence. Some externalist critics of the JTB account argue that, since animals and infants have knowledge, internal justification is not a necessary condition of knowledge. According to these critics, justification should be construed as an external matter, namely as a belief's property of being the outcome of a reliable cognitive process. Other critics of the JTB approach dispense with an account of justification altogether and instead offer an analysis of knowledge using alternative concepts.

*See also* CAUSAL THEORIES IN EPISTEMOLOGY; COHERENTISM; EPISTEMIC LUCK; EVIDENCE; EXTERNALISM/INTERNALISM; FOUNDATIONALISM; GETTIER PROBLEM; PRAGMATISM; PROPOSITIONAL KNOWLEDGE; RELIABILISM; RELATIVISM; SUBJECTIVISM.

MATTHIAS STEUP

# K

---

**Kant, Immanuel (1724–1804)** Kant is often regarded as the greatest of the modern philosophers. He spent his entire life in or near the East Prussian city of Königsberg (now Kaliningrad), holding the post of Professor of Philosophy at the University of Königsberg from 1770 onwards. His best-known works are his three *Critiques*: the *Critique of Pure Reason* (1781, with a second edition in 1787), which deals with epistemology and metaphysics; the *Critique of Practical Reason* (1788), which deals with ethics, and the *Critique of Judgement* (1790), which deals with aesthetics and teleology. Other important works are the *Dissertation on the Form and Principles of the Sensible and Intelligible World* (1770), *Prolegomena to Any Future Metaphysics* (1783), *Foundations of the Metaphysics of Morals* (1785), *Metaphysical Foundations of Natural Science* (1786), and *Religion Within the Limits of Reason Alone* (1793). His contributions to epistemology are contained mainly in the first of the three *Critiques*.

The *Critique of Pure Reason* has both positive and negative aims: its task is to “institute a tribunal which will assure to reason its lawful claims, and dismiss all groundless pretensions” (A xi). In other words, Kant seeks to determine the scope and possibility of a priori knowledge, defending such knowledge against sceptical suspicion in areas where it is legitimate and exposing its lack of credentials in areas where it is not.

## SYNTHETIC A PRIORI JUDGEMENTS

In the Introduction to the *Critique*, Kant draws three important distinctions: a priori vs. empirical, necessary vs. contingent, and analytic vs. synthetic. Each of these has its own entry in this volume, which the reader

should consult. Briefly, a proposition or judgement is knowable a priori if it is knowable without relying on experience; otherwise it is empirical. A true proposition is necessary if it is true not just in the actual world, but in any possible world; otherwise it is contingent. Kant believes these two distinctions divide up the field of truths in exactly the same way: a proposition is necessary if and only if it is knowable a priori and contingent if and only if it is knowable (if at all) only empirically.

The third distinction, between analytic and synthetic judgements, cuts across the two distinctions just mentioned. In an analytic judgement, the predicate belongs to the concept of the subject (as in “all bodies are extended”); in a synthetic judgement, the predicate lies outside the subject concept (as in “all bodies are heavy”). For our purposes, we can perhaps say that an analytic judgement is one in which the predicate belongs to the definition of the subject concept, or can at least be derived from the subject concept using only definitions and logical laws. That would bring Kant’s analytic/synthetic distinction into alignment with more recent versions of the distinction, e.g., Frege (see FREGE). Many philosophers – including Leibniz (see LEIBNIZ) and Hume (see HUME) among Kant’s predecessors and the logical positivists among his successors – have held that a priori knowledge is to be had only of analytic propositions. This view, if true, would take much of the mystery out of a priori knowledge. Kant, however, was convinced that there are important classes of propositions that are both a priori and synthetic. Such propositions are to be found in abundance in arithmetic and geometry (e.g.  $7 + 5 = 12$ , a straight line is the shortest distance between two points), and they include as well certain framework propositions



of natural science (e.g. every event has a cause). In addition, most of the propositions of traditional metaphysics (e.g., the soul is a substance, the compound must be composed of the simple) are synthetic and *purportedly* a priori. Kant therefore set the following as the central question of the *Critique of Pure Reason*: how are synthetic a priori propositions possible?

#### GEOMETRICAL KNOWLEDGE

What assures us of the truth of propositions of geometry (*see* GEOMETRY)? We know many of them because we prove them from others, but how do we ascertain the truth of those we do not need to prove? It is not by finding that they are analytic, for “however long [one] meditates on [the subject] concept, he will never produce anything new” (A716/B744). Nor is it by relying on observation (e.g. inspecting dozens of triangular objects and measuring their angle sums), for that would not yield “universality, still less necessity” (A718/B746). The answer is rather that we exhibit to ourselves by “pure intuition” (imagination or visualization) an object answering to the concept and read off further properties not contained in the concept itself. For example, a cube is defined as a regular solid composed of six square faces; there is nothing said about the number of edges. Yet by visualizing a cube and counting, we can see that any cube must have twelve edges. A simple act of intuition thus convinces us that we will never encounter a cube having more than twelve edges.

How is it possible for intuition thus to “anticipate” its objects, to give us knowledge of the properties of objects in advance of our experiential encounters with them? To answer this question, Kant instituted his “Copernican Revolution” in philosophy:

If intuition must conform to the constitution of the objects, I do not see how we could know anything of the latter *a priori*; but if the object (as object of the senses) must conform to our faculty of intuition, I have no difficulty in conceiving such a possibility (B xvii).

Ptolemy tried to explain the apparent motion of the heavenly bodies about the earth by attributing this motion to the bodies themselves; Copernicus fared better by explaining the apparent motion by referring to the observer’s own motion. Comparing his own strategy with that of Copernicus, Kant proposed to explain many of the observed features of objects by reference to traits of the observer rather than to traits of the objects themselves. In the case at hand, he sought to explain the geometrical properties of objects in terms of the structure of the human knower, or what Kant called our “form of intuition”.

Now what manner of object must conform to our own form of intuition? It is hard indeed to see why *things in themselves* – that is, things whose existence is in no way dependent on human cognition – should so conform. Hence Kant unhesitatingly draws the conclusion that the objects of geometrical knowledge must not be things in themselves:

If the object (the triangle) were something in itself, apart from any relation to you, the subject, how could you say that what necessarily exist in you as subjective conditions for the construction [i.e. exhibition in pure intuition] of a triangle, must of necessity belong to the triangle itself?

(A48/B65)

The objects of geometrical knowledge potentially include all spatial configurations – everything that exists in space. Hence, it is a corollary of Kant’s Copernican strategy that spatial features cannot belong to things in themselves. This is the doctrine he calls transcendental idealism: things in space (and as he also holds, things in time) are merely appearances, not things in themselves.

By an appearance, Kant means something that exists only as the object of a representation (or perhaps as the possible object of a representation). Since appearances exist only if they are representable, and since they are representable only if they conform with the laws of Euclidean geometry, it can be known in advance that Euclid’s laws hold of all appearances.

Many objections have been raised against Kant’s account, more than can be discussed

here. Some have cited the rise of alternative geometries as casting doubt on the alleged necessity of Euclid's geometry. Some have questioned whether intuition plays any legitimate role in geometrical knowledge. Some have denied that the laws of geometry are *a priori*, assimilating them rather to the laws of physics, which are known in virtue of the way they help us to systematize empirical data. Some, notably Quine (*see* QUINE), have challenged the analytic–synthetic distinction and questioned whether anything at all is *a priori*.

I shall discuss here only one objection to Kant's view, a composite of objections raised by Russell and Moore. Kant sought to account for the necessity of arithmetical and geometrical truths, and for our ability to know them *a priori*, by appeal to the structure of our cognitive faculties. But it is contingent that our faculties are the way they are; so if Kant is right, might we not wake up tomorrow and find that cubes have thirteen edges? Or if that is *not* possible, if we can rule out *a priori* any such change in our constitution, what is the ground of that piece of *a priori* knowledge? It is evidently unaccounted for by Kant's theory.

This is a difficult dilemma, and unless Kant wants to water down the necessity he claims for mathematics (to something like "truth in all worlds experienceable by beings constituted as we are now"), there may be no escaping it. It is worth pointing out, however, that a similar objection applies with equal force to the leading theory of *a priori* knowledge developed in the twentieth century as an alternative to Kant's. This is conventionalism, the theory that necessary truth and *a priori* knowledge are the products of convention. If the role of convention is merely to map certain sentences onto pre-existing necessarily true propositions, the objection I am about to raise does not apply. But most conventionalists have been more ambitious, seeing convention as the very source of necessity. Since it is contingent that we have the conventions we do instead of some others, these more thoroughgoing conventionalists have made the same mistake as Kant – they have tried to account for the necessary in terms of the contingent.

## INTUITION AND CONCEPT

According to Kant, our knowledge arises from two fundamentally different faculties of the mind, sensibility and understanding. He criticized his predecessors for running these faculties together, Leibniz for treating sensing as a confused mode of understanding and Locke for treating understanding as an abstracted mode of sensing. Kant held that each of the faculties operates with its own distinctive type of mental representation. Concepts, the instruments of the understanding, are mental representations that apply potentially to many things in virtue of their possession of a common feature. Intuitions, the instruments of sensibility, are representations that refer to just one thing and to that thing directly. Intuitions play the role that is played in Russell's philosophy by "acquaintance" (*see* KNOWLEDGE BY ACQUAINTANCE/BY DESCRIPTION). Through intuitions objects are given to us, Kant said; through concepts they are thought.

It is a famous Kantian thesis that knowledge is yielded neither by intuitions nor by concepts alone, but only by the two in conjunction. "Thoughts without content are empty," he says in an often quoted remark, and "intuitions without concepts are blind" (A51/B75). Exactly what Kant means by the remark is a debated question, however, answered in different ways by scholars who bring different elements of Kant's text to bear on it. A minimal reading is that it is only propositionally structured knowledge that requires the collaboration of intuition and concept; this view allows that intuitions without concepts constitute some kind of non-judgemental awareness. A stronger reading is that it is reference or intentionality that depends on intuition and concept together, so that the blindness of intuition without concept is its referring to no object. A more radical view yet is that intuitions without concepts are indeterminate, a mere blur, perhaps nothing at all. This last interpretation, though admittedly suggested by some things Kant says, is at odds with his official view about the separation of the faculties.

## A PRIORI CONCEPTS

“A priori origin is manifest in certain concepts,” Kant writes, “no less than in judgements” (B5). Kant was thus not only a judgement rationalist, i.e. a believer in synthetic a priori judgements (see RATIONALISM), but also a concept rationalist, i.e. a believer in a priori concepts. He believed that there are certain concepts that are not abstracted from experience (nor compounded out of concepts so abstracted), but which are applicable to objects of experience none the less. These are his categories, of which substance and cause are perhaps the most important. He realized the puzzling nature of these concepts in a letter to his pupil Marcus Herz in 1772. How is it possible, he asked, for a concept to “relate to an object” if the concept has not been derived from objects (as happens with empirical concepts), nor the object brought into being by the concept (as Kant believed happens with the divine understanding)? This is the problem he addressed in the “transcendental deduction of the categories”. It is impossible to do more here than give the barest outline of the transcendental deduction. Kant’s starting point is the *unity of apperception* (see APPERCEPTION) – the fact that all my representations have the property of being co-apprehensible in one consciousness. Kant argues that representations have such unity only because they have been synthesized according to rules encapsulated in the categories. As a result, (1) there are objects to which the representations refer, and (2) the categories apply to these objects. If successful, the transcendental deduction simultaneously demonstrates the “objective validity” of the categories – their applicability to objects of experience – and the objectivity of our experience – its having objects corresponding to it, rather than being a mere play of representations. Kant’s main concern was the validity of the categories, but the objectivity of experience has seemed a more significant result to many contemporary philosophers, such as Strawson (see STRAWSON), who have seen in Kant the promise of a transcendental argument (see TRANSCENDENTAL ARGUMENTS) against scepticism (see SCEPTICISM). It must be borne

in mind, however, that the ‘reference to objects’ secured by the transcendental deduction is to be understood in accordance with the idealist strategy of Kant’s Copernican Revolution. The objects are not mind-independent things in themselves, but phenomenal objects constructed out of patterns of representations.

## THE LIMITS OF A PRIORI KNOWLEDGE

As was said at the beginning, Kant’s purposes in the *Critique of Pure Reason* are both constructive and critical. He wants to give a theory that will explain and defend our legitimate claims to a priori knowledge, but at the same time to expose the pretensions of those that are spurious. It turns out that the factors making a priori knowledge possible (our forms of intuition in the case of arithmetic and geometry, our category-governed modes of synthesis in the case of natural science) can operate only on sense-given materials, so that our a priori knowledge is limited to the world of experience. The *source* of a priori knowledge is not experience, but its only legitimate *subject matter* is objects of possible experience – such is Kant’s compromise between rationalism and empiricism. In the portion of the *Critique* entitled “Transcendental Dialectic”, he argues that we can have no knowledge of the properties of the soul, the outer limits and inmost nature of the cosmos, or the existence of God. As matters transcending possible experience, these can be matters of faith or speculation, but not of knowledge.

See also A PRIORI KNOWLEDGE in Part I; ANALYTICITY; IDEALISM; IN ITSELF/FOR ITSELF; MATHEMATICAL KNOWLEDGE; NOUMENAL/PHENOMENAL; REALISM; SELF-KNOWLEDGE AND SELF-IDENTITY; TRANSCENDENTAL ARGUMENTS.

## WRITINGS

*Kritik der reinen Vernunft* (Riga, 1781), 2nd edn (Riga, 1787); trans. N. Kemp Smith as *Critique of Pure Reason* (London: Macmillan, 1964).

*Dissertation on the Form and Principles of the Sensible and Intelligible World*, trans. G. B. Kerferd and D. E. Walford, in *Kant: Selected Pre-Critical Writings* (Manchester: Manchester University Press, 1968).

*Prolegomena zu einer jeden Metaphysik*; trans. L. W. Beck as *Prolegomena to Any Future Metaphysics* (Indianapolis, IN: Bobbs-Merrill, 1950).

## BIBLIOGRAPHY

Allison, H.: *Kant's Transcendental Idealism* (New Haven, CT: Yale University Press, 1973).

Ameriks, K.: "Recent Work on Kant's Theoretical Philosophy," *American Philosophical Quarterly* 19 (1982), 1–24.

Aquila, R.: *Representational Mind: A Study of Kant's Theory of Knowledge* (Bloomington, IN: Indiana University Press, 1983).

Beck, L. W.: "Did the Sage of Königsberg Have No Dreams?," in his *Essays on Kant and Hume* (New Haven, CT: Yale University Press, 1978), 38–60.

Bennett, J.: *Kant's Analytic* (Cambridge: Cambridge University Press, 1966).

Bird, G.: *Kant's Theory of Knowledge* (London: Routledge and Kegan Paul, 1962).

Broad, C. D.: *Kant: An Introduction* (Cambridge: Cambridge University Press, 1978).

George, R.: "Kant's Sensationism," *Synthese* 47 (1981), 229–55.

Guyer, P.: *Kant and the Claims of Knowledge* (Cambridge: Cambridge University Press, 1987).

Paton, H. J.: *Kant's Metaphysics of Experience* (London: Allen and Unwin, 1936).

Prichard, H. A.: *Kant's Theory of Knowledge* (Oxford: Clarendon Press, 1909).

Sellars, W.: *Science and Metaphysics: Variations on Kantian Themes* (London: Routledge and Kegan Paul, 1968).

Smith, N. Kemp: *A Commentary on Kant's Critique of Pure Reason*, 2nd edn (London: Macmillan, 1923).

Strawson, P. F.: *The Bounds of Sense* (London: Methuen, 1966).

Van Cleve, J.: *Problems from Kant* (Oxford University Press, 1999).

Wolff, R. P.: *Kant's Theory of Mental Activity* (Cambridge, MA: Harvard University Press, 1963).

JAMES VAN CLEEVE

**KK-thesis** The thesis that knowing entails knowing that one knows, often symbolized in

epistemic logic as " $Kp \rightarrow KKp$ ", where "K" stands for the concept of knowing. According to the KK-thesis, the logic of knowledge resembles the modal system S4. The KK-thesis was introduced into the contemporary epistemological discussion by Jaakko Hintikka in *Knowledge and Belief* (1962) (see HINTIKKA), but a tacit or an explicit acceptance of the thesis has been part of many philosophers' views about knowledge since Plato and Aristotle. The validity of the thesis is sensitive to shifts in the sense of "know"; it has often been thought to characterize a strong concept of knowledge (e.g. knowledge based on conclusive grounds), or active as opposed to implicit knowledge.

## BIBLIOGRAPHY

Hintikka, J.: *Knowledge and Belief: An Introduction to the Logic of the Two Notions* (Ithaca, NY: Cornell University Press, 1962).

*Synthese* 21: 2 (1970).

RISTO HILPINEN

**knower paradox** see PARADOX OF THE KNOWER.

**knowledge** see CAUSAL THEORIES IN EPISTEMOLOGY; COHERENTISM; DIFFERENT CONSTRUCTIONS IN TERMS OF "KNOWS"; FOUNDATIONALISM; GETTIER PROBLEM; KNOWLEDGE AND BELIEF; PROPOSITIONAL KNOWLEDGE; RELIABILISM.

**knowledge and belief** According to most epistemologists, knowledge entails belief, so that I cannot know that such and such is the case unless I believe that such and such is the case. Others think this *entailment thesis* can be rendered more accurately if we substitute for belief some closely related attitude. For instance, several philosophers would prefer to say that knowledge entails psychological certainty (see CERTAINTY) (Prichard, 1950; Ayer, 1956) or conviction (Lehrer, 1974) or acceptance (Lehrer, 1989). None the less, there are arguments against all versions of the thesis that knowledge requires having a belief-like attitude toward the known. These arguments

are given by philosophers who think that knowledge and belief (or a facsimile) are mutually incompatible (the *incompatibility thesis*), or by ones who say that knowledge does not entail belief, or vice versa, so that each may exist without the other, but the two may also coexist (the *separability thesis*).

The incompatibility thesis is sometimes traced to Plato (*see* PLATO) in view of his claim that knowledge is infallible while belief or opinion is fallible (*Republic* 476–9). But this claim would not support the thesis. Belief might be a component of an infallible form of knowledge in spite of the fallibility of belief. Perhaps knowledge involves some factor that compensates for the fallibility of belief.

A. Duncan-Jones (1966; cf. also Vendler, 1978) cites linguistic evidence to back up the incompatibility thesis. He notes that people often say “I don’t believe she is guilty, I *know* she is!” and the like, which suggests that belief rules out knowledge. However, as Lehrer (1974) indicates, the above exclamation is only a more emphatic way of saying, “I don’t *just* believe she is guilty, I know she is!” where “just” makes it especially clear that the speaker is signalling that she has something more salient than mere belief, not that she has something inconsistent with belief, namely knowledge. Compare: “You didn’t hurt him, you killed him!”

H. A. Prichard (1950) offers a defense of the incompatibility thesis which hinges on the equation of knowledge with certainty (both infallibility and psychological certitude) and the assumption that when we believe in the truth of a claim we are not certain about its truth. Given that belief always involves uncertainty while knowledge never does, believing something rules out the possibility of knowing it. Unfortunately, however, Prichard gives us no good reason to grant that states of belief are never ones involving confidence. Conscious beliefs clearly involve some level of confidence; to suggest that we cease to believe things about which we are *completely* confident is bizarre.

A. D. Woozley (1953) defends a version of the separability thesis. Woozley’s version, which deals with psychological certainty rather than belief *per se*, is that knowledge

can exist in the absence of confidence about the item known, although knowledge might also be accompanied by confidence as well. Woozley remarks that the test of whether I know something is “what I can do, where what I can do may include answering questions”. On the basis of this remark he suggests that even when people are unsure of the truth of a claim, they might know that the claim is true. We unhesitatingly attribute knowledge to people who give correct responses on examinations even if those people show no confidence in their answers. Woozley acknowledges, however, that it would be *odd* for those who lack confidence to claim knowledge. It would be peculiar to say, “I am unsure whether my answer is true; still, I know it is correct.” But this tension Woozley explains using a distinction between conditions under which we are justified in making a claim (such as a claim to know something), and conditions under which the claim we make is true. While “I know such and such” might be true even if I am unsure whether such and such holds, none the less it would be inappropriate for me to claim that I know that such and such unless I were sure of the truth of my claim.

Colin Radford (1966) extends Woozley’s defence of the separability thesis. In Radford’s view, not only is knowledge compatible with the lack of certainty, it is also compatible with a complete lack of belief. He argues by example. In one example, Jean has forgotten that he learned some English history years prior and yet he is able to give several correct responses to questions such as “When did the Battle of Hastings occur?” Since he forgot that he took history, he considers his correct responses to be no more than guesses. Thus when he says that the Battle of Hastings took place in 1066 he would deny having the *belief* that the Battle of Hastings took place in 1066. *A fortiori* he would deny being sure (or having the right to be sure) that 1066 was the correct date. Radford would none the less insist that Jean knows when the Battle occurred, since clearly he remembers the correct date. Radford admits that it would be inappropriate for Jean to *say* that he knew when the Battle of Hastings occurred, but,



like Woozley, he attributes the impropriety to a fact about when it is and is not appropriate to *claim* knowledge. When we claim knowledge, we ought at least to believe that we have the knowledge we claim, or else our behavior is “intentionally misleading”.

Those who agree with Radford’s defence of the separability thesis will probably think of belief as an inner state that can be detected through introspection. That Jean lacks beliefs about English history is plausible on this Cartesian picture since Jean does not find himself with any beliefs about English history when he seeks them out. One might criticize Radford, however, by rejecting the Cartesian view of belief (*see* BELIEF). One could argue that some beliefs are thoroughly unconscious, for example. Or one could adopt a behaviourist (*see* BEHAVIOURISM) conception of belief, such as Alexander Bain’s (1859), according to which having beliefs is a matter of the way people are disposed to behave (and hasn’t Radford already adopted a behaviourist conception of knowledge?). Since Jean gives the correct response when queried, a form of verbal behaviour, a behaviourist would be tempted to credit him with the belief that the Battle of Hastings occurred in 1066.

D. M. Armstrong (1973) takes a different tack against Radford. Jean does know that the Battle of Hastings took place in 1066. Armstrong will grant Radford that point. In fact, Armstrong suggests that Jean believes that 1066 is *not* the date the Battle of Hastings occurred, for Armstrong equates the belief that such and such is just possible but no more than just possible with the belief that such and such is not the case. However, Armstrong insists, Jean also believes that the Battle *did* occur in 1066. After all, had Jean been mistaught that the Battle occurred in 1060, and had he forgotten being “taught” this and subsequently “guessed” that it took place in 1060, we would surely describe the situation as one in which Jean’s false belief about the Battle became unconscious over time but persisted as a memory trace that was causally responsible for his guess. Out of consistency, we must describe Radford’s original case as one in which Jean’s *true* belief became unconscious but persisted long enough to cause his guess.

Thus while Jean consciously believes that the Battle did not occur in 1066, unconsciously he does believe it occurred in 1066. So after all Radford does not have a counterexample to the claim that knowledge entails belief.

Armstrong’s response to Radford was to reject Radford’s claim that the examinee lacked the relevant belief about English history. Another response is to argue that the examinee lacks the knowledge Radford attributes to him (cf. Sorensen, 1982). If Armstrong is correct in suggesting that Jean believes both that 1066 is and that it is not the date of the Battle of Hastings, one might deny Jean knowledge on the grounds that people who believe the denial of what they believe cannot be said to know the truth of their belief. Another strategy might be to liken the examinee case to examples of ignorance given in recent attacks on externalist (*see* EXTERNALISM/INTERNALISM) accounts of knowledge (needless to say, externalists themselves would tend not to favour this strategy). Consider the following case developed by Bonjour (1985): For no apparent reason, Samantha believes that she is clairvoyant. Again for no apparent reason, she one day comes to believe that the President is in New York City, even though she has every reason to believe that the President is in Washington, DC. In fact, Samantha is a completely reliable clairvoyant, and she has arrived at her belief about the whereabouts of the President through the power of her clairvoyance. Yet surely Samantha’s belief is completely irrational. She is not justified in thinking what she does. If so, then she does not know where the President is. But Radford’s examinee is little different. Even if Jean lacks the belief which Radford denies him, Radford does not have an example of knowledge that is unattended with belief. Suppose that Jean’s memory had been sufficiently powerful to produce the relevant belief. As Radford says, Jean has every reason to suppose that his response is mere guesswork, and so he has every reason to consider his belief false. His belief would be an irrational one, and hence one about whose truth Jean would be ignorant.

*See also* PROPOSITIONAL KNOWLEDGE.

BIBLIOGRAPHY

- Armstrong, D. M.: *Belief, Truth and Knowledge* (Cambridge: Cambridge University Press, 1973).  
 Ayer, A. J.: *The Problem of Knowledge* (Harmondsworth: Penguin Books, 1956).  
 Bain, A.: *The Emotions and the Will* (London: Longmans, Green, 1859).  
 Bonjour, L.: *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1985).  
 Duncan-Jones, A.: "Further Questions about 'know' and 'think,'" in *Philosophy and Analysis*, ed. M. MacDonald (Oxford: Blackwell, 1966).  
 Lehrer, K.: *Knowledge* (Oxford: Oxford University Press, 1974).  
 Lehrer, K.: "Knowledge Reconsidered," in *Knowledge and Skepticism*, ed. M. Clay and K. Lehrer (Boulder, Colo.: Westview, 1989).  
 Prichard, H. A.: *Knowledge and Perception* (Oxford: Clarendon Press 1950).  
 Radford, C.: "Knowledge – by examples," *Analysis* 27 (1966), 1–11.  
 Sorensen, R. A.: "Knowing, Believing, and Guessing," *Analysis* 42 (1982), 212–13.  
 Vendler, Z.: *Res Cogitans* (Ithaca, NY: Cornell University Press, 1978).  
 Woozley, A. D.: "Knowing and Not Knowing," *Proceedings of the Aristotelian Society* 53 (1952–3), 151–72.

STEVEN LUPER

**knowledge by acquaintance/by description**

The expressions "knowledge by acquaintance" and "knowledge by description", and the distinction they mark between knowing *things* and knowing *about* things, are now generally associated with Bertrand Russell (see RUSSELL). However, John Grote and Hermann von Helmholtz had earlier and independently used essentially similar terminology to mark the same distinction, and William James (see JAMES) adopted Grote's terminology in his investigation of the distinction. Philosophers have perennially investigated this and related distinctions using varying terminology.

GROTE AND HELMHOLTZ

Grote introduced the distinction by noting that natural language "distinguishes between

these two applications of the notion of knowledge, the one being *γινῶναι*, *noscere*, *kennen*, *connaître*, the other being, *εἰδεῖναι*, *scire*, *wissen*, *savoir*" (Grote, 1865, p. 60). On Grote's account, the distinction is a matter of degree, and there are three sorts of dimensions of variability: epistemic, causal and semantic.

We know things by experiencing them, and knowledge of acquaintance (Russell changed the proposition to "by") is epistemically prior to and has a relatively higher degree of epistemic justification than knowledge about things. Indeed, sensation has "the one great value of trueness or freedom from mistake" (1900, p. 206).

A thought (using that term broadly, to mean any mental state) constituting knowledge of acquaintance with a thing is more or less causally proximate to sensations caused by that thing, while a thought constituting knowledge about the thing is more or less distant causally, being separated from the thing and experience of it by processes of attention and inference. At the limit, if a thought is maximally of the acquaintance type, it is the first mental state occurring in a perceptual causal chain originating in the object to which the thought refers, i.e. it is a sensation. The things presented to us in sensation and of which we have knowledge of acquaintance include ordinary objects in the external world, such as the sun.

Grote contrasted the imagistic thoughts involved in knowledge of acquaintance with things, with the judgements involved in knowledge about things, suggesting that the latter but not the former are contentful mental states. Elsewhere, however, he suggested that every thought capable of constituting knowledge of or about a thing involves a form, idea, or what we might call conceptual propositional content, referring the thought to its object. Whether contentful or not, thoughts constituting knowledge of acquaintance with a thing are relatively indistinct, although this indistinctness does not imply incommunicability. On the other hand, thoughts constituting knowledge about a thing are relatively distinct, as a result of "the application of notice or attention" to the

"confusion or chaos" of sensation (1900, pp. 206–7). Grote did not have an explicit theory of reference, the relation by which a thought is *of* or *about* a specific thing. Nor did he explain how thoughts can be more or less indistinct.

Helmholtz held unequivocally that all thoughts capable of constituting knowledge, whether "knowledge which has to do with Notions" (*Wissen*) or "mere familiarity with phenomena" (*Kennen*), are judgements or, we may say, have conceptual propositional contents. Where Grote saw a difference between distinct and indistinct thoughts, Helmholtz found a difference between precise judgements which are expressible in words and equally precise judgements which, in principle, are not expressible in words, and so are not communicable (Helmholtz, 1962, pp. 269–75).

#### JAMES

James was influenced by Helmholtz and, especially, by Grote (James, 1975, pp. 17–18; 1890, vol. 1, p. 221 n). Adopting the latter's terminology, James agreed with Grote that the distinction between knowledge of acquaintance with things and knowledge about things involves a difference in the degree of vagueness or distinctness of thoughts, though he, too, said little to explain how such differences are possible. At one extreme is knowledge of acquaintance with people and things, and with sensations of colour, flavour, spatial extension, temporal duration, effort and perceptible difference, unaccompanied by knowledge about these things. Such pure knowledge of acquaintance is vague and inexplicit. Movement away from this extreme, by a process of notice and analysis, yields a spectrum of less vague, more explicit thoughts constituting knowledge about things.

However, the distinction was not merely a relative one for James. He was more explicit than Grote in not imputing content to every thought capable of constituting knowledge of or about things. At the extreme where a thought constitutes pure knowledge of

acquaintance with a thing, there is a complete absence of conceptual propositional content in the thought (which is a sensation, feeling, or percept), which also renders the thought incommunicable. James's reasons for positing an absolute discontinuity between pure knowledge of acquaintance and knowledge at all about things seem to have been that any theory adequate to the facts about reference must allow that some reference is not conceptually mediated; that conceptually unmediated reference is necessary if there are to be judgements at all about things and, especially, if there are to be judgements about relations between things; and that any theory faithful to the common person's "sense of life" must allow that some things are directly perceived.

James made a genuine advance over Grote and Helmholtz by analysing the reference relation holding between a thought and the specific thing of or about which it is knowledge. In fact, he gave two different analyses. On both analyses, a thought constituting knowledge about a thing refers to and is knowledge about "a reality, whenever it actually or potentially terminates in" a thought constituting knowledge of acquaintance with that thing (1975, pp. 27–8). The two analyses differ in their treatments of knowledge of acquaintance. On James's first analysis, reference in both sorts of knowledge is mediated by causal chains. A thought constituting pure knowledge of acquaintance with a thing refers to and is knowledge of "whatever reality it directly or indirectly operates on and resembles" (1975, p. 27). The concepts of a thought "operating on" a thing or "terminating in" another thought are causal, but where Grote found chains of efficient causation connecting thought and referent James found teleology and final causes. On James's later analysis, the reference involved in knowledge of acquaintance with a thing is direct. A thought constituting knowledge of acquaintance with a thing either *is* that thing, or has that thing as a constituent, and the thing and the experience of it are identical (1975, ch. 2; 1976, chs 1 and 2).

James further agreed with Grote that pure knowledge of acquaintance with things (i.e.

sensory experience) is epistemically prior to knowledge about things. While the epistemic justification involved in knowledge about things rests on the foundation of sensation, all thoughts about things are fallible and their justification is augmented by their mutual coherence. James was unclear about the precise epistemic status of knowledge of acquaintance. At times, thoughts constituting pure knowledge of acquaintance are said to possess "absolute veritableness" (1890, vol. 1, p. 189) and "the maximal conceivable truth" (1975, p. 87), suggesting that such thoughts are genuinely cognitive and that they provide an infallible epistemic foundation. At other times, such thoughts are said not to bear truth-values, suggesting that "knowledge" of acquaintance is not genuine knowledge at all, but only a non-cognitive necessary condition of genuine knowledge, viz. knowledge about things (1976, p. 102). Russell understood James to hold the latter view.

#### RUSSELL

Russell agreed with Grote and James on the following points. First, knowing things involves experiencing them. Second, knowledge of things by acquaintance is epistemically basic and provides an infallible epistemic foundation for knowledge about things. (Like James, Russell vacillated about the epistemic status of knowledge by acquaintance, and it eventually was replaced at the epistemic foundation by the concept of noticing.) Third, knowledge about things is more articulate and explicit than knowledge by acquaintance with things. Fourth, knowledge about things is causally removed from knowledge of things by acquaintance, by processes of reflection, analysis and inference (1911, 1913, 1959).

But Russell also held that the term "experience" must not be used uncritically in philosophy, on account of the "vague, fluctuating and ambiguous" meaning of the term in its ordinary use. The precise concept found by Russell "in the nucleus of this uncertain patch of meaning" is that of direct occurrent experience of a thing, and he used the term "acquaintance" to express this relation,

though he used that term technically, and not with all of its ordinary meaning (1913, pt 1, ch. 1). Nor did he undertake to give a constitutive analysis of the relation of acquaintance, though he allowed that it may not be unanalysable, and did characterize it as a generic concept. If the use of the term "experience" is restricted to expressing the determinate core of the concept it ordinarily expresses, then we do not experience ordinary objects in the external world, as we commonly think and as Grote and James held we do. In fact, Russell held, one can be acquainted only with one's sense-data (*see* SENSE-DATA) (i.e. particular colours, sounds, etc.), one's occurrent mental states, universals, logical forms, and (perhaps) oneself.

Russell agreed with James that knowledge of things by acquaintance "is essentially simpler than any knowledge of truths, and logically independent of knowledge of truths" (1912, p. 46; 1929, p. 115). The mental states involved when one is acquainted with things do not have propositional contents. Russell's reasons here seem to have been similar to James's. Conceptually unmediated reference to particulars is necessary for understanding any proposition mentioning a particular (e.g. 1918–19, p. 33) and, if scepticism about the external world is to be avoided, some particulars must be directly perceived (1911, p. 119). Russell vacillated about whether or not the absence of propositional content renders knowledge by acquaintance incommunicable.

Russell agreed with James that different accounts should be given of reference as it occurs in knowledge by acquaintance and in knowledge about things, and that in the former case reference is direct. But Russell objected on a number of grounds to James's causal account of the indirect reference involved in knowledge about things. Russell gave a descriptive rather than a causal analysis of that sort of reference: A thought is about a thing when the content of the thought involves a definite description uniquely satisfied by the thing referred to. Indeed, he preferred to speak of knowledge of things by description, rather than of knowledge about things.

Russell advanced beyond Grote and James by explaining how thoughts can be more or less articulate and explicit. If one is acquainted with a complex thing without being aware of or acquainted with its complexity, the knowledge one has by acquaintance with that thing is vague and inexplicit. Reflection and analysis can lead one to distinguish constituent parts of the object of acquaintance and to obtain progressively more distinct, explicit, and complete knowledge about it (1913, 1918–19, 1950, 1959).

#### SUMMARY CONSIDERATIONS

Apparent facts to be explained about the distinction between knowing things and knowing about things are these. Knowledge about things is essentially propositional knowledge, where the mental states involved refer to specific things. This propositional knowledge can be more or less complete, can be justified inferentially and on the basis of experience, and can be communicated. Knowing things, on the other hand, involves experience of things. This experiential knowledge provides an epistemic basis for knowledge about things, and in some sense is difficult or impossible to communicate, perhaps because it is more or less vague.

If one is unconvinced by James's and Russell's reasons for holding that experience of and reference to things are at least sometimes direct, it may seem preferable to join Helmholtz in asserting that knowing things and knowing about things both involve propositional attitudes. To do so would at least allow one the advantages of unified accounts of the nature of knowledge (propositional knowledge would be fundamental) and of the nature of reference (indirect reference would be the only kind). The two kinds of knowledge might yet be importantly different if the mental states involved have different sorts of causal origins in the thinker's cognitive faculties, involve different sorts of propositional attitudes, and differ in other constitutive respects relevant to the relative vagueness and communicability of the mental states.

See also EXPERIENCE; FOUNDATIONALISM; the GIVEN; INTROSPECTION.

#### BIBLIOGRAPHY

- Grote, J.: *Exploratio Philosophica*, part I (Cambridge: Cambridge University Press, 1865); part II, ed. J. B. Mayor (Cambridge: Cambridge University Press, 1900).
- Helmholtz, H. von: "The Recent Progress of the Theory of Vision," trans. P. H. Pye-Smith, in *Popular Scientific Lectures*, ed. M. Kline (New York: Dover, 1962), 93–115.
- James, W.: *The Principles of Psychology*, 2 vols (New York: Henry Holt, 1890).
- James, W.: *The Meaning of Truth: A Sequel to Pragmatism* (Cambridge, MA: Harvard University Press, 1975).
- James, W.: *Essays in Radical Empiricism* (Cambridge, MA: Harvard University Press, 1976).
- Russell, B.: "On Denoting," *Mind* 14 (1905), 479–493; reprinted in *Logic and Knowledge: Essays 1901–1950*, ed. R. C. Marsh (London: Allen and Unwin, 1956).
- Russell, B.: "Knowledge by Acquaintance and Knowledge by Description," *Proceedings of the Aristotelian Society* 11 (1910–11), 108–28. Reprinted as ch. 10 in his *Mysticism and Logic and Other Essays* (New York: Longmans, Green, 1918); adapted as ch. 5 in his *The Problems of Philosophy* (1912).
- Russell, B.: *Theory of Knowledge: The 1913 Manuscript*, ed. E. R. Eames and K. Blackwell (London: Allen and Unwin, 1984). Chs 1–3 of part 1 published as "On the Nature of Acquaintance," *The Monist* 24 (1914), 1–16, 161–87, 435–53; reprinted in *Logic and Knowledge*.
- Russell, B.: *Our Knowledge of the External World*, revised edn (London: Allen and Unwin, 1929).
- Russell, B.: "The Philosophy of Logical Atomism," *The Monist* 28 (1918), 495–527, and 29 (1919), 32–63, 190–222, 345–80; reprinted in *Logic and Knowledge*.
- Russell, B.: *An Inquiry into Meaning and Truth* (London: Allen and Unwin, 1950).
- Russell, B.: *My Philosophical Development* (London: Allen and Unwin, 1959).

DAVID B. MARTENS

**knowledge how, who, why etc.** see DIFFERENT CONSTRUCTIONS IN TERMS OF "KNOWS".



### knowledge-seeking by questioning

Much of traditional epistemology is devoted to the study of the justification or, more generally, the evaluation of the beliefs we have on the basis of some given body of evidence. Recently, belief revision has claimed its place as a further chapter of epistemology. In contrast, relatively little attention has been devoted to the epistemology of knowledge acquisition. One (usually tacit) reason for this neglect is the belief that the most important types of knowledge acquisition, e.g. the discovery of a new scientific theory, are not subject to rules, and hence cannot be studied logically or epistemologically.

In direct opposition to such traditional views, theories of knowledge-seeking by questioning seek to develop explicit logical models of knowledge acquisition. The natural range of application of such theories is, in fact, wider than the title suggests. For consider a knowledge-seeking argument, conceptualized as a sequence of propositions. Take any proposition in the sequence: where does it come from? In some cases, it is a logical consequence of the earlier stages of the argument, and hence does not introduce any new information into the argument; in others, it is not. In such cases some new information enters into the argument. In order to evaluate the argument, we have to know what the source of the new information is and why the arguer resorted to this particular oracle (i.e. source of information). But if we know both of these two things, we might as well think of the new information as an answer to a question addressed by the inquirer to that oracle. Thus a theory of information acquisition by questioning can serve as a framework for argumentation theory in general.

The most explicit form of a theory of knowledge-seeking by questioning is Jaakko Hintikka's interrogative model of inquiry. In spite of the name, it is really a model schema which can be varied in different ways and which leaves open certain parameters needed to specify a model fully.

The interrogative model of inquiry can be thought of as an explicit modern version of the Socratic *elenchus* or method of questioning.

Like the Socratic method, it can be cast into the form of a game, with the important proviso that in the interrogative model the answerer need not be a human interlocutor. In its simplest form, interrogative inquiry is accordingly represented as a two-person game, where the players are the *inquirer* (who need not be an individual investigator but can be a research team or even the scientific community) and a single source of information called *the oracle*. The inquirer is trying to prove a given conclusion *C* starting from a given initial premise *T*. The inquirer has two sorts of moves available to him or her. The inquirer can either draw a logical conclusion from *T* and from the results so far established (a *logical inference move*), or she or he can put a question to the oracle (an *interrogative move*). For the purpose, the inquirer must have previously established (or assumed) the *presupposition* of the question. For instance, if the disjunction ( $S_1 \vee S_2$ ) has been established, the inquirer can use it as a presupposition for the question, "Is  $S_1$  or  $S_2$  true?" If the oracle answers, the inquirer has made progress: now he or she knows not only that the disjunction is true (or can be treated as being true for the purposes of the argument), but which disjunct makes it true.

Formally, the course of an interrogative game can be recorded in a semantical *tableau* in the sense of E. W. Beth. The initial situation involves an initial premise *T* and a fixed ultimate conclusion *C*. At each stage the inquirer can choose to make either a *logical inference move* or an *interrogative move*. The rules for logical inference moves are Beth's *tableau* rules modified so as not to allow any traffic from the right column to the left one. (Subformula property is preserved.) If the presupposition of a question occurs in the left column, the inquirer may make an interrogative move and address the corresponding question to the oracle. If the oracle gives a (conclusive) answer, the answer is added as an extra premise to the left column. The game is played with respect to a given model *M* of the underlying first-order language, and in the simplest version all of the oracle's answers are assumed to be true in *M*, provided only that a true answer is possible in the first place.

The inquirer wins the game iff (if and only if) he or she closes the *tableau*, and the closure rules are *mutatis mutandis* the same as in the deductive case. Iff the inquirer can win the game no matter what the oracle does, C is said to be an interrogative consequence of T in M, in short

(1)  $M: T \vdash C$

For the logical theory of questions relied on here, see Hintikka (1976) and (1984).

Several variations are possible here:

1. The aim of the inquiry may be to answer a question rather than to prove a fixed conclusion. It is easy to see how one can try to answer a propositional question by means of interrogative inquiry: one constructs a separate *tableau* for each (propositional) answer. It is far from obvious how a wh-question can be answered interrogatively; see (e) below.
2. The range of answers the oracle will give is usually assumed to be fixed throughout the inquiry, but it can initially be chosen in different ways.
3. The oracle may be chosen differently in different applications. It can be an actual person, for instance, a witness in a court of law or a patient in a diagnostic interview, but it can also be nature as a target of observations and experiments, computer memory or one's own tacit knowledge. There may also be several different oracles.
4. Instead of assuming that the oracle's answers are always true (whenever they can be), we can assume that they can be false. Then the inquirer must be given the option at each stage to reject (at least temporarily) an earlier answer (or initial premise), together with those steps that depend on it. By the same token, the inquirer is allowed to restore a previously rejected answer or premise (together with its rejected dependents).

What insights are suggested or given rise to by the theory of knowledge-seeking by questioning? The following are some partial answers:

(a) In so far as even a simple form of the interrogative model leads to an interesting theory, one of the main dogmas of recent epistemology and philosophy of science will be refuted, viz. that there cannot be a rational (logical) theory of discovery, only a theory of justification or evaluation.

What can be shown, however, is that there normally are no mechanical (recursive) rules of discovery. Hence this new "logic of discovery" does not contradict the impossibility of subjecting discovery to (mechanical) rules; on the contrary, it enables us to prove that impossibility.

(b) In interrogative games with uncertain answers, the interrogative process itself can be used to evaluate oracles and their answers for reliability, assuming of course some amount of suitable prior information of their credibility. Hence the interrogative model of inquiry can serve as a framework for discussing the self-correcting character of knowledge-seeking methods.

(c) The relation (1) of interrogative provability can be compared with the relation of logical consequence:

(2)  $T \vdash C$

and with the notion of truth in a model:

(3)  $M \models C$

In a certain sense, (1) is between (2) and (3). If no answers are available, (1) reduces to (2). If all questions and answers are available (in a sense that can be made precise), (3) is the case iff

(4)  $M: \emptyset \text{ OS } \models C$

i.e. iff C can be proved interrogatively without any premises ( $\emptyset$  is the empty set).

Thus the interrogative model can serve as a framework for examining the idea of truth as the ideal limit of inquiry.

(d) Heuristically, (1) can be studied by trying to prove for it analogues to well-known metatheorems of deductive logic. This strategy has already proved to be fruitful. For instance, a form of Craig's interpolation

theorem has been shown to hold for (1) (see Hintikka, 1991a).

(e) The interrogative counterpart to the logical notion of definability (on the basis of a given theory) turns out to be a generalization of a notion well known from the methodology of several particular disciplines, viz. the notion of identifiability (see Hintikka, 1991a).

Here the strategy mentioned in (d) serves us especially well. In ordinary logic, the interpolation theorem has as its corollary Beth's theorem which serves as a cornerstone of a theory of definability. Likewise, the interrogative extension of the interpolation theorem has as its corollary an extension of Beth's theorem which can serve as a starting-point of a general logical theory of identifiability. This theory shows among other things how one can try to answer wh-questions by means of interrogative inquiry (see DIFFERENT CONSTRUCTIONS IN TERMS OF 'KNOWS').

(f) The most interesting kind of restriction on the answers the oracle will give is in terms of their maximal allowed quantificational complexity. The different allowed maximal complexity can range from quantifier-free answers (the atomistic case) to A-answers (a prefix of universal quantifiers only) to AE-answers all the way to quantificationally unrestricted answers.

These different restrictions characterize different types of inquiry. For instance, purely observational inquiry is atomistic whereas clinical inquiry can be thought of as being unrestricted in this dimension.

(g) This line of thought can be pushed further. It is generally but mistakenly assumed in effect that empirical scientific inquiry is atomistic. If that were the case, general conclusions, e.g. scientific theories, could not be derived interrogatively without strong (general) antecedent premises. This fact has in effect prompted some of the characteristic models of scientific theorizing, including the inductivist and the hypothetico-deductive one.

Both become redundant if the result of a controlled experiment (establishing the dependence between the controlled variable and the observed one) is construed à la Kant as nature's answer to an inquirer's question. For such an answer is at least of

the AE-complexity, and hence can logically imply general truths. In general, experimental inquiry is characterized by at least the AE-complexity of its answers (see Hintikka, 1988).

This analysis of the logic of experimental inquiry throws light on the actual methodology of scientists, e.g. Isaac Newton's actual methodology and on his methodological views.

(h) Unlike ordinary first-order logic, interrogative inquiry does not allow cut-elimination or the elimination of *modus ponens*. In other words, tautological extra premises ( $S \vee \sim S$ ) can strengthen interrogative inquiry, essentially by serving as presuppositions of yes-or-no questions. The introduction of such a premise can be interpreted as extending the inquirer's range of attention, for the main function of a new tautological premise ( $S \vee \sim S$ ) is to enable the inquirer to ask the yes-or-no question 'Is it the case that S or not?'

(i) The interrogative model facilitates the study of strategies of knowledge-seeking in contrast to the evaluation of beliefs on the basis of already acquired data. One important result in this direction is that strategies of questioning are largely parallel, especially when the oracle's answers are forthcoming and true, to strategies of deduction. (Cf. here Hintikka, 1989a.) This brings out an element of truth in the old conception of logic as the gist of all good reasoning.

In another direction, the question can be raised as to whether many of the rules of inductive and statistical reasoning should be construed as strategic rules rather than rules definitory of the "game" of inquiry.

(j) The interrogative model can be used to distinguish kinds of knowledge from each other, such as active, tacit, virtual, potential, etc., knowledge. It also yields a framework for studying suggestions like Ramsey's according to which knowledge equals true belief obtained by an appropriate method (see Hintikka, 1989b).

(k) When the aim of the game is to answer a question and not to prove a predetermined conclusion, we have to distinguish the *principal question* to be answered through the entire interrogative inquiry from the *operative*

questions whose answers (when available) are used for that purpose. This makes it possible to construe interrogative knowledge-seeking as a multi-level process where the operative question of a higher-level inquiry are principal questions for a series of lower-level interrogative inquiries.

The mistake of trying to ask the principal question rather than an operative one is precisely what the so-called fallacy of *petitio principii* meant (see Robinson, 1971). In general, the interrogative model helps to put the entire traditional theory of fallacies in a new light (see Hintikka 1987).

An early popular exposition of the interrogative approach to inquiry is Hintikka and Hintikka (1982). A textbook treatment of the interrogative model and its applications is Hintikka and Bachman (1990). An alternative treatment of some of the same matters is found in Rescher (1977).

#### BIBLIOGRAPHY

- Hintikka, J.: *The Semantics of Questions and the Questions of Semantics*, Acta Philosophica Fennica 28, no. 4 (1976).
- Hintikka, J.: "Questioning as a Philosophical Method," in *Principles of Philosophical Reasoning*, ed. J. H. Fetzer (Totowa, NJ: Rowman and Allanheld, 1984), 25–43.
- Hintikka, J.: "The Fallacy of Fallacies," *Argumentation* 1 (1987), 211–38.
- Hintikka, J.: "What Is the Logic of Experimental Inquiry?," *Synthese* 74 (1988), 173–90.
- Hintikka, J.: "The Role of Logic in Argumentation," *The Monist* 72 (1989a), 3–24.
- Hintikka, J.: "Knowledge Representation and the Interrogative Model of Inquiry," in *Knowledge and Skepticism*, ed. M. Clay and K. Lehrer (Boulder, Colo.: Westview Press, 1989b), 155–83.
- Hintikka, J.: "Toward a General Theory of Identification," in *Definitions and Definability: Philosophical Perspectives*, ed. J. Fetzer, D. Shatz and G. Schlesinger (Dordrecht: Kluwer, 1991a).
- Hintikka, J.: "Overcoming Metaphysics through Logical Analysis of Language," *Dialectica* (1991b).
- Hintikka, J. and Bachman, J.: *What If . . . ? Toward Excellence in Reasoning* (Mountain View, CA: Mayfield, 1990).
- Hintikka, J. and Hintikka, M. B.: "Sherlock Holmes Confronts Modern Logic: Toward a Theory of Information-seeking through Questioning," in *Argumentation: Approaches to Theory Formation*, ed. E. M. Barth and J. L. Martens (Amsterdam: John Benjamins, 1982), 55–70.
- Rescher, N.: *Dialectics: A Controversy-oriented Approach to the Theory of Knowledge* (Albany, NY: SUNY Press, 1977).
- Robinson, R.: "Begging the question," *Analysis* 31 (1971), 113–17.

JAAKKO HINTIKKA

# L

---

## **Leibniz, Gottfried Wilhelm (1646–1716)**

German philosopher and mathematician. Leibniz's fundamental philosophical interests were in metaphysics, natural theology, logic and the philosophy of science. Leibniz did not produce a work that constitutes his theory of knowledge, nor did he consider in detail the basic epistemological issues that so moved Descartes and Locke. Commenting on Descartes' alleged proof of the existence of material objects in article one of the second part of *The Principles of Philosophy*, Leibniz affirmed that Descartes' argument is weak, and added: "It would be better not to try" (Loamier, p. 391). Note also that there is no section in Leibniz's *New Essays on Human Understanding* – essentially a commentary on Locke's *An Essay Concerning Human Understanding* – corresponding to Chapter 1 of Book I of Locke's *Essay*, in which Locke outlined the epistemological purposes of the *Essay*. None the less, Leibniz made important contributions to epistemology, some of which make plausible the usual listing of Leibniz under the rubric rationalist (see RATIONALISM).

Leibniz's account of truth has the following striking consequence: a proposition is true just in case it is conceptually true, i.e. true in virtue of a relation of containment holding among its concepts. Leibniz was acutely aware that his characterization of truth appears to imply that all truth is necessary truth. (See, for example, Parkinson and Morris, 1973, p. 97.) In his mature work he set out to avoid this unacceptable conclusion by defending the following claims: (1) in a proposition that is necessarily true there is a containment relation among its concepts such that there are definitional analyses of those concepts in virtue of which the original proposition is reducible to an identity, i.e. something of the form  $AB$  is  $A$ , in a finite number of steps; and

(2) in a proposition that is contingently true there is no such analysis, although there are definitional analyses in virtue of which the original proposition converges on an identity. Leibniz took this distinction between finite and infinite analysis to allow him to hold the following doctrines, all of which he thought to be worthy of support:

1. Truth is fundamentally a matter of the containment of the concept of the predicate of a proposition in the concept of its subject.
2. The distinction between necessary truth and contingent truth is absolute, and in no way relative to a corresponding distinction between divine and human sources of knowledge.
3. A proposition is known a priori by a finite mind only if that proposition is a necessary truth. (See, for example, Parkinson and Morris, 1973, p. 98.)

Hence, although Leibniz commenced with an account of truth that one might expect to lead to the conclusion that all knowledge is ultimately a priori knowledge (see A PRIORI KNOWLEDGE in Part I), he set out to avoid that conclusion.

Leibniz's account of our knowledge of contingent truths is remarkably similar to what we would expect to find in an empiricist's epistemology. Leibniz claimed that our knowledge of particular contingent truths has its basis in sense perception. He argued that our knowledge of universal contingent truths can not be based entirely on simple enumerative inductions, but must be supplemented by what he called "the conjectural method a priori", which he described as follows:

The conjectural method a priori proceeds by hypotheses, assuming certain causes, perhaps,



without proof, and showing that the things that happen would follow from those assumptions. A hypothesis of this kind is like the key to a cryptograph, and the simpler it is, and the greater the number of events that can be explained by it, the more probable it is.

(Loamier, 1969, p. 283)

Leibniz's conception of the conjectural method a priori is a precursor of the hypothetico-deductive method. He placed emphasis on the need for a formal theory of probability, in order to formulate an adequate theory of our knowledge of contingent truths.

Leibniz sided with his rationalist colleagues, e.g. Descartes, in maintaining, contrary to the empiricists, that, since thought is an essential property of the mind, there is no time at which a mind exists without a thought, a perception. But Leibniz insisted on a distinction between having a perception and being aware of it. He argued forcefully on both empirical grounds and conceptual grounds that finite minds have numerous perceptions of which they are not aware at the time at which they have them. (See, for example, Remnant and Bennett, 1981, pp. 53–5.)

Leibniz's rationalism in epistemology is most evident in his account of our a priori knowledge, that is, according to (3), our knowledge of necessary truths (see NECESSARY/CONTINGENT). One of Leibniz's persistent criticisms of Locke's empiricism (see LOCKE; EMPIRICISM) is the thesis that Locke's theory of knowledge provides no explanation of how we know of certain propositions that they are not only true, but necessarily true. Leibniz argued that Locke offered no adequate account of how we know propositions to be true whose justification does not depend upon experience; hence, that Locke had no acceptable account of our a priori knowledge. Leibniz's diagnosis of Locke's failing was straightforward: Locke lacked an adequate account of our a priori knowledge because, on Locke's theory, all the material for the justification of beliefs must come from experience, thus overlooking what Leibniz took to be the source of our a priori knowledge, namely, what is innate to the mind. Leibniz summarized his dispute with Locke thus:

Our differences are on matters of some importance. It is a matter of knowing if the soul in itself is entirely empty like a writing tablet on which nothing has as yet been written (tabula rasa) . . . and if everything inscribed there comes solely from the senses and experience, or if the soul contains originally the sources of various concepts and doctrines that external objects merely reveal on occasion. . . .

(Remnant and Bennett, 1981, p. 48)

Leibniz argued for the second alternative, the theory of innate doctrines and concepts. (See, for example *ibid.*, pp. 69–108.)

The thesis that some concepts are innate to the mind is crucial to Leibniz's philosophy. He held that the most basic metaphysical concepts, e.g. the concepts of substance and causation, are innate. Hence, he was unmoved by the inability of empiricists to reconstruct full-blown versions of those concepts from the materials of sense experience.

See also INNATE IDEAS.

#### WRITINGS

Loemker, L. E., ed. and trans.: *Gottfried Wilhelm Leibniz: Philosophical Papers and Letters* (Dordrecht: Reidel, 1969).

Parkinson, G. H. R. and Morris, M., ed. and trans.: *Leibniz: Philosophical Writings* (London: Dent, 1973).

Remnant, P. and Bennett, J., ed. and trans.: *Leibniz: New Essays on Human Understanding* (Cambridge: Cambridge University Press, 1981).

#### BIBLIOGRAPHY

- Broad, C. D.: *Leibniz: An Introduction*, ed. C. Lewy (Cambridge: Cambridge University Press, 1975), ch. 5, "Psychology and Theory of Knowledge."  
 Jolley, N.: *Leibniz and Locke: A Study of the New Essays on Human Understanding* (Oxford: Clarendon Press, 1984).  
 McRae, R.: *Leibniz: Perception, Apperception, and Thought* (Toronto: Toronto University Press, 1976).  
 Rescher, N.: *Leibniz: An Introduction to His Philosophy* (Totowa, NJ: Rowman and Littlefield, 1979), ch. 2, "Human Knowledge."

R. C. SLEIGH, JR

**Lewis, Clarence Irving (1883–1964)**

Lewis was educated at Harvard and taught there from 1920 until his retirement in 1953. His first work was in logic, and his investigation of what he called “strict implication” is still of interest. Lewis also wrote extensively on ethics, but is best known for his work in epistemology.

Lewis defended a strong form of foundationalism (see FOUNDATIONALISM); he held that beliefs about the external world are at most highly probable, and that their probability derives from beliefs about sense-experience which are not just probable but certain. Lewis frequently claims that his position is the only alternative to coherentism (see COHERENTISM), and should be adopted because no beliefs can be probable unless some are certain. Unfortunately, Lewis found these claims so obvious that his support for them is cursory and rests heavily on metaphor. His most explicit argument for the claim that probability requires certainty runs as follows: if the probability of  $h$ , given  $e$ , is  $n$ , and the probability of  $e$  is  $m$  ( $m < 1$ ), we should reassess the probability of  $h$  as  $n \times m$  (Lewis, 1952, p. 172). If this reasoning is correct, and every  $h$  rests on evidence which is itself less than certain, then the probability of  $h$  will indeed be “whittled down to nothing” (ibid.). But this argument rests on dubious assumptions and, like similar arguments in Hume and Russell (see HUME; RUSSELL), leads to the conclusion that the probability of every  $h$  approaches 0, which violates the principle that the probabilities of  $h$  and not  $\neg h$  must sum to 1. The argument also conflicts with Lewis’s views on memory (see MEMORY). The reliability of memory is crucial to inductive inference, but no beliefs of the form “I remember eating breakfast” are certain, and attempts to ground them on the certainty (see CERTAINTY) of judgements of the form “I seem to remember eating breakfast” are doomed to circularity. Lewis meets this problem by assigning memory beliefs an initial probability less than 1 but greater than 0.5, and noting that sets of “congruent” memory beliefs can make a remembered event more probable than any of the memory beliefs taken by itself. This is an ingenious proposal,

but it conflicts with the claim that probability must ultimately derive from certainty.

Another of Lewis’s central claims is that the meaning of every empirical judgment is equivalent to some (very large) set of judgements of the form “If such-and-such sense-experiences occurred, and such-and-such volitions followed, then in all probability such-and-such sense-experiences would ensue”. Lewis calls these conditionals “terminating judgements”, and holds that they can be conclusively verified and thus confirm our judgements about the external world. The claim that empirical judgements entail terminating judgements has been challenged by Chisholm (1948) (see CHISHOLM), and Lewis’s rebuttal (1948) raises doubts about the consistency of his claims that: (1) non-terminating judgements entail terminating judgments, and (2) terminating judgements can be conclusively verified.

Despite these problems, and others, Lewis’s writings provide an exceptionally rich and cogent version of classical empiricism (see EMPIRICISM) as revised by a very fine philosopher well aware of its strengths and weaknesses.

See also the GIVEN; INFINITE REGRESS ARGUMENT; PHENOMENALISM.

**WRITINGS**

- with Langford, C. H.: *Symbolic Logic* (New York: Appleton-Century, 1932).
- Mind and the World Order* (New York: Scribner’s, 1929).
- An Analysis of Knowledge and Valuation* (LaSalle, IL: Open Court, 1946).
- “Professor Chisholm and Empiricism,” *Journal of Philosophy* 45 (1948), 517–24.
- “The Given Element in Empirical Knowledge,” *Philosophical Review* 61 (1952), 168–75.

**BIBLIOGRAPHY**

- Chisholm, R. M.: “The Problem of Empiricism,” *Journal of Philosophy* 45 (1948), 512–7.
- Schilpp, P. A., ed.: *The Philosophy of C. I. Lewis* (LaSalle, IL: Open Court, 1968).

JOHN TROYER

**limits of human knowledge** The issue of the extent and limits of human knowledge is a perplexing one. We can reliably estimate the amount of gold or oil yet to be discovered because we know the earth's extent as a priori given, and can thus establish a proportion between what has been explored and what has not. But we cannot comparably estimate the amount of knowledge yet to be discovered (both because there is no *measure* of what is known and because we have no reliable information regarding new knowledge yet to come). There is thus no way of establishing a proportion between what we know and what we do not.

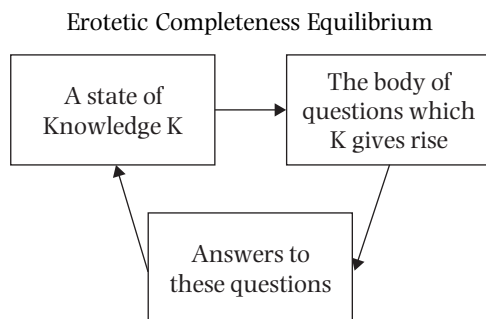
The idea of the question-oriented completeness of the body of knowledge prevailing at a particular historical juncture can be construed in four alternative ways: (1) *Perceived Q-completeness*: every then-asked question has a then-given answer. (2) *Weak Q-completeness*: every then-asked question has a then-available answer. (3) *Strong Q-Completeness: Completeness in Principle*: every then-askable question has a then-available answer. (4) *An Unrealistic Case*: Every then-askable question has a then-given answer. The concept of strong Q-completeness points directly to the idea of an *erotetic equilibrium* between questions and answers. Such an *erotetic equilibrium* subsists when the questions that can be raised on the basis of the concepts and theses of a body of accepted knowledge can be answered with recourse to this same body of knowledge, as per the situation of Figure 1.

But it is sobering to realize that even the strong Q-completeness of a state of knowledge

K does not necessarily betoken its comprehensiveness or sufficiency, but might simply reflect the paucity of the range of questions we are in a position to contemplate on its basis. When the range of our knowledge is sufficiently restricted, then its Q-completeness will merely reflect this restrictedness rather than its intrinsic adequacy. For even as its capacity to resolve our questions counts as a merit of a body of knowledge, so also does its capacity to raise new questions of significance and "depth". Completeness is a hallmark of pseudoscience, which is generally so contrived that the questions that are allowed to be raised are the questions the projected machinery is in a position to resolve.

The idea of incompletionable scientific progress is wholly compatible with the view that *every* question that can be asked at each and every particular state is going to be answered – or dissolved – at some future state: it does not commit one to the idea that there are any unanswerable questions placed altogether beyond the limits of possible resolution. It thus suffices for the prospect of endless cognitive progress to rely on Kant's idea that old problems when solved or dissolved give birth to others whose inherent significance is of no lesser magnitude than that of their predecessors. No recourse to *insolubilia* need be made to maintain the incompleteness of our scientific knowledge.

Someone may ask: "But will it still be appropriate to persist in maintaining the incompleteness of scientific knowledge when science can predict *everything*?" The reply is simply that science will *never* be able to predict literally everything: the very idea of predicting *everything* is simply unworkable since future scientific discoveries always lie beyond the reach of present science. We can only make predictions about matters that lie, at least broadly speaking, within our cognitive horizons; Newton could not have predicted findings in quantum theory any more than he could have predicted the outcome of American presidential elections. The only viable limits to knowability are those which root in knowledge – that is, in a model of nature which entails that certain sorts of things – the outcome of stochastic processes, for example – are in principle unknowable.



**Figure 1**

On the supposition that we are not discussing God, but dealing with finite, imperfect knowers, one must *reject* the contention that all truths are known. But this plausible concession has problematic consequences. Consider the thesis that all truths are known,  $(\forall p)(p \rightarrow Kp)$ . If this is false, that is, if  $\neg(\forall p)(p \rightarrow Kp)$ , then obviously  $(\exists p)(p \& \neg Kp)$  is true. Accordingly, let it now be supposed that there is a particular value of “ $p$ ”, say  $p_0$ , for which this holds, so that:  $p_0 \& \neg Kp_0$ . But we have it in general that  $(\exists p) \neg \Diamond K(p \& \neg Kp)$  or, equivalently,  $\neg(\forall p) \Diamond K(p \& \neg Kp)$ . This emerges from the following reasoning:

- (1)  $(\exists p) \Diamond K(p \& \neg Kp)$  by assumption
- (2)  $(\exists p) \Diamond (Kp \& K(\neg Kp))$  from (1) by the principle:  $K(p \& q) \rightarrow (Kp \& Kq)$
- (3)  $(\exists p) \Diamond (Kp \& \neg Kp)$  from (2) by the principle:  $Kp \rightarrow p$ .

Since (3) is self contradictory its negation is established and that of (1) follows in its wake – as claimed at the start. And now, combining the two preceding findings we have:  $(p_0 \& \neg Kp_0) \& \neg \Diamond K(p_0 \& \neg Kp_0)$ .

And this at once yields  $(\exists p)(p \& \neg \Diamond Kp)$  since the proposition  $p_0 \& \neg Kp_0$  itself represents a specific value of “ $p$ ” for which this thesis obtains. Accordingly,  $(\exists p)(p \& \neg Kp)$  entails  $(\exists p)(p \& \neg \Diamond Kp)$ .

What does this rather striking finding mean? It means that if we want to reject (for finite knowers) the clearly untenable thesis  $p \rightarrow Kp$ , then we must accept not just  $(\exists p)(p \& \neg Kp)$  but even  $(\exists p)(p \& \neg \Diamond Kp)$ . For finite knowers there are not just unknown but even unknowable facts. For if there indeed is an item of fact that is unknown – that is, if  $p_0 \& \neg Kp_0$  – then this particular fact is one that cannot possibly be known in its full specificity, seeing that  $K(p_0 \& \neg Kp_0)$  leads *ad absurdum* by way of the above demonstration. Certain facts about the nature of their own ignorance must necessarily be outside the cognitive range of finite knowers. (However, the incapacity reflected in this unknowability means that, as regards finite knowers,  $(\forall x)(\exists p)(p \& \neg \Diamond Kp)$ ; it does *not* entail  $(\exists p)(\forall x)(p \& \neg \Diamond Kp)$ , that is, establish the existence of *particular* unknowable facts.)

To identify concretely an insoluble question in natural science, we would have to show that a certain scientifically appropriate question  $Q$  is such that its resolution lies “in principle” beyond the reach of science. But how could we possibly establish that a question  $Q$  will continue to be both *raisable* and *unanswerable* in every future state of science, seeing that we cannot now circumscribe the changes that science might undergo in the future? The best we can do here and now is put  $Q$ ’s resolvability beyond the power of any future state of science that looks to be a real possibility *from where we stand*. But the ground we stand on here is always to some extent shifting. If a question belongs to science at all – if it reflects the sort of issue that science might possibly resolve in principle and in theory – then we shall never be in a position to put it beyond the reach of possible future states of science as such. The course of wisdom accordingly lies with the stance that there are no particular scientific questions – and certainly no presently identifiable ones – that science cannot resolve as a matter of principle.

The inherent unpredictability of its internal changes is an ineradicable feature of science. It sets real science apart from the closed structures of pseudoscience, whose deficiencies are reflected precisely in the “elegance” with which *everything* falls much too neatly into place. As regards the question-resolving potential of real science, we can set no a priori restrictions, but have to be flexible. Nobody can say what natural science will and will not be able to do, simply because the science of today is unable to speak decisively for that of tomorrow.

#### BIBLIOGRAPHY

- du Bois-Reymond, E.: *Über die Grenzen des Naturekennens* (Leipzig: 1916).
- Kant, I.: *Critique of Pure Reason*, trans N. Kemp Smith (London: Macmillan, 1964).
- McMullin, E.: “Limits of Scientific Inquiry,” in *Science and the Modern World*, ed. J. C. Steinhardt (New York: Plenum, 1966).
- Rescher, N.: *The Limits of Science* (Berkeley and Los Angeles, CA: University of California Press, 1984).

Wigner, E. P.: "The Limits of Science," *Proceedings of the American Philosophical Society* 94 (1950), 422–7.

NICHOLAS RESCHER

**linguistic understanding** The most influential idea in the theory of meaning in the past hundred years is the thesis that the meaning of an indicative sentence is given by its truth-conditions. On this conception, to understand a sentence is to know its truth-conditions. The conception was first clearly formulated by Frege (*see* FREGE), was developed in a distinctive way by the early Wittgenstein (*see* WITTGENSTEIN), and is a leading idea of Davidson (*see* DAVIDSON). The conception has remained so central that those who offer opposing theories characteristically define their position by reference to it.

The conception of meaning as truth-conditions need not and should not be advanced as being in itself a complete account of meaning. For instance, one who understands a language must have some idea of the range of speech acts conventionally performed by the various types of sentence in the language, and must have some idea of the significance of various kinds of speech act. The claim of the theorist of truth-conditions should rather be targeted on the notion of content: if two indicative sentences differ in what they strictly and literally say, then this difference is fully accounted for by the difference in their truth-conditions. It is this claim, and its attendant problems, which will be the concern of this article.

The meaning of a complex expression is a function of the meaning of its constituents. This is indeed just a statement of what it is for an expression to be semantically complex. It is one of the initial attractions of the conception of meaning as truth-conditions that it permits a smooth and satisfying account of the way in which the meaning of a complex expression is a function of the meaning of its constituents. On the truth-conditional conception, to give the meaning of an expression is to state the contribution it makes to the truth-conditions of sentences in which it occurs. For singular terms – proper names, indexicals, and certain

pronouns – this is done by stating the reference of the term in question. For predicates, it is done either by stating the conditions under which the predicate is true of arbitrary objects, or by stating the conditions under which arbitrary atomic sentences containing it are true. The meaning of a sentence-forming operator is given by stating its contribution to the truth-conditions of a complete sentence, as a function of the semantic values of the sentences on which it operates. For an extremely simple, but nevertheless structured, language, we can state the contributions various expressions make to truth conditions as follows:

- A1: The referent of "London" is London.
- A2: The referent of "Paris" is Paris.
- A3: Any sentence of the form "*a* is beautiful" is true if and only if the referent of *a* is beautiful.
- A4: Any sentence of the form "*a* is larger than *b*" is true if and only if the referent of *a* is larger than the referent of *b*.
- A5: Any sentence of the form "It's not the case that *A*" is true if and only if it is not the case that *A* is true.
- A6: Any sentence of the form "*A* and *B*" is true if and only if *A* is true and *B* is true.

The principles A1–A6 form a simple theory of truth for a fragment of English. In this theory, it is possible to derive these consequences: that "Paris is beautiful" is true if and only if Paris is beautiful (from A2 and A3); that "London is larger than Paris and it is not the case that London is beautiful" is true if and only if London is larger than Paris and it is not the case that London is beautiful (from A1–A5); and in general, for any sentence *A* of this simple language, we can derive something of the form "'*A*' is true if and only if *A*".

The theorist of truth conditions should insist that not every true statement about the reference of an expression is fit to be an axiom in a meaning-giving theory of truth for a language. The axiom "London" refers to the city in which there was a huge fire in 1666 is a true statement about the reference of "London". It is a consequence of a theory which substitutes this axiom for A1 in our simple truth theory that "London is beautiful"



is true if and only if the city in which there was a huge fire in 1666 is beautiful. Since a subject can understand the name "London" without knowing that last-mentioned truth condition, this replacement axiom is not fit to be an axiom in a meaning-specifying truth theory. It is, of course, incumbent on a theorist of meaning as truth conditions to state the constraints on the acceptability of axioms in a way which does not presuppose any prior, non-truth conditional conception of meaning. We will return to this crux later.

Among the many challenges facing the theorist of truth conditions, two are particularly salient and fundamental. First, the theorist has to answer the charge of triviality or vacuity. Second, the theorist must offer an account of what it is for a person's language to be truly describable by a semantic theory containing a given semantic axiom.

We can take the charge of triviality first. In more detail, it would run thus: since the content of a claim that the sentence "Paris is beautiful" is true amounts to no more than the claim that Paris is beautiful, we can trivially describe understanding a sentence, if we wish, as knowing its truth-conditions, but this gives us no substantive account of understanding whatsoever. Something other than grasp of truth conditions must provide the substantive account. The charge rests upon what has been called the redundancy theory of truth, the theory which, somewhat more discriminately, Horwich calls the minimal theory of truth (*see* TRUTH). The minimal theory states that the concept of truth is exhausted by the fact that it conforms to the equivalence principle, the principle that for any proposition *p*, it is true that *p* if and only if *p*. Many different philosophical theories of truth will, with suitable qualifications, accept that equivalence principle. The distinguishing feature of the minimal theory is its claim that the equivalence principle exhausts the notion of truth. It is indeed now widely accepted, both by opponents and supporters of truth conditional theories of meaning, that it is inconsistent to accept both the minimal theory of truth and a truth conditional account of meaning (*see* Davidson, 1990; Dummett, 1959; Horwich, 1990). If the claim that the

sentence "Paris is beautiful" is true is exhausted by its equivalence to the claim that Paris is beautiful, it is circular to try to explain the sentence's meaning in terms of its truth conditions. The minimal theory of truth has been endorsed by Ramsey, Ayer, the later Wittgenstein, Quine, Strawson, Horwich and – confusingly and inconsistently if this article is correct – Frege himself. But is the minimal theory correct?

The minimal theory treats instances of the equivalence principle as definitional of truth for a given sentence. But in fact it seems that each instance of the equivalence principle can itself be explained. The truths from which such an instance as

"London is beautiful" is true if and only if  
London is beautiful

can be explained are precisely A1 and A3 above. This would be a pseudo-explanation if the fact that "London" refers to London consists in part in the fact that "London is beautiful" has the truth-condition it does. But that is very implausible: it is, after all, possible to understand the name "London" without understanding the predicate "is beautiful". The idea that facts about the reference of particular words can be explanatory of facts about the truth conditions of sentences containing them in no way requires any naturalistic or any other kind of reduction of the notion of reference. Nor is the idea incompatible with the plausible point that singular reference can be attributed at all only to something which is capable of combining with other expressions to form complete sentences. That still leaves room for facts about an expression's having the particular reference it does to be partially explanatory of the particular truth condition possessed by a given sentence containing it. The minimal theory thus treats as definitional or stipulative something which is in fact open to explanation. What makes this explanation possible is that there is a general notion of truth which has, among the many links which hold it in place, systematic connections with the semantic values of subsentential expressions.

A second problem with the minimal theory is that it seems impossible to formulate it without at some point relying implicitly on features and principles involving truth which go beyond anything countenanced by the minimal theory. If the minimal theory treats truth as a predicate of anything linguistic, be it utterances, types-in-a-language, or whatever, then the equivalence schema will not cover all cases, but only those in the theorist's own language. Some account has to be given of truth for sentences of other languages. Speaking of the truth of language-independent propositions or thoughts will only postpone, not avoid, this issue, since at some point principles have to be stated associating these language-independent entities with sentences of particular languages. The defender of the minimalist theory is likely to say that if a sentence *S* of a foreign language is best translated by our sentence *p*, then the foreign sentence *S* is true if and only if *p*. Now the best translation of a sentence must preserve the concepts expressed in the sentence. Constraints involving a general notion of truth are pervasive in a plausible philosophical theory of concepts. It is, for example, a condition of adequacy on an individuating account of any concept that there exist what I have called a "Determination Theory" for that account – that is, a specification of how the account contributes to fixing the semantic value of that concept. The notion of a concept's semantic value is the notion of something which makes a certain contribution to the truth conditions of thoughts in which the concept occurs. But this is to presuppose, rather than to elucidate, a general notion of truth.

It is also plausible that there are general constraints on the form of such Determination Theories, constraints which involve truth and which are not derivable from the minimalist's conception. Suppose that concepts are individuated by their possession conditions (*see* CONCEPTS). One such plausible general constraint is then the requirement that when a thinker forms beliefs involving a concept in accordance with its possession condition, a semantic value is assigned to the concept in such a way that the belief is true. Some general principles involving truth

can indeed, as Horwich has emphasized, be derived from the equivalence schema using minimal logical apparatus. Consider, for instance, the principle that "Paris is beautiful and London is beautiful" is true if and only if "Paris is beautiful" is true and "London is beautiful" is true. This follows logically from the three instances of the equivalence principle: "Paris is beautiful and London is beautiful" is true if and only if Paris is beautiful and London is beautiful; "Paris is beautiful" is true if and only if Paris is beautiful; and "London is beautiful" is true if and only if London is beautiful. But no logical manipulations of the equivalence schema will allow the derivation of that general constraint governing possession conditions, truth and the assignment of semantic values. That constraint can of course be regarded as a further elaboration of the idea that truth is one of the aims of judgement.

We now turn to the other question, "What is it for a person's language to be correctly describable by a semantic theory containing a particular axiom, such as the axiom A6 above for conjunction?" This question may be addressed at two depths of generality. At the shallower level, the question may take for granted the person's possession of the concept of conjunction, and be concerned with what has to be true for the axiom to correctly describe his language. At a deeper level, an answer should not duck the issue of what it is to possess the concept. The answers to both questions are of great interest; we will take the lesser level of generality first.

When a person means conjunction by "and", he is not necessarily capable of formulating the axiom A6 explicitly. Even if he can formulate it, his ability to formulate it is not the causal basis of his capacity to hear sentences containing the word "and" as meaning something involving conjunction. Nor is it the causal basis of his capacity to mean something involving conjunction by sentences he utters containing the word "and". Is it then right to regard a truth theory as part of an unconscious psychological computation, and to regard understanding a sentence as involving a particular way of deriving a theorem from a truth theory at

some level of unconscious processing? One problem with this is that it is quite implausible that everyone who speaks exactly the same language has to use exactly the same algorithms for computing the meaning of a sentence. In the past thirteen years, thanks particularly to the work of Davies and Evans, a conception has evolved according to which an axiom like A6 is true of a person's language only if there is a common component in the explanation of his understanding of each sentence containing the word "and", a common component which explains why each such sentence is understood as meaning something involving conjunction (Davies, 1987). This conception can also be elaborated in computational terms: I suggested that for an axiom like A6 to be true of a person's language is for the unconscious mechanisms which produce understanding to draw on the information that a sentence of the form "A and B" is true if and only if A is true and B is true (Peacocke, 1986). Many different algorithms may equally draw on this information. The psychological reality of a semantic theory thus involves, in Marr's (1982) famous classification, something intermediate between his level one, the function computed, and his level two, the algorithm by which it is computed. This conception of the psychological reality of a semantic theory can also be applied to syntactic and phonological theories. Theories in semantics, syntax and phonology are not themselves required to specify the particular algorithms which the language user employs. The identification of the particular computational methods employed is a task for psychology. But semantic, syntactic and phonological theories are answerable to psychological data, and are potentially refutable by them – for these linguistic theories do make commitments to the information drawn upon by mechanisms in the language user.

This answer to the question of what it is for an axiom to be true of a person's language clearly takes for granted the person's possession of the concept expressed by the word treated by the axiom. In the example of the axiom A6, the information drawn upon is that sentences of the form "A and B" are true if and only if A is true and B is true. This

informational content employs, as it has to if it is to be adequate, the concept of conjunction used in stating the meaning of sentences containing "and". So the computational answer we have returned needs further elaboration if we are to address the deeper question, which does not want to take for granted possession of the concepts expressed in the language. It is at this point that the theory of linguistic understanding has to draw upon a theory of concepts. I would argue that it has to draw upon a theory of the conditions for possessing a given concept (*see* CONCEPTS). Let us continue to fix on the example of conjunction. It is plausible that the concept of conjunction is individuated by the following condition for a thinker to possess it:

the concept *and* is that concept *C* to possess which a thinker must meet the following condition: he finds inferences of the following forms compelling, does not find them compelling as a result of any reasoning, and finds them compelling because they are of these forms:

$$\frac{pCq}{p} \quad \frac{pCq}{q} \quad \frac{pq}{pCq}$$

Here *p* and *q* range over complete propositional thoughts, not sentences. When axiom A6 is true of a person's language, there is a global dovetailing between this possession condition for the concept of conjunction and certain of his practices involving the word "and". For the case of conjunction, the dovetailing involves at least this:

If the possession condition for conjunction entails that the thinker who possesses the concept of conjunction must be willing to make certain transitions involving the thought *p* & *q*, and if the thinker's sentence "A" means that *p* and his sentence "B" means that *q*, then: the thinker must be willing to make the corresponding linguistic transitions involving the sentence "A and B".

This is only part of what is involved in the required dovetailing. Given what we have already said about the uniform explanation of the understanding of the various occurrences

of a given word, we should also add that there is a uniform (unconscious, computational) explanation of the language user's willingness to make the corresponding transitions involving the sentence "A and B".

This dovetailing account returns an answer to the deeper question, because neither the possession condition for conjunction, nor the dovetailing condition which builds upon that possession condition, takes for granted the thinker's possession of the concept expressed by "and". The dovetailing account for conjunction is an instance of a more general schema, which can be applied to any concept. The case of conjunction is of course exceptionally simple in several respects. Possession conditions for other concepts will speak not just of inferential transitions, but of certain conditions in which beliefs involving the concept in question are accepted or rejected; and the corresponding dovetailing condition will inherit these features. This dovetailing account has also to be underpinned by a general rationale linking contributions to truth conditions with the particular possession conditions proposed for concepts. It is part of the task of the theory of concepts to supply this in developing Determination Theories for particular concepts.

In some cases, a relatively clear account is possible of how a concept can feature in thoughts which may be true though unverifiable. The possession condition for the quantificational concept *all natural numbers* can in outline run thus: this quantifier is that concept  $Cx \dots x \dots$  to possess which the thinker has to find any inference of the form

$$\frac{CxFx}{Fn}$$

compelling, where  $n$  is a concept of a natural number, and does not have to find anything else essentially containing  $Cx \dots x \dots$  compelling. The straightforward Determination Theory for this possession condition is one on which the truth of such a thought  $CxFx$  ensures that the displayed inference is always truth-preserving. This requires that  $CxFx$  is true only if all natural numbers are F. That all natural numbers are F is a condition which can hold without our being able to establish that

it holds. So an axiom of a truth theory which dovetails with this possession condition for universal quantification over the natural numbers will be a component of a realistic, non-verificationist theory of truth conditions (see REALISM).

Finally, this response to the deeper question allows us to answer two challenges to the conception of meaning as truth-conditions. First, there was the question left hanging earlier, of how the theorist of truth-conditions is to say what makes one axiom of a semantic theory correct rather than another, when the two axioms assign the same semantic values, but do so by means of different concepts. Since the different concepts will have different possession conditions, the dovetailing accounts, at the deeper level, of what it is for each axiom to be correct for a person's language will be different accounts. Second, there is a challenge repeatedly made by the minimalist theorist of truth, to the effect that the theorist of meaning as truth-conditions should give some non-circular account of what it is to understand a sentence, or to be capable of understanding all sentences containing a given constituent. For each expression in a sentence, the corresponding dovetailing account, together with the possession condition, supplies a non-circular account of what it is to understand any sentence containing that expression. The combined accounts for each of the expressions which comprise a given sentence together constitute a non-circular account of what it is to understand the complete sentence. Taken together, they allow the theorist of meaning as truth-conditions fully to meet the challenge.

See also CONCEPTS; OSTENSIVE DEFINITION; TRUTH.

#### BIBLIOGRAPHY

- Davidson, D.: "The Structure and Content of Truth," *Journal of Philosophy* 87 (1990), 279–328.  
 Davies, M.: "Tacit Knowledge and Semantic Theory: Can a Five Per Cent Difference Matter?," *Mind* 96 (1987), 441–62.

- Dummett, M.: "Truth" (1959): reprinted in his *Truth and Other Enigmas* (London: Duckworth, 1978).
- Horwich, P.: *Truth* (Oxford: Blackwell, 1990).
- Marr, D.: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (San Francisco, CA: Freeman, 1982).
- Peacocke, C.: "Explanation in Computational Psychology: Language, Perception and Level 1.5," *Mind and Language* 1 (1986), 101–23.
- Peacocke, C.: *A Study of Concepts* (Cambridge, MA: MIT Press, 1992).
- Wittgenstein, L.: *Philosophical Investigations* (Oxford: Blackwell, 1953), esp. §136.

CHRISTOPHER PEACOCKE

**literature and knowledge** In the traditional approach to the question of the relation between these two multifarious notions, literary works of art are assessed as possible means to knowledge defined along "classical" lines (propositional knowledge or "knowledge that", typically construed as true and justified belief). Three major stances may be identified: (1) condemnations of literature as a source of irrationality for author and audience alike (e.g. Platonic attacks on poetic *mimesis*); (2) defences of literary autonomy based on the idea that knowledge is neither hindered nor advanced by literature because the two move on separate tracks (e.g. Ingarden, 1950, 1973); and (3) various contentions that literary works do in fact contribute to knowledge. A weak version of the latter position holds that some literary works can be used to provide valuable illustrations of knowledge that has been already been formulated outside literature (Simon, 1983, pp. 32–3). Another sort of claim is that theoretically oriented readings of literary works can contribute to the formation of new hypotheses in the human sciences, hypotheses that may then be empirically evaluated through non-literary means (Beardsley, 1958, pp. 429–31; Livingston, 1988, 1991). For example, reading an Icelandic saga's insightful depictions of revenge could help someone improve a theoretical model of the actions and motives typical of blood feuds, in which case the literary work's implicit heuristic value would be revealed. A stronger thesis is that some

literary works convey significant and even systematic knowledge discovered by their authors, an example being Girard's (1965) contention that a number of novelists have expressed genuine insights into the imitative nature of desire.

In the stances just evoked, literature is taken to be one among several effective means to a single kind of epistemic end. It is often contended, however, that "properly literary" features of a work (e.g. its fictional status, narrative form, use of dialogue, pseudonyms, allegory, tropes, etc.) are somehow necessary or essential to literature's contributions to knowledge. A typical claim along these lines is that Nietzsche required the fictional persona of Zarathustra, a quasi-biblical style, parables and mythological imagery to formulate aspects of his thought. Yet if the knowledge in question is propositional and can be paraphrased faithfully, it is not clear why literary devices are necessary or essential to its discovery or formulation. One way of responding to this objection is to contend that literature's cognitive value should not in any case be evaluated in terms of paraphrasable contributions to propositional knowledge. Instead, literature is prized as an indispensable means of realizing various other epistemically valuable results. Literature's distinct and valuable mode of cognition has been identified, for example, as "knowing by vicarious living through", a sort of knowledge by acquaintance that amounts to "knowing what an experience is like" (Walsh, 1969, p. 129). Several variations on this theme have been proposed, but all such theories have the problem of making convincing claims about epistemic states that cannot by definition be restated in the non-literary language in which the theories themselves are couched. As a result, literature's putative unique cognitive value is in danger of becoming an epistemic *je ne sais quoi* towards which a non-literary discourse can only point, this pointing being neither a literary nor a propositional form of knowledge. On the other hand, a clear and detailed argumentative essay that purports to unpack the philosophical content of *Zarathustra* can hardly convince us that the parables and imagery are indispensable.



Another major approach to the topic finds its point of departure in a rejection of the “classical” conception of knowledge in terms of which literary cognition has typically been assessed. In a moderate version, this rejection is not taken to entail a generalized deflation of epistemic values. Instead, the credit formerly reserved to what is deemed a narrow and unrealistic model of knowledge is generously transferred across a broad range of symbolic activities, sheltered beneath the liberal rubric of “understanding”. Once the goals of truth, certainty and rational justification have been abandoned, literature can be granted its place alongside the other arts and sciences as a worthwhile mode of cognition, characterized by non-denotative forms of reference (Goodman and Elgin, 1988). Many literary rejections of “classical” epistemology are more extreme swerves away from reason, rationality and science having been common in literary culture since the beginning of romanticism. Literature is said to free us from the shackles of a pedestrian reason chained to facts, consistency, evidence, justification and other dull strictures. In some versions, literature’s radical “displacements” of so-called knowledge are fancied to involve a privileged access to special domains, including various dimensions of subjectivity – emotions, “qualia”, primary processes, the properties of an essential feminine “difference” manifested only in a particular form of *écriture* – or such metaphysical items as *die Erde*, the body and the “non-oppositional difference”, a post-structuralist grail. Sometimes literature is said to have the special merit of actively manifesting the impossibility of the kind of knowledge to which other discourses, such as philosophy and the sciences, blindly lay claim. The fragments of textuality embody a kind of epistemic virtue, be it a matter of a radical irony, a generalized negativity (Blanchot, 1968), or a sceptical *arrhepsia* or *epoche* (see EPOCHE). In more extreme tendencies, literary theorists proclaim all epistemological demarcations and criteria to be always already overdetermined by social, pragmatic, psychodynamic and/or discursive conditions: literature and knowledge alike are phantasms; they are simply tokens in a struggle for

distinction and power; or they are dissolved within a sea of chaotic textuality, subject to a generalized “illogic” of dispersion and indeterminacy.

The conjunction of the terms “literature” and “knowledge” also raises the issue of the kinds of knowledge produced by literary scholars. The literary disciplines rightly figure among what Miller (1987) calls the “self-questioning” fields (those lacking solid realist foundations, where the latter are a matter of reliable descriptions of relevant domain-specific causal processes). Critical practice is indeed quite diverse, and it is not obvious, to say the least, that its results are typically true, justified, systematic or progressive. In such a context, critical theory’s reflexion on the basic assumptions, methods and aims of literary enquiry is a crucial part of the discipline. The central debate in critical theory has concerned what critics call the question of the “validity” (i.e. veracity) of interpretation, which amounts to asking whether and how a critic can assign the right meaning, or a range of correct meanings, to a text or work. Some designate the author’s intended meanings as the sole target of valid elucidations, while others have labelled this a fallacy. Yet this debate itself rests on assumptions that may be challenged. One such assumption is that critics advance literary knowledge whenever they perform a new correct reading of a literary text or work, and this independent of the correctness of the doctrine thereby elucidated. Showing that literary item *x* meant *y* is deemed a genuine finding, even if the message drawn from the bottle is irrational or is flatly contradicted by the best contemporary theories relevant to the topic. An important assumption supporting such megaphone criticism seems to be that writing and publishing elaborate interpretations is a good, and perhaps even the best manner of appreciating literature, an additional assumption being that engaging in such appreciations is the primary goal of literary research. Two relevant counterclaims are that elaborate interpretive raids do not always contribute to genuine aesthetic experiences, and that scholarly explanations of literary phenomena are not in any case

reducible to the performance of aesthetical appreciations and/or interpretations of works. Literary phenomena also have psycho-social conditions and consequences that would not be adequately captured by even the most successful appreciations and interpretations. Participating in critical appreciations is one kind of task, but describing and explaining the actions constitutive of literary facts and institutions – including the practice of appreciation – is another (Schmidt, 1980–2). Often the two are conflated. We may also speak of a prevalent “poetic fallacy”, the attitude that scholarly research on literary topics should somehow reproduce and exemplify the qualities of the artefacts and practices to which they refer, instead of being guided by norms and aims typical of rational-empirical enquiry.

The prevalent tendency to look at literature as a collection of autonomous works of art requiring elaborate interpretation is relatively recent, and its conceptual foundations are anything but unproblematic (Todorov, 1973, 1982). Critics who remain committed to the tasks of appreciation and interpretation (as opposed to enquiry into the social and psychological history of literary practices and institutions) should pay more attention to the practical conditions that are necessary not only to the production, but to the critical individuation of literary works of art (Currie, 1989; Davies, 1991). It is far from obvious that works can be adequately individuated as objectively identifiable types of token texts or inscriptions, as is often supposed. No semantic function – not even a partial function – maps all types of textual inscriptions onto works of art: some types of inscriptions are not correlated with works at all, and some types of inscriptions may be correlated with more than one work (as in Borges’ *Pierre Menard*). Nor is there even a partial function mapping works onto types of inscriptions: some works may be correlated with more than one type of inscription (e.g. cases where there are different versions of the same work). Particular correlations between text types and works are in practice guided by pragmatic factors involving aspects of the attitudes (beliefs, motives, plans, etc.) of the

agent(s) responsible for the creation of the artefacts in a given context. The latter point entails some non-trivial constraints on literary appreciation in so far as the identification of important types of aesthetic qualities is linked to knowledge of these pragmatic factors. For example, reference to such factors may be needed to determine the assignment of an inscription to a particular genre, as in cases where a text may be read as fantastic fiction, as a hoax, or as earnest supernaturalism. Reference to the author’s context can determine the nature of a work’s stylistic properties: features recognized as highly innovative and unconventional in one context could be viewed as archaic and imitative in another, e.g. should it be learned that the inscription was a twentieth-century work and not a faithful reprinting of an eighteenth-century original. The upshot is that even an aesthetic orientation towards literary phenomena cannot be sundered from action descriptions and the related assumptions about agency. Bringing these assumptions into the foreground reveals the shortcomings of prevalent romantic notions of the “autonomous”, “organic” and “self-organizing” nature of literary works.

Pragmatic factors should also be stressed in a discussion of the cognitive value of literary works and of critics’ interpretations of them. Texts or symbolic artefacts are not the sorts of items that can literally embody or contain the kinds of intentional attitudes that are plausible candidates for the title of knowledge, and this on a wide range of understandings of the latter. If it is dubious that texts and works can know or fail to know anything at all, attention should be shifted to relations between the artefacts and their users, those authors and readers whose relevant actions and attitudes may literally be said to manifest epistemic states and values. Writing was for Proust a genuine form of research, and moreover, one that led to greater self-knowledge, even if the resultant artefacts are not always (and certainly not only) reliable reports of the author’s actual experiences. Moreover, there is no good reason to believe that every reader’s experience of Proust’s writings deserves to be called knowledge or even understanding, although in some hands these works may

very well result in some valuable epistemic results.

See also HISTORICAL KNOWLEDGE; SOCIAL SCIENCES.

# BIBLIOGRAPHY

- Beardsley, M. C.: *Aesthetics: Problems in the Philosophy of Criticism* (New York: Harcourt, Brace, 1958).
- Blanchot, M.: *L'Espace littéraire* (Paris: Gallimard, 1968).
- Currie, G.: *An Ontology of Art* (New York: St Martin's Press, 1989).
- Davies, D.: "Works, Texts and Contexts: Goodman on the Literary Artwork." *Canadian Journal of Philosophy* 21 (1991), 331–45.
- Girard, R.: *Deceit, Desire, and the Novel: Self and Other in Literary Structure*, trans. Y. Freccero (Baltimore, MD: Johns Hopkins University Press, 1965).
- Goodman, N. and Elgin, C. Z.: *Reconceptions in Philosophy and Other Arts and Sciences* (Indianapolis, IN: Hackett, 1988).
- Ingarden, R.: "Des différentes conceptions de vérité dans l'oeuvre d'art," *Revue d'esthétique*, 2 (1950), 162–80.
- Ingarden, R.: *The Cognition of the Literary Work of Art*, trans. R. A. Crowley and K. R. Olson (Evanston, IL: Northwestern University Press, 1973).
- Livingston, P.: *Literary Knowledge: Humanistic Inquiry and the Philosophy of Science* (Ithaca, NY: Cornell University Press, 1988).
- Livingston, P.: *Literature and Rationality: Conceptions of Agency in Theory and Fiction* (Cambridge: Cambridge University Press, 1991).
- Miller, R. W.: *Fact and Method: Explanation, Confirmation and Reality in the Natural and the Social Sciences* (Princeton, NJ: Princeton University Press, 1987).
- Schmidt, S. J.: *Grundriss der empirischen Literaturwissenschaft*, 2 vols (Braunschweig: Vieweg, 1980–2); vol. 1 trans. R. de Beaugrande as *Foundation for the Empirical Study of Literature* (Hamburg: Helmut Buske, 1982).
- Simon, H. A.: *Reason in Human Affairs* (Stanford, CA: Stanford University Press, 1983).
- Todorov, T.: "The Notion of Literature," *New Literary History* 5 (1973), 5–16.
- Todorov, T.: *Theories of the Symbol*, trans. C. Porter (Ithaca, NY: Cornell University Press, 1982).
- Walsh, D.: *Literature and Knowledge* (Middletown, CT: Wesleyan University Press, 1969).

PAISLEY LIVINGSTON

**Locke, John (1632–1704)** English philosopher. Aside from his *Two Treatises of Government* (1690) and *The Reasonableness of Christianity* (1695), Locke is known for his epistemological masterpiece, *An Essay concerning Human Understanding* (1690). Implicitly believing that we have been put here by God in this world with some expectation of an after-life in another, Locke's aim in this work is to discover what kind of thing God has fitted us to know, and so how we should direct and use our intellect and understandings. "My purpose", he tells us, is "to enquire into the original, certainty, and extent of human knowledge; together, with the grounds and degrees of belief, opinion, and assent" (1.1.2).

In opposition to some at the time Locke vigorously denies, throughout Book 1 of the *Essay*, that any of our knowledge – whether theoretical or ethical – is innate and there in our minds from the very outset (see INNATE IDEAS). The mind at birth, he says, is like "white paper" (2.1.2) (see TABULA RASA); and as a consequence he is often described as an empiricist (see EMPIRICISM). He certainly thinks that all our ideas (see IDEA) are derived from experience; but they are only what he calls "the materials of knowledge" (2.1.2). Unlike ideas, knowledge itself is *not* "made out to us by our senses" (Locke, 1671, p. 157). Somewhat after the style of Descartes (see DESCARTES), knowledge is very much a product of reason and understanding for Locke. In line with what a rationalist might say (see RATIONALISM), without reason all we have is belief, not knowledge. "Reason must be our last judge and guide in everything" (4.19.14).

In Book 2 Locke substantiates the claim that all our ideas, all the materials of knowledge, come from experience – in either of its two forms, of sensory perception of the material world and of reflection on the operations of our own minds. His argument is facilitated by a distinction between simple and complex ideas – the former being unanalysable and indefinable, the latter being mentally constructible out of simples. Complex ideas are of various sorts – substances (e.g. gold, lead, horses) which represent things in the material world, modes (e.g. triangle, gratitude)

which are “dependences on, or affections of substances” (2.12.4), and relations (e.g. parent of, whiter than). In the course of showing how any idea has its source in experience Locke gives us analyses of various philosophically important notions: perception, solidity, memory, space, time, number, infinity, volition, substance, cause and effect, personal identity. He also makes a distinction between primary and secondary qualities (see PRIMARY AND SECONDARY QUALITIES). There is considerable implicit criticism in these chapters of various views of Descartes (see DESCARTES): the identification of extension as the whole essence of material substance, and the denial of the vacuum; the belief that thought and personal identity necessarily require an immaterial substance; the view that the mind is always thinking. Nevertheless it was his reading of Descartes which first led Locke away from the verbal trivialities and mystifications (as they seemed) of the then-prevailing Aristotelian scholasticism to which he had been exposed as a student in Oxford. It was Descartes too from whom Locke takes the term “idea”, whose frequent use (for which he begs pardon (1.1.8)) is such a feature of the *Essay*.

Many of the “vague and insignificant forms of speech, and abuse of language” of the Scholastics to which Locke refers in his preface come in for criticism in Book 3, “On Words”. Here he rejects the view that the general and classificatory words which form such a large part of our language stand for “real essences” or “substantial forms” which, by being embodied in things, make them to be of one sort of another. According to this view, the world, prior to any intellectual activity of ours, is already divided up into various sorts of thing. In opposition to this, Locke argues that “all things, that exist . . . [are] particulars” (3.3.1), that classification is a matter of human interests and convenience, and that general words stand for “nominal essences”, mental abstract ideas which we ourselves construct. Generality and universality, he says, “belong not to the real existence of things; but are the inventions and creatures of the understanding, made by it for its own use” (3.3.11).

In Book 4, knowledge is defined as the intellectual “perception” of connections, or agreements and disagreements, between ideas. Such connections may be immediate and intuitively perceived, or mediated by other ideas and so a matter of demonstration. Where we can perceive such connections we have certain and universal knowledge; where we cannot, we lack “knowledge”, and, at best, have “belief” or “opinion”. This account fits well with our knowledge in a priori subjects such as mathematics and geometry (see A PRIORI KNOWLEDGE in Part I). These, because of their systematic development, were standard seventeenth-century examples of “science”. But, like others of the “new philosophers” of the time, Locke did not accept the details of the Scholastic account of *scientia*, according to which scientific knowledge involved a rigidly defined structure, and is arranged and developed according to strict canons of syllogistic reasoning which has as its premisses abstract maxims and definitions of the real essences of things (see AQUINAS). Science for Locke is, as in mathematics and geometry, a matter of deduction, but he places no particular value on syllogistic techniques, and the associated apparatus of Scholastic logic.

Yet though he rejects much of what the Scholastics say about real essences and of their associated account of science Locke does not reject completely any idea of real essence. He retains the general idea that there is something about a mode, such as a triangle, or a substance, such as gold, which accounts for and explains the characteristic properties of those things. Indeed, it is because we know the real essences of geometrical figures that we can discern necessary connections in geometry – as between our complex idea of the real essence of a triangle (“three lines, including a space”) (2.31.6) and its various properties.

On the other hand we can, in general, discern no necessary connections between our ideas of substances and their properties. “Beliefs” about the properties of, for example, gold depend on observation and experiment, and not on the a priori contemplation of our own abstract ideas. The reason for this lack of

knowledge, with its consequence that “natural philosophy”, the study of the properties of the things of the material world, “is not capable of being made a science” (4.12.10) is that we are ignorant of the relevant substantial real essences. Locke’s conception of the real essences of substances differs from that of the scholastics, according to which they are “real definitions”; rather, he sees them in terms of the classical Greek atomic theory of matter of Leucippus, Democritus and Epicurus which was revived in the seventeenth century. If we knew the real essences of substances then chemistry would be as *a priori* as mathematics. In the absence of detailed ideas of these things, however, our beliefs about the properties of substances are to be based on careful and systematic observation and experiments. In advocating this “historical, plain method” (1.1.2) for natural philosophy and in his belief in the corpuscular theory of matter Locke aligns with natural philosophers of the time such as Robert Boyle, people to whom in his preface to the *Essay*, Locke refers as “master-builders” of “the commonwealth of learning”. He, like them, was a Fellow of the anti-scholastic Royal Society of London for the Improving of Natural Knowledge.

In response to sceptical arguments about the existence of the external world (see PROBLEM OF THE EXTERNAL WORLD) and the possibility of our living a perpetual dream, Locke reacts in his practical way. “No body can, in earnest, be so sceptical, as to be uncertain of the existence of those things which he sees and feels” (4.11.3) (see SCEPTICISM, MODERN). His account of this “sensitive knowledge” (4.2.14) does not fit his definition of knowledge as the perception of connections between ideas. However, though its certainty is not so great as that from intuition (see INTUITION IN EPISTEMOLOGY) and demonstration it still, he says, deserves the name of knowledge. Traditionally, Locke’s theory of perception has been taken to be a representational realism (see REPRESENTATIVE REALISM) – according to which, since the mind “perceives nothing but its own ideas” (4.4.3), our perception of the external world is indirect. In recent years, however, it has been argued that he is a direct realist (see DIRECT REALISM).

Geometry is not the only case where real knowledge can be had (see GEOMETRY). Morality is another, since there too the relevant ideas are modes whose real essences we do or could know. According to Locke, moral principles express the will of God, and so a science of morality would be based partly on a proof of the existence of God and partly on the idea that as His dependants we should obey his will, as children should obey that of their parents. Despite the urgings of friends Locke said he had neither the time nor the ability to produce a systematic moral science himself. Nevertheless, some individual moral truths have been worked out by reason, though many people, “perplexed in the necessary affairs of life” (1.3.25), have no time for this. It is fortunate, then, that such people can get morality from the revelations of the Bible; though of course what they acquire in that way is not moral knowledge but only belief.

There are other ways in which the beliefs based on the Bible are inferior and secondary to reasoned knowledge. Revelation is actually answerable to reason in that reason is needed to decide what is a genuine revelation. If something is the message of God then undoubtedly it is true; but is it the message of God? Furthermore, any religious or moral knowledge that is necessary for salvation can be acquired by our natural faculties alone and without the assistance of revelation. This is a claim which gives Locke a place in the history of the development of a deistic natural religion.

The upshot of his enquiry is that there are things that we do not know, things about which we can only form beliefs and things about which we must remain in ignorance. Natural philosophy as a whole will never be a science; and Locke is more agnostic than Descartes about the possibility of ever knowing how the mind acts on the body or the body on the mind, or of whether thought necessarily needs an immaterial soul. But some things we do know – and our beliefs are often not foundationless.

Locke’s general picture is that what we do know, and the things we justifiably believe, answer to our true needs and real interests. We



are not in ignorance of our duties and obligations to each other and to God; and we have beliefs enough for everyday practicalities. "Men have reason to be well satisfied with what God hath thought fit for them, since he has given them . . . whatsoever is necessary for the conveniences of life, and information of virtue; and has put within the reach of their discovery the comfortable provision for this life and the way that leads to a better" (1.1.5). Not only should we thank God for what we have, but we should also be less greedy and "more cautious in meddling with things exceeding our comprehension" (1.1.4).

Much of what is in the Essay, such as its stress on the descriptive "historical, plain method" (11.2) in natural philosophy, its advocacy of the corpuscular theory of matter, its attacks on the Scholastics, its stress on working things out for oneself and its dislike of authority, its giving reason a central place in religion, and its criticisms of various teachings of Descartes on the twin substances of thought and extension, belong to his time as well as to him. But though much of what he says can thus be found in lesser contemporaries, he brings together much of the thought of the seventeenth century in a fresh and coherent way.

See also ANALYTICITY; INTUITION AND DEDUCTION; REPRESENTATIVE REALISM.

#### WRITINGS

*Draft A of Locke's Essay Concerning Human Understanding* (1671), transcribed by P. H. Nidditch (Sheffield: Department of Philosophy, University of Sheffield, 1980).  
*An Essay Concerning Human Understanding* (1690), ed. P. H. Nidditch (Oxford: Clarendon Press, 1975).

#### BIBLIOGRAPHY

Aaron, R. I.: *John Locke*, 3rd edn (Oxford: Clarendon Press, 1971).  
 Alexander, P.: *Ideas, Qualities and Corpuscles: Locke and Boyle on the External World* (Cambridge: Cambridge University Press, 1985).  
 Gibson, J.: *Locke's Theory of Knowledge and Its*

*Historical Relations* (Cambridge: Cambridge University Press, 1960).  
 Hall, R. and Woolhouse, R.: *Eighty Years of Locke Scholarship: A Bibliographical Guide* (Edinburgh: Edinburgh University Press, 1983).  
 Schapiro, B. J.: *Probability and Certainty in Seventeenth-century England* (Princeton, NJ: Princeton University Press, 1983).  
 Tipton, I., ed.: *Locke and Human Understanding* (Oxford: Oxford University Press, 1977).  
 Woolhouse, R. S.: *Locke* (Brighton: Harvester Press, 1983).  
 Yolton, J. W.: *Locke and the Compass of the Human Understanding* (Cambridge: Cambridge University Press, 1970).

R. S. WOOLHOUSE

**logical construction** The phrase was coined by Russell: "Wherever possible, logical constructions are to be substituted for inferred entities" (Russell, 1914, p. 115) (see RUSSELL). Instead of thinking of our world as containing "metaphysical monsters" like material, mind-independent continuants, we should think of these as logical constructions out of sense-data (see SENSE-DATA). For Russell, thinking of material things as logical constructions amounted to a revision, rather than merely a description, of ordinary views (which incorporate the "savage superstitions of cannibals" (Russell, 1925, p. 143)). Russell's constructions were sets. For Carnap (see CARNAP), who also used the notion, constructions were sums (to be consistent with his nominalism). A recent exponent of logical construction is Quine, who holds up as a model the definition ("construction") of an ordered pair  $\langle x, y \rangle$  as the set  $\{x, \{x, y\}\}$  (see his 1960, §53). An important feature of constructions is that, although one can "paraphrase" everything one wanted to say about the original entity in terms of the construction, substituting an expression for the one by an expression for the other does not always preserve intuitive truth, let alone meaning. Thus "each of  $\langle x, y \rangle$ 's members is an individual" is intuitively true, whereas "each of  $\{x, \{x, y\}\}$ 's members is an individual" is false. To accept the construction is to forgo the intuition.

See also PHENOMENALISM.

## BIBLIOGRAPHY

- Carnap, R.: *The Logical Structure of the World: Pseudo-problems in Philosophy*, trans. R. A. George (London: Routledge and Kegan Paul, 1967).
- Russell, B.: "The Relation of Sense-data to Physics," *Scientia* 4 (1914); reprinted in his *Mysticism and Logic* (New York: Longmans, 1918; London: Allen and Unwin, 1963).
- Russell, B.: "Mind and Matter," *Nation and Athenaeum* 37 (1925); reprinted in his *Portraits from Memory* (London: Readers Union, 1958).
- Quine, W. V.: *Word and Object* (Cambridge, MA: MIT Press, 1960).

R. M. SAINSBURY

**logical empiricism** *see* LOGICAL POSITIVISM; REICHENBACH.

**logical positivism** A loosely defined movement or set of ideas (sometimes also called "logical empiricism") which coalesced in Vienna in the 1920s and early 1930s and found many followers and sympathizers elsewhere and at other times. It was a dominant force in philosophy, at least in English-speaking countries, into the 1960s, and its influence, if not specific theses, remains present in the views and attitudes of many philosophers. It was "positivism" in its adherence to the doctrine that science is the only form of knowledge and that there is nothing in the universe beyond what can in principle be scientifically known. It was "logical" in its dependence on developments in logic and mathematics in the early years of this century which were taken to reveal how a priori knowledge of necessary truths is compatible with a thorough-going empiricism.

The exclusiveness of a scientific worldview was to be secured by showing that everything beyond the reach of science is strictly or "cognitively" meaningless, in the sense of being incapable of truth or falsity, and so not a possible object of cognition. This required a criterion of meaningfulness, and it was found in the idea of empirical verification. A sentence is said to be cognitively meaningful if and only if it can be verified or falsified in experience. This is not meant to

require that the sentence be conclusively verified or falsified, since universal scientific laws or hypotheses (which are supposed to pass the test) are not logically deducible from any amount of actually observed evidence. The criterion is accordingly to be understood to require only verifiability or falsifiability, in the sense of empirical evidence which would count either for or against the truth of the sentence in question, without having to logically imply it. Verification or confirmation is not necessarily something that can be carried out by the person who entertains the sentence or hypothesis in question, or even by anyone at all at the stage of intellectual and technological development achieved at the time it is entertained. A sentence is cognitively meaningful if and only if it is in principle empirically verifiable or falsifiable.

Anything which does not fulfil this criterion is declared literally meaningless. There is no significant "cognitive" question as to its truth or falsity; it is not an appropriate object of enquiry. Moral and aesthetic and other "evaluative" sentences are held to be neither confirmable nor disconfirmable on empirical grounds, and so are cognitively meaningless. They are at best expressions of feeling or preference which are neither true nor false. Whatever is cognitively meaningful and therefore factual is value-free. The positivists claimed that many of the sentences of traditional philosophy, especially those in what they called "metaphysics", also lack cognitive meaning and say nothing that could be true or false. But they did not spend much time trying to show this in detail about the philosophy of the past. They were more concerned with developing a theory of meaning and of knowledge adequate to the understanding and perhaps even the improvement of science.

The logical positivist conception of knowledge in its original and purest form sees human knowledge as a complex intellectual structure employed for the successful anticipation of future experience. It requires, on the one hand, a linguistic or conceptual framework in which to express what is to be categorized and predicted and, on the other, a factual element which provides that abstract form with

content. This comes, ultimately, from sense experience. No matter of fact that anyone can understand or intelligibly think to be so could go beyond the possibility of human experience, and the only reasons anyone could ever have for believing anything must come, ultimately, from actual experience.

The general project of the positivistic theory of knowledge is to exhibit the structure, content, and basis of human knowledge in accordance with these empiricist principles. Since science is regarded as the repository of all genuine human knowledge, this becomes the task of exhibiting the structure, or as it was called, the "logic" of science. The theory of knowledge thus becomes the philosophy of science. It has three major tasks: (1) to analyse the meanings of the statements of science exclusively in terms of observations or experiences in principle available to human beings; (2) to show how certain observations or experiences serve to confirm a given statement in the sense of making it more warranted or reasonable; (3) to show how non-empirical or a priori knowledge of the necessary truths of logic and mathematics is possible even though every matter of fact which can be intelligibly thought or known is empirically verifiable or falsifiable.

1. The slogan "the meaning of a statement is its method of verification" expresses the empirical *verification theory of meaning*. It is more than the general criterion of meaningfulness according to which a sentence is cognitively meaningful if and only if it is empirically verifiable. It says in addition what the meaning of each sentence is: it is all those observations which would confirm or disconfirm the sentence. Sentences which would be verified or falsified by all the same observations are empirically equivalent or have the same meaning.

A sentence recording the result of a single observation is an observation or "protocol" sentence. It can be conclusively verified or falsified on a single occasion. Every other meaningful statement is a "hypothesis" which implies an indefinitely large number of observation sentences which together exhaust its meaning, but at no time will all of them have been verified or falsified. To give an "analysis"

of the statements of science is to show how the content of each scientific statement can be reduced in this way to nothing more than a complex combination of directly verifiable "protocol" sentences (*see VERIFICATIONISM*).

2. The observations recorded in particular "protocol" sentences are said to confirm those "hypotheses" of which they are instances. The task of confirmation theory is therefore to define the notion of a confirming instance of a hypothesis and to show how the occurrence of more and more such instances adds credibility or warrant to the hypothesis in question. A complete answer would involve a solution to the problem of induction (*see PROBLEMS OF INDUCTION*); to explain how any past or present experience makes it reasonable to believe in something that has not yet been experienced (*see HUME*).

3. Logical and mathematical propositions, and other necessary truths, do not predict the course of future sense experience. They cannot be empirically confirmed or disconfirmed. But they are essential to science, and so must be accounted for. They are one and all "analytic" in something like Kant's sense: true solely in virtue of the meanings of their constituent terms (*see KANT*). They serve only to make explicit the contents of and the logical relations among the terms or concepts which make up the conceptual framework through which we interpret and predict experience. Our knowledge of such truths is simply knowledge of what is and what is not contained in the concepts we use (*see ANALYTICITY*).

Experience can perhaps show that a given concept has no instances, or that it is not a useful concept for us to employ. But that would not show that what we understand to be included in that concept is not really included in it, or that it is not the concept we take it to be. Our knowledge of the constituents of and the relations among our concepts is therefore not dependent on experience; it is a priori. It is knowledge of what holds necessarily, and all necessary truths are "analytic". There is no synthetic a priori knowledge (*see A PRIORI KNOWLEDGE* in Part I).

The anti-metaphysical empiricism (*see EMPIRICISM*) of logical positivism requires that there be no access to any facts beyond sense

experience. The appeal to analyticity succeeds in accounting for knowledge of necessary truths only if analytic truths state no facts, and our knowledge of them does not require non-sensory awareness of matters of fact. The reduction of all the concepts of arithmetic, for example, to those of logic alone, as was taken to have been achieved in Whitehead and Russell's *Principia Mathematica*, showed that the truths of arithmetic were derivable from nothing more than definitions of their constituent terms and general logical laws. Frege would have called them "analytic" for that reason alone. But for a complete account positivism would have also to show that general logical laws state no facts.

Under the influence of their reading of Wittgenstein's *Tractatus Logico-Philosophicus* (see WITTGENSTEIN), the positivists regarded all necessary and therefore all analytic truths as "tautologies". They do not state relations holding independently of us within an objective domain of concepts. Their truth is "purely formal"; they are completely "empty" and "devoid of factual content". They are to be understood as made true solely by our decisions to think and speak in one way rather than another, as somehow true "by convention". A priori knowledge of them is in this way held to be compatible with there being no non-sensory access to a world of things beyond sense experience.

The full criterion of meaningfulness therefore says that a sentence is cognitively meaningful if and only if either it is analytic or it is in principle empirically verifiable or falsifiable.

See also A PRIORI KNOWLEDGE in Part I; ANALYTICITY; CARNAP; EMPIRICISM; NEURATH; PHILOSOPHICAL KNOWLEDGE; SCHLICK; VIENNA CIRCLE.

## BIBLIOGRAPHY

- Ayer, A. J.: *Language, Truth and Logic* (London: Gollancz, 1946).  
 Ayer, A. J., ed.: *Logical Positivism* (Glencoe, IL: Free Press, 1959).  
 Carnap, R.: *The Logical Structure of the World:*

- Pseudoproblems in Philosophy*, trans. R. George (London: Routledge and Kegan Paul, 1967).  
 Carnap, R.: *The Logical Syntax of Language*, trans. A. Smeaton (Paterson, NJ: Littlefield, Adams, 1959).  
 Carnap, R.: *Logical Foundations of Probability* (Chicago, IL: University of Chicago Press, 1950).  
 Feigl, H. and Sellars, W., eds: *Readings in Philosophical Analysis* (New York: Appleton-Century-Crofts, 1949).  
 Hempel, C.: *Aspects of Scientific Explanation* (New York: The Free Press, 1965).  
 Wittgenstein, L.: *Tractatus Logico-Philosophicus*, trans. C. K. Ogden (London: Routledge and Kegan Paul, 1922).

BARRY STROUD

**logicism** First formulated by Gottlob Frege and Bertrand Russell (see FREGE; RUSSELL), logicism is the view that number theory is reducible in its entirety to pure, deductive logic. It involves the claim that arithmetical concepts are logical concepts; that arithmetical truths are truths of logic; and that proofs in arithmetic are valid to the extent that they are logically, deductively valid. If this view could be substantiated it would show that arithmetical knowledge has the same origin, content, and justification as our knowledge of logical truths.

See also A PRIORI KNOWLEDGE in Part I; INTUITION AND DEDUCTION; MATHEMATICAL KNOWLEDGE.

## BIBLIOGRAPHY

- Frege, G.: *The Foundations of Arithmetic*, trans. J. L. Austin (Oxford: Blackwell, 1974).  
 Frege, G.: *The Basic Laws of Arithmetic*, trans. M. Furth (Berkeley, CA: University of California Press, 1964).  
 Russell, B.: *The Principles of Mathematics* (London: Allen and Unwin, 1903).  
 Whitehead, A. N. and Russell, B.: *Principia Mathematica* (Cambridge: Cambridge University Press, 1910).

DAVID BELL

**lottery paradox** Due originally to Henry Kyburg (Kyburg, 1961 and 1970), the lottery

paradox is a demonstration that certain appealing principles about rational acceptance (or perhaps justified belief or knowledge) turn out to be logically inconsistent. The key assumptions are:

- A. If it is very likely that a conclusion is true, then it is rational to accept that conclusion.
- B. If it is rational to accept that  $p$  is the case and it is rational to accept that  $q$  is the case, then it is rational to accept that both  $p$  and  $q$  are the case.
- C. It is never rational to accept propositions you realize to be logically inconsistent.

How do these assumptions lead to paradox? Suppose there is a fair lottery with a thousand tickets. It is determined that one ticket will be drawn as the winner; however, the probability that each particular ticket will lose is very high (i.e., .999). Let us imagine that the tickets are numbered from 1 to 1,000. Since it is overwhelmingly probable that Ticket 1 will lose, by Assumption (A) you should conclude that it will, indeed, lose. Now, the same thing can be said about each of the other tickets in turn; hence, you should accept that Ticket 1 will lose, and you should accept that Ticket 2 will lose, and so on. But, according to Assumption (B), this means that you should accept that Ticket 1 will lose and that Ticket 2 will lose . . . *and* that Ticket 1,000 will lose. That is, it is rational for you to believe that *all* the tickets will lose, which is to say that no ticket will win. The trouble is that you have already accepted that some ticket or other will be chosen as the winner; and it seems, then, that it is now rational for you to accept something logically inconsistent: namely, that no ticket will win and that some ticket will win. However, such an outcome conflicts with Assumption (C). It is worth noting that, in arriving at this point, nothing essential turned on the choice of .999 as a sufficient degree of likelihood to warrant rational belief. The required odds can be set as high as we like; so long as they are less than one, we can generate the paradox just by increasing the number of tickets in the imagined lottery.

This general result has elicited a wide range of responses. An extreme position has been taken by Richard Jeffrey (Jeffrey, 1970). Jeffrey holds that it is misguided from the start to try to frame a notion of rational acceptance in probabilistic terms. For Jeffrey, probabilities (understood as degrees of confidence) are, in effect, all there is to rational belief formation. We do not simply believe or reject a proposition; we just become inclined to risk more or less on its truth (see *PROBABILITY, THEORIES OF*).

Other philosophers, who remain committed to providing an account of rational acceptance, have taken the Lottery Paradox to show that one or more of the assumptions (A)–(C) is faulty or incorrect. Kyburg himself has urged that we deny Assumption (B) (Kyburg, 1970). That is, you may well rationally accept separately that Ticket 1 will lose and that Ticket 2 will lose, and so on; however, you may not accept the conjunction that Ticket 1 will lose *and* Ticket 2 will lose *and* . . . (at least for sufficiently long conjunctions). This restriction has a plausible motivation in the facts about probability. For it is a theorem that the probability of a conjunction is less than or equal to that of its conjuncts. And, if high probability is the basis for rational acceptance, we should expect that a conjunction, when significantly less probable than its conjuncts, will be unacceptable even though the conjuncts themselves are acceptable.

Still, one can marshal considerations in favour of Assumption (B). Mark Kaplan (1981) has suggested that, without it, we could not account for the recognized force of arguments by *reductio ad absurdum*. That a contradiction can be derived from the conjunction of things I believe gives me reason to change one or more of those beliefs, and I could not evade the critical force of such a demonstration simply by forswearing acceptance of the conjunction (while retaining belief in the conjuncts). Another difficulty is that to distinguish between believing a conjunction and believing its conjuncts seems to commit one to the view that the objects of belief have a particular syntactic structure or other (here, that of conjunction). Such a view is controversial, and it would be desirable to find a solution to



the Lottery Paradox that did not depend upon a particular theory about the character of intentional mental states (Stalnaker, 1984).

A case has also been made for abandoning Assumption (C). To be sure, an out-and-out contradiction is not rationally acceptable. But it has been claimed that the conflict between accepting that each ticket individually will lose and accepting that some ticket will win is of a different, less objectionable sort. This tension is seen as an unavoidable aspect of cognitive life – which figures in the preface paradox (*see* PREFACE PARADOX) as well (Klein, 1985).

The alternative to denying Assumptions (B) and (C) is to reject (A). There are several possibilities here. One might draw the lesson from the Lottery Paradox that rational acceptance (or at least the kind of rational belief necessary for knowledge) requires absolute certainty rather than mere likelihood. That is, you should not believe anything for which there is a chance, however small, that you might be wrong. In particular, given that there is some (perhaps tiny) probability that “Ticket 1 (or 2, or 3, etc.) will lose” is false, you should not accept any of these propositions (*see* Heidelberger, 1963; Dretske, 1981). The threat of paradox is obviated, but still, it would seem, at a very high cost. Almost every belief we hold is such that there is some chance of being wrong about it; we can be perfectly certain about very little. Thus, if the Lottery Paradox shows that rational acceptance requires absolute certainty, then, by extension, the Paradox seems to demonstrate that we may rationally accept almost nothing.

A more moderate response would be to maintain that high probability may license rational acceptance given certain additional constraints or qualifications. Thus, it has been proposed (Lehrer, 1980; Pollock, 1983; and others) that we may rationally accept a likely conclusion so long as it is part of some overall coherent scheme of beliefs; in the Lottery Paradox, accepting claims that individual tickets will lose would violate this condition, and so is barred. A different approach, due to Isaac Levi, relativizes the notion of acceptance to a context of inquiry. Levi holds that we may accept that Ticket 1

will lose, if the question asked concerns the outcome with respect to Ticket 1 in particular; however, whether *all* tickets (Ticket 1 and Ticket 2, etc.) will lose is another matter, and must be dealt with independently of the former (Levi, 1967).

#### RELATED ISSUES

In its original form, the Lottery Paradox brings to light the inconsistency of certain epistemic principles, among which are (A) If it is very likely that a conclusion is true, then it is rational to accept that conclusion. The paradox might motivate us to reject (A), and thus to deny that the holder of a ticket in a large, fair lottery can rationally accept the proposition that her ticket will lose. But, in any case, it seems highly intuitive that the ticket-holder does not know that her ticket will lose even if, unsurprisingly, it does.

This sort of intuition bears on a new epistemological problem, illustrated by the Car Theft Case (Vogel, 2004; *see also* Harman, 1974). The Car Theft Case arises in connection with a compelling epistemic principle, the Closure Principle for Knowledge. According to this principle, if a subject knows a proposition P, and knows that Q is a logical consequence of P, then the subject must also know, or be in a position to know, Q. Now consider the following example. You own a car which you parked a few hours ago on a side street in a major metropolitan area. You remember clearly where you left it. Do you know where your car is? A natural reaction is to say that you do. But it is true that every day hundreds of cars are stolen in the major cities of the United States. Do you know that your car has not been stolen? Many people have the intuition that you would not know that. In other words: You know the proposition (p) “My car is now parked on (say) Avenue A”. You also know that this proposition entails (q) “My car has not been stolen and driven away from where it was parked”. Yet, it seems, you do not know q, despite the fact that it is for you a clear logical consequence of p, which you do know. We have, then, what appears to be a counter-example to the

Closure Principle. Either the Closure Principle must be abandoned, which would be highly unattractive, or the judgments about what you know and do not know in the Car Theft Case must be reconsidered.

The Car Theft Case and aspects of the Lottery Paradox are related in certain ways, and an appreciation of these is essential to a resolution of the problems raised by the Car Theft Case. A key point is why it seems that you do not know that your car has not been stolen and driven away. Plausibly, you fail to know this for the just same reason that a person does not know that her ticket will lose the lottery. The two situations are structurally similar. Your car is one of many parked on the street in the city, and a few of these will be chosen, essentially at random, by car thieves and stolen. In this way, parking your car is like buying a lottery ticket, and having your car picked out by thieves and stolen is the unfortunate counterpart of having your ticket being picked as the lottery winner. Just as you do not know that your ticket will not win, you do not know that your car will not be stolen.

This sort of phenomenon, brought to light by the Car Theft Case, turns out to be very common. Call a proposition that is relevantly similar to the proposition that one's ticket will lose the lottery a *lottery proposition*. Exactly what features are essential to the lottery set-up and why that set-up precludes knowledge are difficult and controversial questions. Nevertheless, we do have a strong tendency to say (at least when the presence of lottery-like elements is apparent) that (D) people do not know lottery propositions. Further reflection discloses that (E) very many propositions about the world at large, which we take ourselves to know, entail lottery propositions. For example, I ordinarily take myself to know that the entrance to San Francisco Bay is now spanned by the Golden Gate Bridge. The proposition that the entrance to San Francisco Bay is now spanned by the Golden Gate Bridge entails that a meteorite did not strike and destroy the bridge five minutes ago. The latter is a lottery proposition, and it's easy to see how to generate similar lottery propositions in other

cases. (D) and (E) together with the Closure Principle for Knowledge seem to enforce the conclusion that people know virtually nothing about the world at large as it is now – a repellent result.

Suppose we set aside the possibility of abandoning the Closure Principle (see Vogel, 2004, for an argument that this would not help in any case). How, then, can we do justice to the motivations for denying that we know lottery propositions as well as the motivations for holding that we do have knowledge of the world outside our immediate environment? Again, take the Car Theft Case as a template. One approach is that some kind of shift takes place with the respect to the possibility that your car has been stolen. Initially, that possibility is somehow out of mind or out of consideration, but subsequently it becomes prominent and pertinent in some way (the exact nature of the shift is highly debated). This claim is endorsed by contextualists. They maintain generally that the truth-conditions for knowledge-ascriptions can vary from context to context. In particular, some contexts are such that lower, more relaxed standards for knowledge are in force; other contexts are such that higher, more stringent standards apply. According to contextualists, the Car Theft Case is puzzling because we do not note that a change of context, and a concomitant change in the standards for knowledge, occurs as it unfolds. In the initial context, the standards for knowledge are more easily satisfied, and the sentences "You know where your car is" and "You know that your car hasn't been stolen and driven away" are both true. Then, when the possibility of car theft becomes explicit, the context changes and the standards for knowledge are elevated. The two sentences are false with respect to the new context. The contextualist thus means to avoid wholesale skepticism by allowing for knowledge of both propositions in the initial context, while recognizing that, at least in the second context, we do not know the pertinent lottery proposition (see Cohen, 1998; Lewis, 1996; De Rose, 1996).

An alternative to contextualism is called *subject sensitive invariantism* (hereafter "SSI";

for articulations of this view, see Hawthorne, 2004, and Stanley, 2005). Advocates of SSI agree with contextualists that there is some bar to knowing lottery propositions, and that we can avoid far-ranging skepticism nevertheless. They agree further that an epistemically significant shift occurs in situations like the Car Theft Case. But, according to SSI, what shifts is not the epistemic standards employed by knowledge-attributors, but rather features of the subject's psychological and pragmatic situation. According to SSI, such features enter into the conditions for knowledge as much as traditionally recognized factors like truth or justification. An advocate of SSI might treat the Car Theft Case as follows (see Hawthorne, 2004). At first, the possibility that your car has been stolen is not psychologically salient to you. You know both that your car is where you parked it and that it hasn't been stolen and driven away. Later, however, the possibility of car theft does come to mind, and in this situation you know neither that your car is where you parked it nor that it hasn't been stolen and driven away. According to SSI, the factors which govern knowledge in this way are features of the *subject's* situation. The circumstances of knowledge-attributors other than yourself are irrelevant – contrary to what contextualists maintain.

The respective merits of SSI and contextualism are much debated, raising many key issues in epistemology and the philosophy of language. Other ways to treat knowledge of lottery propositions have also been proposed (see Vogel, 2004). In any event, coming to terms with our knowledge or ignorance of lottery propositions has become a central issue for epistemology, and doing so may have a significant impact on our general views about the nature of knowledge and justification.

In this brief survey, I have only hinted at the wealth of issues and ideas associated with the Lottery Paradox. There are various psychological and technical aspects of the Paradox that are of great interest, and there are important complexities that arise when one considers lottery situations in which the chances of winning vary from entrant to

entrant. The interested reader may consult the sources listed below for further details.

See also BAYESIAN EPISTEMOLOGY in Part I; EVIDENCE; LEWIS; RATIONALITY; SURPRISE EXAMINATION PARADOX.

# BIBLIOGRAPHY

- Bovens, L. and Hawthorne, J.: "The Preface, the Lottery, and the Logic of Belief," *Mind* 108 (1999), 241–64.
- Cohen, S.: "Contextualist Solutions to Epistemological Problems: Scepticism, Gettier, and the Lottery," *Australasian Journal of Philosophy* 76 (1998), 289–306.
- DeRose, K.: "Knowledge, Assertion and Lotteries," *Australasian Journal of Philosophy* 74 (1996), 568–80.
- Dretske, F. I.: *Knowledge and the Flow of Information* (Cambridge, MA: MIT Press, 1981).
- Douven, I.: "Nelkin on the Lottery Paradox," *Philosophical Review* 112 (2003), 395–404.
- Eells, E.: "On a Recent Theory of Rational Acceptance," *Philosophical Studies* 44 (1983), 331–43.
- Harman, G.: *Thought* (Princeton, NJ: Princeton University Press, 1974).
- Hawthorne, J.: *Knowledge and Lotteries* (Oxford: Oxford University Press, 2004).
- Heidelberger, H.: "Knowledge, Certainty, and Probability," *Inquiry* 6 (1963), 242–50.
- Jeffrey, R. C.: "Dracula Meets Wolfman: Acceptance vs. Partial Belief," in *Induction, Acceptance, and Rational Belief*, ed. M. Swain (Dordrecht: Reidel, 1970), 157–85.
- Kaplan, M.: "A Bayesian Theory of Rational Acceptance," *Journal of Philosophy* 78 (1981), 305–30.
- Klein, P.: "The Virtues of Inconsistency," *The Monist* 68 (1985), 105–35.
- Kyburg, Jr., H. E.: *Probability and the Logic of Rational Belief* (Middletown, CT: Wesleyan University Press, 1961).
- Kyburg, Jr., H. E.: "Conjunctivitis," in *Induction, Acceptance, and Rational Belief*, ed. M. Swain (Dordrecht: Reidel, 1970), 55–82.
- Lehrer, K.: "Coherence and the Racehorse Paradox," *Midwest Studies in Philosophy* 5 (1980), 183–91.
- Levi, I.: *Gambling with Truth* (Cambridge, MA: MIT Press, 1967).
- Lewis, D.: "Elusive Knowledge," *Australasian Journal of Philosophy* 74 (1996), 549–67.

- Nelkin, D. K.: "The Lottery Paradox, Knowledge, and Rationality," *Philosophical Review* 109 (2000), 373–409.
- Pollock, J.: "Epistemology and Probability," *Synthese* 55 (1983), 231–52.
- Sanford, D. H.: "Lotteries, Horseraces, Probability, and Projection" (abstract), *Nous* 17 (1983), 70–1.
- Stalnaker, R.: *Inquiry* (Cambridge, MA: MIT Press, 1984).

- Stanley, J.: *Knowledge and Practical Interests* (Oxford: Oxford University Press, 2005).
- Vogel, J.: "Speaking of Knowledge," *Philosophical Issues* 14 (2004).

JONATHAN VOGEL

**luck, epistemic**    *see* EPISTEMIC LUCK.

# M

---

**Marxism** Marx had no patience for the traditional projects of epistemology: the attempt to refute scepticism by establishing a foundation of certainty for human knowledge, and the search for philosophical criteria by which to adjudicate knowledge claims. He rejected the methodological solipsism (*see* SOLIPSISM) of the epistemologist's customary point of departure: the Cartesian picture of a disembodied mind, reflecting on its own self-awareness, and seeking grounds to believe that its ideas accurately represent an "external world". For Marx, this approach unintelligibly supposes that the subject can attain a form of self-awareness prior to and independent of a relation to nature and society. Marx shared Hegel's contempt for the idea that the philosopher may occupy a privileged position, somehow outside our evolving conceptual scheme, from which to assess the relation our concepts bear to reality itself. At the centre of his alternative vision, Marx places the concept of *activity* or *practice*. He urges that we replace the abstract contemplating self of classical epistemology with a historically situated subject, engaged in an active relation to the natural and social environment, and intelligible only in light of that relation. Marx thus dismisses philosophical disputes over the possibility of knowledge as "purely *scholastic*". Whether objective truth may be attributed to human thought is, he maintains, ultimately "a *practical* question" ("Theses on Feuerbach").

Marx's position forms part of a general critique of speculative philosophy, developed in the *Economic and Philosophical Manuscripts of 1844*, the "Theses on Feuerbach" (1845), the German Ideology (with Engels) (1845–6), and other writings. Philosophy, at its most progressive, seeks to characterize the human essence so as to determine the circumstances

in which human flourishing is possible. Human misery, however, has its cause, not in intellectual confusions about the nature of man, but in the alienated character of existing economic and social life. Indeed, the philosopher's characteristic questions, and his inevitably empty answers, are symptoms of this alienation rather than a potential cure. Philosophy is thus at best impotent, and at worst a source of ideological mystification that serves to perpetuate the status quo.

Almost all of Marx's diverse followers have endorsed both his hostility to epistemology and his emphasis on "practice". This does not mean, however, that they have neglected issues of the nature and origin of knowledge. On the contrary, Marx's legacy has prompted intense theoretical discussions of scientific method, of objectivity (*see* OBJECTIVITY), of the relation between natural and social scientific modes of explanation (*see* EXPLANATION), of necessity and prediction, of the nature of "false consciousness", and many other issues which are "epistemological" in a broad sense. There is little consensus on these questions; indeed, political disagreements between Marxist factions have often been accompanied by differences in their stances on such theoretical concerns. The various interpretations of Marx's vision of the relation of "subject" and "object", individual and world, form a spectrum with, at one end, austere varieties of *scientific realism* that accentuate Marx's confidence in the power of science to render objective reality transparent, and at the other, forms of *anthropocentrism*, which represent human practice as in some sense constituting the reality we come to know.

Scientific realist interpretations, which have their precedent in Engels, Plekhanov and Lenin, begin from the assertion that an "external", material world exists as an



objective reality prior to, and independently of, human beings. In virtue of its independent existence, this world may be described as a reality of “things-in-themselves”. *Contra* Kant, however, “things-in-themselves” are knowable, and as we acquire knowledge of the world, so it is transformed from a “thing-in-itself” into a “thing-for-us” (see IN ITSELF/FOR ITSELF). Our primary access to reality is through sense-perception. Some Marxists of this persuasion subscribe to representative realism (see REPRESENTATIVE REALISM), on which we are directly acquainted only with sensations caused in us by objects, and others to direct realism (see DIRECT REALISM), where the senses yield direct access to material objects themselves (Lenin’s position is intriguingly ambiguous between the two). All agree, however, that on the basis of our senses we construct theories that are able to reflect the world accurately: objective truth is possible. At any point in history, our theories are only “relatively” true, since they capture the truth only partially. However, as history progresses, so our theories tend towards “absolute” truth. Marx’s emphasis on activity finds its reflection in the claim that “practice is the criterion of truth”: through action and experimentation we verify our conceptions. Accordingly, sceptical objections are also dismissed by appeal to “practice”, either on the grounds that the sceptic’s doubts are empty because they have no impact upon action, or because once we have established, by all the usual practices, that *p* is true, there is no longer room for genuine doubt that *p*. Finally, this approach holds that it is science, and not philosophy, that determines the nature of reality. Philosophy’s role is to generalize the results of the sciences (thus Engels makes his controversial claim that the development of nature is governed by dialectical laws, not as a metaphysical conclusion, but as a generalization from instances of scientific explanation).

At the opposite end of the spectrum, *anthropocentric* readings emphasize Marx’s claim that not only is the human subject an active being, but the object world itself must also be conceived as “human sensuous activity, practice” (“Theses on Feuerbach”). Marx takes

human praxis to have a world-transforming character. By acting upon reality, human beings change its very nature: the world they confront is no longer brutally physical in kind; it is a “humanized” environment. The strongest version of this view (attributed to Marx by Habermas, and by the Soviet philosopher E. V. Ilyenkov) takes Marx to be addressing a Kantian question: What are the necessary conditions for the possibility of experience and thought? Marx’s answer invokes activity, the material transformation of the natural world, as the essential precondition of objectivity. Everything Kant treats as a priori forms of thought must in fact be construed as “forms of the self-consciousness of social beings”, embodied in the material, linguistic and intellectual activities of the community (Ilyenkov). On this “transcendental pragmatism”, human individuals become thinking beings as they appropriate these forms of activity in socialization (a claim developed in the Soviet psychological tradition, especially by Vygotsky and Leontiev). This is the sense in which the human essence is “the ensemble of social relations”, and human agents socially-constituted beings.

What sense can such a position make of the world as it is “in itself” independent of human activity? Since the world is said to be given to us only through incorporation into our activities, there can be no transcendent perspective from which we may determine how the world is prior to that incorporation. Some Marxists have concluded that “no ‘reality’ exists in and for itself, but only in historical relation to human beings who modify it” (Gramsci). Nevertheless, the Marxist can argue that the distinction between features of the world that depend on our activity and those which do not is an empirical distinction drawn within the framework of our activities, and that the criteria by which we determine these features are just those of our everyday and scientific practices. Most Marxists, however, have pursued a less quietist line, either by proposing to analyse empirical truth in terms of activity (i.e. a set of “activity conditions” may be given for each sentence setting out the constraints its acceptance would have on possible activity),

or by offering a functional account of truth where, as Kolakowski puts it, "knowledge has no epistemological value distinct from its value as an organ of human self-affirmation". Kolakowski attributes to Marx a "generic subjectivism" on which "truth" is ultimately a property of beliefs which promote human liberation, and "falsehood" a property of forms of ideological mystification which serve to perpetuate human alienation. (Similar interpretations have provoked a range of contrasting views, from the crudest identification of truth with class interest, to the sophisticated *ideologiekritik* of the Frankfurt School.) It remains controversial, however, whether such anthropocentric readings can make real sense of a world independent of human agency, and they are often accused of collapsing into an idealism incompatible with Marx's avowed naturalism (see NATURALISM) and materialism.

Although not all of Marx's followers may be represented as subscribing to pure scientific realism or anthropocentrism, it is often fruitful to read those who cannot (e.g. Althusser, Lukacs, Marcuse) as seeking an intermediate position that would reconcile the insights of both stances. However, whether it is possible to give substantive theoretical content to the concept of activity within the terms of a realist epistemology remains a live issue in the Marxist tradition.

See also CONTINENTAL EPISTEMOLOGY.

## BIBLIOGRAPHY

- Engels, F.: *Anti-Dühring*, trans. E. Burns (Moscow: Progress, 1947).
- Engels, F.: *Dialectics of Nature*, trans. C. Dutt (Moscow: Progress, 1934).
- Gramsci, A.: *Il materialismo storico e la filosofia di Benedetto Croce* (Turin: Einaudi, 1949).
- Habermas, J.: *Knowledge and Human Interests*, trans. J. Shapiro (London: Heinemann, 1971).
- Ilyenkov, E.: "The Concept of the Ideal," in *Philosophy in the USSR: Problems of Dialectical Materialism*, trans. R. Daglish (Moscow: Progress, 1977).
- Kolakowski, L.: *Main Currents of Marxism*, 3 vols (Oxford: Oxford University Press, 1978).
- Lektorsky, V. A.: *Subject, Object, Cognition* (Moscow: Progress, 1984).
- Lenin, V. I.: *Materialism and Empirio-criticism*, in his *Collected Works* (London: Lawrence and Wishart, 1960–72), vol. 14.
- Marx, K.: *Economic and Philosophical Manuscripts of 1844*, in K. Marx and F. Engels, *Collected Works* (CW) (London: Lawrence and Wishart), vol. 3.
- Marx, K.: "Theses on Feuerbach," in CW, vol. 5.
- Marx, K. and Engels, F.: *German Ideology*, in CW, vol. 5.

DAVID BAKHURST

**mathematical knowledge** The following paradox is presented by mathematical knowledge.

Mathematics is, historically, perhaps the earliest science. For many thinkers mathematical knowledge, by virtue of its seeming absolute certainty, has served as an ideal or paradigm for all the sciences. For example, the mathematical method was extended by rationalistic scientists like Galileo and Descartes to the realm of what we today call physics. Even if we do not go so far as to regard physics as the "mathematics of motion", mathematical knowledge seems to be indispensable for modern scientific knowledge – the mathematically illiterate cannot read the papers of Dirac, Einstein, or Feynman. We can say, therefore, that mathematics is at least continuous with scientific knowledge.

Yet mathematics seems continuous also with metaphysics. Indeed, mathematics seems not to deal with nature – its subject matter could be variously described as "ideal", or "abstract". Figures like triangles and spheres, etc. are ideal – they are perfectly shaped, and have no breadth. They seem to be the limit of some infinite process unattainable in the actual world. Numbers, on the other hand, are abstract: they are not, apparently, the idealization of any actual objects. Furthermore, the very certainty of mathematical knowledge seems to set it apart from empirical knowledge. Kant put the matter polemically in his "good company" argument: one cannot reject metaphysics without rejecting mathematics. This argument, of course, was intended as an *ad hominem* argument against

the empiricists (like David Hume) who, being pro-science, would never reject mathematics (see EMPIRICISM).

Latter-day “naturalists” are also subject to Kant’s “good company” argument. As formulated by Paul Benacerraf, the argument goes: for the naturalist (see NATURALISM), who takes seriously the findings of modern science, knowledge is a causal interaction between knower and environment. (The causal interaction involved may be seen as an energy transfer between an individual knower and his environment, or – as in “evolutionary epistemology” – a process of natural selection that shapes an entire species.) But mathematical objects do not participate in causal interactions. Hence mathematical knowledge is impossible, unless we drop naturalism. (It is interesting that something like this argument already occurs in Plato’s *Sophist* at §248: “If knowing is to be acting on something, it follows that what is known must be acted upon by it, and, so on this showing, reality when it is being known by the act of knowledge must, in so far as it is known, be changed owing to being so acted upon – and that, we say, cannot happen to the changeless.”)

We can, therefore, sum up the paradox of mathematical knowledge as follows: without mathematical knowledge, there is no scientific knowledge – yet the epistemology (“naturalism”) suggested by scientific knowledge seems to make mathematical knowledge impossible!

In what follows, I shall outline the various strategies to deal with the paradox of mathematical knowledge that have been suggested by philosophers throughout the ages.

#### REALIST STRATEGIES

There is a non-causal relationship between the soul, or mind, of humanity, and the world of mathematics. The naturalist epistemology is inadequate. (This need not involve rejection of “naturalism” *in toto*, however.) Of course, this idea is the basis of Plato’s entire metaphysics (see PLATO), but many mathematicians have felt the same way. G. H. Hardy

(1929) and Roger Penrose (1989), for example, speak of “seeing” that a mathematical proposition is true, proof being necessary only to persuade others. In our century, the great logician Gödel (1948) endorsed the view that there is some other connection between ourselves and reality than sense perception, and that this “mathematical intuition” can account for mathematical knowledge. In fact Gödel’s discoveries in logic have been used to support the realist position: Gödel’s theorem has been interpreted as showing that, for any serious system of mathematical axioms, the mathematician can know a mathematical truth that does not follow from those axioms; realists argue that the only way this could be true is by mathematical intuition. This argument can be resisted, however, since it presupposes what is doubtful: that we know that the axioms of mathematics are jointly consistent. (The unprovable truth, call it *G*, says, roughly, “I am not provable” – but if the axioms are inconsistent, then everything is a provable, so *G* is false.) Granted, one could argue that we know that the axioms of mathematics are consistent because we intuit their truth – true axioms are perforce consistent. But the appeal to Gödel’s theorem then becomes superfluous and circular.

Though “Platonist” realism in a sense accounts for mathematical knowledge, it postulates such a gulf between both the ontology and the epistemology of science and that of mathematics that realism is often said to make the applicability of mathematics in natural science into an inexplicable mystery.

Recently, therefore, some writers have attempted a broadly realist position based on “structuralism”: mathematics is about structures, not objects. Benacerraf (1965) already suggested such a position in “What Numbers Could Not Be”, but such writers as Michael Resnik (1982) and Penelope Maddy (1980) have developed the position, which can be seen as an “Aristotelian” attempt to reconcile the naturalist epistemology with an attenuated mathematical realism. There is a mathematical intuition, but it is not a separate faculty from empirical sense perception. This idea is supported also by the work of the influential American philosopher of mathematics,

Charles Parsons (1979–80), and by that of the present author (Steiner, 1975). Its proponents claim to make the applicability of mathematics to the empirical world intelligible.

#### KANTIAN STRATEGIES

These argue that mathematical knowledge is a necessary condition of empirical knowledge. Kant himself (*see* KANT) argued that the laws of mathematics are actually constraints on our perception of space and time. In knowing mathematics, then, we know only the laws of our own perception. Physical space in itself, for all we know, may not obey the laws of Euclidean geometry and arithmetic, but the world as perceived by us must. Mathematics is objective – or “intersubjective” – in the sense that it holds good of all perceptions of the whole human race, past, present, and future. For this reason, also, there is no problem with the applicability of mathematics in empirical science – or so the Kantians claim.

Kant’s view of mathematical knowledge is often regarded as having been refuted by the discovery of non-Euclidean geometry, and curved spaces, but these geometries are “locally Euclidean” (i.e. a curved region looks flatter and flatter the smaller it gets), and Kant could have made the more modest claim that any single field of vision is, *a priori*, a locally Euclidean space.

At any rate, the modern branch of mathematics known as topology, developed by the famous mathematician Henri Poincaré, can be regarded as that part of geometry for which the Kantian thesis remains a viable option.

Poincaré (1907) was a Kantian in arithmetic as well. For him the law of “mathematical induction” was the essence of arithmetic: any property *P* of zero which is “hereditary” (i.e. it holds of  $n + i$  whenever it holds for  $n$ ) holds of every natural number. This principle is justified by noting that if *P* holds of zero then it holds of *i*, so by *modus ponens*, it actually holds of *i*. By continual use of *modus ponens* and hereditariness we know that we can eventually “arrive at” any number  $n$  and show that *P* holds of  $n$ . This is the kind of self-knowledge

of which Kant spoke; Poincaré held, inspired by Kant, that it cannot be reduced to “logic”; on the contrary, mathematical induction is a principle about what logic can do. Poincaré knew of Frege’s and Russell’s efforts (as part of their “logicism”) to convert the principle of mathematical induction to a logical definition: (roughly)  $n$  is a natural number just if it is subject to the law of induction (*see* FREGE; RUSSELL; LOGICISM). But he regarded this definition as circular.

In the 1920s, of the present century, two outstanding logicians – Hilbert and Brouwer – argued for competing versions of Kantianism: “formalism” and “intuitionism”. Both men accepted that the ultimate content of mathematics is intuition, and that classical mathematics goes beyond the merely intuitive (for example, in its acquiescence in infinite totalities), and therefore does not give knowledge. But where Brouwer (1913) advocated replacing classical mathematics by a new kind of mathematics (which he and his followers proceeded to develop), Hilbert (1926) took the conservative approach of justifying the so-called “ideal” proofs of classical mechanics as instruments of discovery.

#### EMPIRICIST STRATEGIES

John Stuart Mill (1843) (*see* MILL) is the most outstanding empiricist (*see* EMPIRICISM) to adopt the radical stance that mathematics is a branch, not of logic, but of physics. The relative certainty – and applicability – that attaches to mathematics results from the great range of empirical confirmation that mathematics enjoys. For geometry, of course, the position is widely accepted today, but it is more difficult to see how one could regard arithmetic as empirical. For example, what would be empirical evidence for  $1234 \times 1234 = 1,522,756$ ? Certainly nothing that would justify our actual conviction concerning this product. Recently some authors, most notably Philip Kitcher (1983), have attempted to refurbish this position by arguing that the axioms of arithmetic can be given an empirical interpretation and supported by evidence. The logicist argues that

mathematics is just a branch of logic, or, more generally, and more traditionally, “analytic truth”. Though this is not an empiricist interpretation of *mathematics*, it is congenial to empiricism, since it appears to give a non-metaphysical account of mathematical knowledge. Empiricists have assumed that once it is proved that “mathematics is logic” the problem of mathematical knowledge no longer arises, since there is no philosophical problem about logical knowledge. Nevertheless, it should be remembered that it was Leibniz who first conjectured, and tried to prove, mathematics is logic – but he had a metaphysical picture of logical knowledge, as being “true in all possible worlds”. Nor was Frege, who invented modern logic, in part, to prove that mathematics is nothing but logic (and thus founded the modern logicist school), an empiricist. Conversely, the common belief that Hume’s view of mathematics as the study of “relations among ideas” prefigures modern logicism is actually due to Kant’s influence. For Kant (in the *Prolegomena*) characterized Hume’s theory as “amounting to” the claim that mathematics is “analytic”, a quite doubtful characterization, in the light of Hume’s explicit declaration (*Treatise* I, ii, 4) that such propositions as “The shortest distance between two points is a straight line” are not true by “definition”. (Perhaps a better translation of Hume’s doctrine into Kantian language would be: mathematics is synthetic a priori. This does not mean that Kant and Hume had the same philosophy of mathematics, though, because their theories of the a priori, i.e. their theories of necessary truth, were quite different.)

Nevertheless, twentieth-century empiricists like the later Russell, Carnap, Ayer and Hempel saw logicism as an appropriate doctrine (see CARNAP; AYER; HEMPEL). They, unlike Leibniz, saw logical validity as a matter of linguistic rules, the rules governing words like “all”, “and” and “not”; knowledge of these was considered free of metaphysics (see LOGICAL POSITIVISM). The problem with this “logicism”, as it has come to be called, was a technical one: no logician was ever able to reduce mathematics to a system of “logic” which could plausibly be called “analytic”

(e.g. classical mathematics can be reduced to set theory, but set theory does not qualify as an “analytic” science).

A neglected virtue of logicism, in my opinion, is that it solves – or dissolves – some of the problems of mathematical applicability. Logicism shows that all mathematical science can be represented in set theory. Thus the only relation between physical objects and mathematical objects we need recognize is that physical objects can be members of sets (sets being mathematical objects). Presumably, if we believe in sets at all, we have no further problem of seeing how physical objects can be members of sets, so some (but not all – see Steiner, 1989) of the problems concerning mathematical applicability disappear. This virtue of logicism does not depend upon our recognizing set theory as “logic”.

#### PRAGMATISM

In pragmatist theories of mathematical knowledge (see PRAGMATISM), the indispensability of mathematics in all other knowledge, especially in physical sciences, is converted into a justification of mathematical “commitment”. The only justification of mathematical assertions is that we can’t help ourselves, if we want to achieve the goals of science and everyday life. While this might be regarded as weak confirmation indeed (and certainly no explanation of the “obviousness” of mathematics, as Parsons has pointed out), pragmatists argue that mathematics is in the same boat as every scientific theory. In this sense, their argument is similar to the “good company” argument of Kant.

Quine (e.g. 1960, 1970, and many other writings) (see QUINE), who has made this pragmatist-“Kantian” argument famous (though Quine’s predecessor at Harvard, C. I. Lewis, already preached a synthesis of Kantianism and pragmatism in *Mind and the World Order*), adds a Deweyite “naturalistic” element: ultimately what justifies mathematics and every justified theory is its usefulness in predicting “surface irritations”. What is striking about Quine’s philosophy of mathematics, however, is that it is explicitly



Platonist in its ontology (though not, of course, in its epistemology). Quine agrees with Frege that modern mathematics is heavily “committed” to abstract objects – and disagrees with Wittgenstein and the British “ordinary language” school, who regard the “commitment” as a manner of speaking, similar, if you will, to the commitments of a politician which nobody takes seriously.

For Quine, again, commitment to abstract objects is justified on pragmatic grounds: we have no choice if we want to do science. However, by combining Platonism, pragmatism and naturalism, Quine seems to make it impossible to give a theory of mathematical discovery. His reasoning can give, at best, a *post facto* pragmatist justification for mathematics once it has been discovered. For Quine has no place, in his philosophy, for “mathematical intuition” either in the Kantian sense or the Platonic sense. Thus, Quine’s picture of mathematical discovery is that of a senseless procedure that accidentally gets *post facto* justification.

I will now present approaches that “solve” the paradox by denying the very existence of mathematical knowledge! According to these approaches, mathematical theorems do not express “truths”, hence there is nothing to “know”. Mathematics can play its role in science and daily life without being “true”.

#### INSTRUMENTALISM

According to this view, mathematics is a tool for making inferences in other fields, but is not itself a science. Perhaps the simplest and most radical form of this view is “fictionalism”. The fictionalist is not interested in intervening in mathematical discourse, but in interpreting it as fiction. The fictionalist argues that, in principle, one could do without mathematics – even in science. But mathematics allows the scientist the use of compact, elegant proofs, of what otherwise would be cumbersome deductions. A recent defence of this position is by Hartry Field (1980). Field argues that one can rewrite physical theories without any reference to “mathematical objects”, and then prove that adding mathematical axioms

does not increase the deductive power of the rewritten theory. He actually shows how one might get “rid of” mathematical objects in a particular theory, namely classical gravitation, and how, by what amounts to a consistency proof, one shows that adding mathematics produces a “conservative extension” of this “nominalistic” theory of gravitation. Field claims explicitly, as a virtue of his fictionalism, that it eliminates the puzzles concerning mathematical applicability, since we have no longer to worry about the alleged gulf between the subject matter of mathematics and that of the natural sciences.

However, Field’s thesis is controversial. Instead of “mathematical objects”, Field’s version of gravitation takes points in space–time as real entities, and some argue that this is out of the frying pan and into the fire. Others protest that there are physical theories that are not space–time theories at all, like quantum mechanics. Some argue that the consistency proof will itself raise the ghosts of the departed mathematical entities up in the “metalanguage”. And there are technical objections based on the use by Field of “higher-order” logic in his version of gravity.

#### CONVENTIONALISM

Conventionalism is the view that mathematical theorems are “true by convention” (*see* CONVENTION). Poincaré argued, for example, that the difference between Euclidean and non-Euclidean geometry is not a factual, but only a conventional, difference. That is, we can adopt either Euclidean or non-Euclidean geometry according to convenience, since geometry is the study of measurements, and measuring instruments are subject to the forces of nature. For example, we can explain the failure of angles of triangles to sum to 180 degrees either by postulating “deforming” forces, or by invoking non-Euclidean geometry. This is another way of saying that there is no such thing as “knowledge” in geometry, unless “knowledge” that such-and-such are the consequences of our conventions is meant. Note that Poincaré does not extend his conventionalism to mathematics

in general: his general point of view in topology (and arithmetic, as we have seen) is not conventionalist, but Kantian.

Thus, for example, Poincaré's position implies that whether a surface is flat or parabolic is a matter of convention; but that the surface is of two dimensions is not conventional at all. No conceivable force could alter our "rulers" in such a way as to cause a two-dimensional surface to appear three-dimensional. Though this position is Kantian, it should be noted that Poincaré gives a biological explanation for our perception of dimension, particularly why we perceive the world in three dimensions and not more.

The later Wittgenstein (1956, 1976) (see WITTGENSTEIN) is often regarded as a conventionalist, though a much more thoroughgoing one than Poincaré, since Wittgenstein makes no distinction between geometry and other branches of mathematics, including arithmetic. And it is true that Wittgenstein often refers to mathematical theorems as conventions; for this reason, he is discussed here.

Yet I have grave reservations about calling Wittgenstein a "conventionalist" (aside from the general problem that he rejected all philosophical "positions" or theories):

1. Wittgenstein does not say that mathematical theories "follow from" conventions. On the contrary, for Wittgenstein, each step in a mathematical proof is a new convention, not just the axioms (as for Poincaré). Conventions, for Wittgenstein, do not "bind" anybody.
2. This apparent anarchical element in Wittgenstein's position, however, can mislead. When Wittgenstein speaks of a theorem as a convention, he does not mean that there is a genuine option to ignore the proof and accept the negation of the theorem. All mathematical conventions, for Wittgenstein, presuppose empirical regularities which, in his words, are then "hardened" into rules. That is, what happens most of the time is regarded as the norm, and deviations are to be explained as mistakes, perturbations, etc. Empirical regularities connected with measuring are "hardened" into theorems of geometry,

while regularities in counting are "hardened" into theorems of arithmetic and number theory.

3. Wittgenstein does side with the conventionalist in one sense, however, in that he regards it as very misleading to speak of mathematical knowledge. To say that someone knows the Pythagorean Theorem is, for Wittgenstein, like saying that someone knows that 12 inches = 1 foot. But there is a tendency for us to regard mathematical knowledge, rather, as like empirical knowledge, a tendency which leads either to empiricist or Platonist theories of mathematics, both of which Wittgenstein rejected.

See also A PRIORI KNOWLEDGE in Part I: GEOMETRY.

#### BIBLIOGRAPHY

- Benacerraf, P. and Putnam, H., eds: *Philosophy of Mathematics: Selected Readings* (Cambridge: Cambridge University Press, 1984).
- Benacerraf, P.: "What Numbers Could Not Be," *Philosophical Review* 74 (1965), 7–73.
- Brouwer, L. E. J.: "Intuitionism and Formalism" (1913); reprinted in Benacerraf and Putnam, 1984.
- Field, H.: *Science without Numbers: A Defense of Nominalism* (Princeton, NJ: Princeton University Press, 1980).
- Gödel, K.: "What Is Cantor's Continuum Problem?," (1964); reprinted in Benacerraf and Putnam, 1984.
- Hardy, G. H.: "Mathematical Proof," *Mind* 38 (1929), 1–25.
- Hilbert, D.: "On the Infinite" (1926); reprinted in Benacerraf and Putnam, 1984.
- Kitcher, P.: *The Nature of Mathematical Knowledge* (Oxford: Oxford University Press, 1983).
- Maddy, P.: "Perception and Mathematical Intuition," *Philosophical Review* 89 (1980), 163–96.
- Mill, J. S.: *A System of Logic* (London: Longmans, 1843).
- Parsons, C.: "Mathematical Intuition," *Proceedings of the Aristotelian Society* 80 (1979–80), 142–68.
- Penrose, R.: *The Emperor's New Mind* (Oxford: Oxford University Press, 1989).
- Poincaré, H.: "Science and Hypothesis" (1907); excerpt reprinted in Benacerraf and Putnam, 1984.

- Quine, W. V.: *Word and Object* (Cambridge, MA: MIT Press, 1960).
- Quine, W. V.: *Philosophy of Logic* (Englewood Cliffs, NJ: Prentice-Hall, 1970).
- Resnik, M. D.: "Mathematics as a Science of Patterns: Epistemology," *Nous* 16 (1982), 95–105.
- Steiner, M.: *Mathematical Knowledge* (Ithaca, NY: Cornell University Press, 1975).
- Steiner, M.: "The Application of Mathematics to Natural Science," *Journal of Philosophy* 86 (1989), 449–80.
- Wittgenstein, L.: *Remarks on the Foundations of Mathematics* (Oxford: Blackwell, 1956).
- Wittgenstein, L.: *Wittgenstein's Lectures on the Foundation of Mathematics*, ed. C. Diamond (Ithaca, NY: Cornell University Press, 1976).

MARK STEINER

**memory** The epistemology of memory and the metaphysics of memory are tightly related. The latter offers an account of the nature of memory while the former examines the conditions under which memory is conducive to justified belief and knowledge. Both the metaphysics and the epistemology of memory are distinct from (although certainly not entirely unrelated to) the psychology of memory. Research in the psychology of memory is aimed at exploring the way memory actually functions, the way memory is realized in neurophysiological states, the circumstances and conditions in which memory is reliable or unreliable, etc. While all of this is of obvious importance and arguable relevance to the philosophical enterprises mentioned above, it is nevertheless distinct. The metaphysics of memory, concerned with giving a characterization of memory at a more general and abstract level, aims at identifying the nature of memory regardless of its particular physical implementation. The epistemology of memory, on the other hand, will benefit from information about the circumstances under which memory is reliable (or not), but such information alone will not answer key normative questions like under what conditions is a memory belief justified and when, if ever, can it count as knowledge.

What, then, is memory? The most obvious answer is that memory is a psychological

process that stores information so that it can be used at a later time. As a very general definition, this will do, but epistemologists are interested in the evaluation of beliefs, and so what we need for present purposes is an account of what it is for a belief to be appropriately categorized as a memory belief. Let us approach this question by clearing up several important points.

First, "S's belief that *p* is a memory belief" does not entail "S remembers that *p*". Memory beliefs need not be true: one can *seem* to remember *p* while *p* is false. But if what one seems to remember is not true one isn't genuinely remembering. Second, "S remembers that *p*" entails the truth of *p*. One can't genuinely remember what is in fact false. Third, "S remembers that *p*" does not entail "S knows that *p*". Suppose that a person *A* is told by source *B* (a source *A* has very good reason to trust as both reliable and knowledgeable) that all *A*'s apparent memories before the age of seven are not veridical, and that they were produced by brain surgery *A* had when he was seven. Unbeknownst to the protagonist, the informant is lying: *A* has never had brain surgery, and his memories from even a young age are generally accurate. *A* now seems to recall blowing out the candles on his birthday cake at the age of six, and indeed his current recollection traces back to this event; in short, he is genuinely remembering blowing out the candles. However, given what *B* has said, together with the fact that *A* has strong reason to think *B* is both reliable and knowledgeable in this regard, *A*'s justification for thinking that he blew out the candles at the age of six is defeated. Consequently, *A* remembers blowing out the candles but he doesn't know that he did (see Goldman, 1979).

In typical cases, a memory belief is often accompanied by an image of the event remembered. Some philosophers (e.g. Aristotle, Hume, and Russell) have suggested that the presence of a memory image is a necessary condition of remembering. For example, if asked where you were when you learned of the events of 11 September 2001, you'll likely recall an image of a friend relaying the

information to you or the radio, TV, or computer screen that carried the first pictures you saw of the burning towers. However, the claim that memory *requires* an image is clearly mistaken. Consider a case in which you are asked the name of your mother while you are intently watching a baseball game. You quickly recall that her name is “Mary” although you don’t call to mind an image either of her or even of the word “Mary”. Or suppose you are asked who was president during the United States civil war. You might quickly respond “Abraham Lincoln” without calling to mind any image at all. Perhaps both of these cases would typically involve images if the subject were to give the matter her complete attention, but it can scarcely be doubted that we frequently recall all kinds of information on the fly without accompanying images.

The two cases described above involve what philosophers call “occurrent memory”. Occurrent memory beliefs are beliefs of which the subject is conscious; they are distinguished from dispositional or stored beliefs. The claim that all memory involves imaging does not assert that all dispositional memory beliefs must have occurrent images, but merely that all occurrent memory beliefs must be accompanied by images.

What, then, is essential for memory belief? Consider a common case in which one recalls an event E, involving a typical memory image, and forms a corresponding memory belief. In virtue of what is this belief a memory belief?

There are two rather different types of answers, one causal-external and the other experiential-internal. According to a causal-external account, a belief will be a memory belief if and only if it has the right kind of causal history – for example, iff it bears the appropriate causal relations to a belief or experience the subject had at an earlier time (see Martin and Deutscher, 1966, for an argument that a causal requirement is crucial for remembering). The motivation for this kind of account is primarily a desire to distinguish between belief kinds by the psychological processes that produce them. Arguably, an occurrent belief that is not causally related in the appropriate

way to past beliefs/experiences can’t be thought to have been produced by memory. The causal view can allow that not all memory is veridical both because the original belief might have been false or because the causal path, while strong enough to be sufficient for memory, might be degraded to the point where the content of the original belief or experience is altered.

In contrast with the causal-external account of memory belief is the experiential-internal view. According to the latter, it is necessary and sufficient for a doxastic state to be a memory belief that it *seems to the subject* (or better “would seem on reflection”) to be something she remembers. Presumably, the great majority of beliefs that subjects would take to be memory beliefs will have the kind of causal connection required by the causal-external position. Where the accounts differ is in cases in which the subject seems to remember that p but in which there is in fact no appropriate causal connection with a past belief or experience. For example, I might seem to remember that the cake at your sixth birthday party was chocolate because I seem to remember being at your party and seeing it, when in fact I wasn’t there at all and the impression that the cake was chocolate has been produced by my imagination. In such a situation, the causal-external account will not deem this an instance of memory belief whereas the experiential-internal account will. Not surprisingly, the account that one accepts here will likely be colored by one’s general epistemic sympathies. The internalist will be tempted to think that, since justificatory questions are primarily perspectival, the epistemic standards for the belief in question will be identical with what the standards would be for a belief with the right causal pedigree, and hence it should be counted as a memory belief. The externalist will likely think that the perspective of the subject is not definitive and that, if the causal connection to a previous belief or experience is not there, then it shouldn’t be counted as a memory belief, however the subject would likely categorize it. There is, I believe, no theory-neutral way to solve this dispute.

Perhaps the most significant epistemic issue concerning memory is its reliability. It is plausible to suppose that if memory is generally reliable, then the beliefs to which it gives rise can count as justified or even knowledge; whereas if memory isn't generally reliable, then memory beliefs fail to count as knowledge and likely will not even be justified. So what are the prospects for showing its general reliability? While some have maintained that, necessarily, memory is reliable (e.g. Malcolm, 1963), their claim can be shown to be false. However, it must be said in their defense that the standard demon world cases that apparently demonstrate the possible, global unreliability of perception won't work so clearly in the case of memory. For in a standard demon world it is sufficient for the falsity of memory beliefs that the input to the perceptual systems (and any other input processes) be false or misleading (mental images don't have truth values and so can't be false but they can be misleading). For, even at its best, memory is primarily a preservation rather than a generative process: if what is put in is false, what comes out is false, too, but not because memory itself is unreliable but because the inputs are unreliable. So the fact that in the demon world the deliverances of memory are false doesn't show that the fault is with memory.

Be this as it may, even if the process of memory is at best conditionally reliable, memory *beliefs* as they are characterized by the experiential-internalist will nevertheless be generally unreliable in a demon world. For everything that the subject seems to remember will turn out to be false, and so memory beliefs will be unreliable given the experiential-internal account of memory belief. And if the demon is clever he can also make the process of memory conditionally unreliable and so unreliable even on the causal-external account. Suppose the demon imparts in the minds of his victims a process that systematically alters the content of the stored memory whenever it is recalled. If, in most instances, the original, occurrent belief and the stored belief are the same in content, and if the stored belief is causally active in (more or less) the right way when the belief is recalled, then the causal-external account of memory

belief will count this as memory belief. And such belief will not be even conditionally reliable. So on both the experiential-internal and the causal-external accounts of memory belief there are worlds in which memory belief is thoroughly unreliable and so the claim that memory is necessarily reliable is false.

Still there are reasons for being skeptical about the prospects for showing memory to be generally reliable. First, Russell (1921) presented an argument for skepticism about that past that has a direct application to memory: Consider the hypothesis that the universe and what it contains was created five minutes ago. Any non-circular demonstration of the overall reliability of memory would have to either be or contain the resources for a successful refutation of Russell's hypothesis. Yet it is generally conceded that there is no such extant refutation.

Even putting aside skeptical hypotheses, there is another reason to think that the reliability of memory can't be demonstrated. Such a demonstration would be in the form of a sound argument that includes no epistemic circularity (i.e. none of the premises is epistemically dependent on memory). Consider even a simple, syllogistic argument. If memory is to play no part in the production of this demonstration of memory's reliability, then the entire argument must be held consciously at once. More than that, however, a demonstration is only as good as the premises it contains. So one will have to be able to show that the premises are true (or at least well justified), too. But, to do that, one will have to push the main argument out of consciousness to make room for the demonstration of the truth of the premises. Even if one can in fact demonstrate the truth of the premises and keep these proofs in consciousness, when one goes back to thinking about the argument the demonstration of the truth of the premises will fade from consciousness. So the subject will have to rely on memory in order justifiably to use the premises in her argument. But this means that the demonstration of the reliability of memory depends crucially on memory being reliable. Hence, any demonstration of the reliability of



memory will in fact depend on the reliability of memory and hence be epistemically circular.

Naturally, having no good proof of reliability should not be taken as an indication of unreliability. Furthermore, internalists about justification believe that (even conditional) reliability is not necessary for justification. So the unreliability of memory would not sway the internalists into thinking that memory beliefs were unjustified (although it would, presumably, indicate that memory beliefs could not be counted as knowledge). What do internalists take to be crucial for the justification of memory beliefs? There are two main sorts of theories: those that take memory beliefs to be immediately justified and those who think their justification requires evidence. The first position holds this principle: If S seems to remember that p, then S is *prima facie* justified in believing that p (cf. Audi, 1995). Although there have been positions that make a memory image necessary, the aforementioned principle shouldn't be so read. It is sufficient for being *prima facie* justified that the subject *seem* to remember the belief. The other kind of internalism requires more: S's memory belief that p is *prima facie* justified only if S has some evidence that supposes p (cf. BonJour, 2002, Feldman and Conee, 1985). It is not thought to be necessary that S have direct evidence that p is true; it will suffice if S has evidence that memory is reliable in conditions like those relevant to the case at hand, and that p is a memory belief. To be plausible, such theories will not require that this argument be consciously entertained by the subject; it will be enough that the evidence is in her cognitive system (and perhaps that it could be accessed should the agent wonder about the justificatory status of her belief). A problem with each of these internalistic theories is that a belief can be unjustified when originally formed and then become justified simply on the basis of being retained. Suppose I know you to be an incurable liar and yet believe you when you tell me that p. At a later time, I recall that p but don't remember when or how I originally formed the belief. Both of the above varieties of internalism

will give the result that the belief is *prima facie* justified.

Against these internalist accounts are externalists who typically claim that, other things being equal, the justificatory status of a memory belief is largely a function of the epistemic status the belief had when originally formed (cf. Senor, 1994; Goldman 1999; for an argument against this view and responses, see Lackey, 2005; Senor, 2007; and Lackey 2007). The epistemic principle that captures this perspective is "Other things being equal, the epistemic status of a memory belief cannot be greater than the status the belief had when it was first formed" (this view has come to be called "preservationism"). On the preservationist theory, memory's job is to preserve belief and epistemic status; when it functions reliably, it does this; when it doesn't, it doesn't. In short, the preservationist view takes epistemic status to have a diachronic component, while those who deny preservationism (whether they require the memory belief to come with an image, to be evidentially supported by other beliefs at the time of recall, or to be *prima facie* justified) take epistemic status to be synchronic. That is, on the latter view what is relevant for the justification of a belief is the current psychological state of the subject.

*See also* CAUSAL THEORIES IN EPISTEMOLOGY; COMMONSENSISM AND CRITICAL COGNITIVISM; CRITERIA AND KNOWLEDGE; HISTORICAL KNOWLEDGE; PROBLEM OF THE CRITERION; TESTIMONY.

#### BIBLIOGRAPHY

- Aristotle: *De Memoria* 450a 13.  
 Audi, R: "Memorial Justification," *Philosophical Topics* 23 (1995), 31–45.  
 BonJour, L.: *Epistemology* (Lanham, MD: Rowman and Littlefield, 2002), ch. 8.  
 Feldman, R. and Conee E.: "Evidentialism," *Philosophical Studies* 48 (1985), 15–34.  
 Goldman, A. I.: "What Is Justified Belief?," in *Justification and Knowledge*, ed. George Pappas (Dordrecht: Reidel, 1979), 1–23.  
 Goldman, A. I.: "Internalism Exposed," *Journal of Philosophy* 96: 6 (1999), 271–93.  
 Hume, D.: *A Treatise of Human Nature*, 1.1.1.3.

- Lackey, J: "Memory as a Generative Process," *Philosophy and Phenomenological Research* 70: 3 (2005), 236–58.
- Lackey, J: "Why Memory Really Is a Generative Source: A Reply to Senor," *Philosophy and Phenomenological Research* 74 (2007), 209–19.
- Malcolm, N.: "Memory and the Past," in *Knowledge and Certainty*, ed. N. Malcolm (Englewood Cliffs, NJ: Prentice Hall, 1963).
- Martin, C. B. and Deutscher, M.: "Remembering," *Philosophical Review* 75 (1966), 161–96.
- Russell, B.: *The Analysis of Mind* (London: Macmillan, 1921).
- Senor, T: "Internalistic Foundationalism and the Justification of Memory Belief," *Synthese* (1994), 453–76.
- Senor, T: "Preserving Preservationism: A Reply to Lackey," *Philosophy and Phenomenological Research* 74: 1 (2007), 199–208.

TOM SENOR

### Merleau-Ponty, Maurice (1908–61)

Merleau-Ponty, French philosopher of the period immediately following World War II, is best-known in epistemology for his analyses of perceptual experience and of the interplay between perception and action, between perception of self and the perception of others, and between perceptual life, taken as a whole, and its various expressions and transformations in language, reflective thought, art, science, and philosophy.

Merleau-Ponty's theory of knowledge begins with a rejection of "the problem of knowledge" in what he takes to be the Cartesian sense. We do not have to respond to radical skepticism; we do not have to seek conclusive reasons to justify an inference beyond perception to knowledge of an "external" world. To think that we do is to demand an inappropriate sort of certainty in perception and to assume an unwarranted dichotomy between perception and world. Merleau-Ponty contends that perception is, in its own way, intrinsically cognitive. "We must not, therefore, wonder whether we really perceive a world, we must instead say: the world is what we perceive" (1962, p. xvi). Not, of course, if "perceive" is understood as the reception of sensory data or some form of explicit judgement based on sense-data; instead, we must

understand perception to be the way in which humans are already, in Heidegger's phrase, "in the world". Perception is precisely our "access" to the world. The task for a theory of knowledge, as Merleau-Ponty sees it, is to explicate the meaning and implications of taking perception in this sense.

In his major work, *The Phenomenology of Perception*, Merleau-Ponty's strategy is to try to refute empiricist and idealist (what he calls 'intellectualist') views of perception (see EMPIRICISM; IDEALISM) by evoking essential features of perceptual experience through the use of exemplary cases from ordinary life, from experimental psychology (centrally from the Gestaltists) and from studies of aphasia and agnosia in brain-injured patients. In this way, he falls broadly within the phenomenological tradition (see CONTINENTAL EPISTEMOLOGY; HUSSERL), but gives it a distinctive methodological direction (See his "Phenomenology and the Sciences of Man" in Edie, 1964). Substantively, his rejection of psychophysical dualism and his insistence equally on a sort of "realism" and on the centrality of the active, bodily and historical subject in all perceptual (and other cognitive) life, give his thought an existential character and a great affinity to pragmatism (see PRAGMATISM; JAMES). His epistemology forms part of a total philosophical anthropology and ontology, a philosophy of society and of history.

Among Merleau-Ponty's distinctive theses are the following: (1) Perception is a developed skill, a "knowing how", not a matter of forming explicit beliefs, in which we, as impersonal, species-specific, living organisms explore, through our sense-organs, an already significant environment; (2) Perceived things are disclosed as unities through perspectival variations – this is their "lived" and always contingent objectivity; (3) Spatial features of things, including depth and distance, are perceived directly (that is, non-inferentially) in reciprocity with bodily motility, and we see the possibilities which things afford for movement and manipulation; (4) Perceived things are "intersensory unities" too – that is, we see their tactile, auditory and other properties; (5) Our awareness of ourselves as individual selves develops after, and upon the basis of a

primordial awareness of the presence of others; (6) We are able directly to perceive the gestural and affective significance of the behavior of other persons; (7) Language and other cultural practices, although constructive and creative, are parasitic upon perceptual practice – they express features of the perceived world while they transform them; (8) Specifically, the sciences, as forms of knowing, constitute their understandings of the world by exercising and transforming the bodily (and instrumental) way in which we ordinarily explore it in perception; (9) Thus the objectivity of science is parasitic upon the objectivity of ordinary perception; and (10) the realities disclosed by well tested scientific theories can be no more real than, although they inform us in distinctive ways about, the things of ordinary perception.

See also EXPERIENCE; PERCEPTUAL KNOWLEDGE; SENSATION/COGNITION.

#### WRITINGS

*Phénoménologie de la Perception* (Paris: Gallimard, 1945); trans. C. Smith as *The Phenomenology of Perception* (New York: Humanities Press, 1962).  
*The Primacy of Perception* ed. J. M. Edie (Evanston, IL: Northwestern University Press, 1964).  
*Le Visible et l'Invisible* (Paris: Gallimard, 1964); trans. A. Lingis as *The Visible and the Invisible* (Evanston, IL: Northwestern University Press, 1968).

#### BIBLIOGRAPHY

Madison, G.: *The Phenomenology of Merleau-Ponty* (Athens, OH: Ohio University Press, 1981).  
 Pietersma, H.: "Merleau-Ponty's Theory of Knowledge," in *Merleau-Ponty: Critical Essays*, ed. H. Pietersma (Washington, DC: Center for Advanced Research in Phenomenology and University Press of America, 1989).

JOHN J. COMPTON

**methodology** A term occurring frequently in philosophy of science, but without any precisely established usage. It suggests

considerations closer to the actual practice of science (e.g. issues associated with statistical testing) than are some of the more general problems in philosophy of science (e.g. the instrumentalism/realism debate). Often it refers to features that differ from one field of science to another, as in the methodology of physics as compared to the methodology of sociology. In contrast to the logic of science, methodology may include heuristic considerations or other aspects of the context of discovery.

WESLEY C. SALMON

**Mill, John Stuart (1806–73)** British philosopher and economist. Mill's epistemology is thoroughly naturalistic (see NATURALISM). Human beings are entirely a part of the natural causal order studied by science. Like Kant, Mill thought this had consequences for knowledge: if minds are a part of nature then no knowledge of the world can be a priori. Grounds for any assertion which has real content must be empirical grounds. Mill thought knowledge remained possible on such a basis; Kant did not.

Kant's distinction between "analytic" and "synthetic" judgements (see ANALYTICITY) is paralleled in Mill's *System of Logic* (1843) by a distinction between "verbal" and "real propositions" and, correspondingly, between "merely apparent" and "real inferences". Verbal propositions have no genuine content; in a merely apparent inference no real inferential move has been made. The conclusion has literally been asserted in the premises.

Unlike Kant, Mill takes this strictly. He does not assume that a proposition derivable from logic by definitional substitutions alone is verbal. He develops an "analysis of language" – of the "import of propositions" – distinguishing between the "connotation" and "denotation" of terms ("names") and applying that distinction to the syntactic forms recognised by syllogistic logic. On this basis he demonstrates that mathematics, and pure logic itself, consist mainly of real propositions and inferences. He points out that logic and mathematics yield new knowledge. They could not do so if they consisted exclusively

of strictly verbal propositions; the syllogism would then be a *petitio principii*. Logical and mathematical knowledge is real, so it must be a posteriori, vindicated ultimately by induction; he takes this view even of the logical laws of contradiction and excluded middle.

This is the first thoroughly naturalistic analysis of deductive reasoning. Mill distinguishes it from “Conceptualism”, which confuses logic and psychology by assimilating propositions to judgements and attributes of objects to ideas, “Realism”, which holds that general terms signify abstract universals, and “Nominalism”, which fails to distinguish connotation from denotation and considers all logic and mathematics to be verbal. (In the contemporary sense however Mill himself is a nominalist: he does not countenance abstract entities, holding that number terms denote aggregates – natural entities – and connote their attributes.)

To accuse Mill of psychologism (*see* PSYCHOLOGISM) is an error; what is true however is that he wanted to explain the facts which lend colour to claims to a priori knowledge. They concern the limits to what we are able to cognise or imagine. Mill seeks to explain them in associationist terms (*see* ASSOCIATION) – not very convincingly. But that does not affect his essential point, which is that when the facts are conceived naturalistically, the step from our inability to represent to ourselves the negation of a proposition to acceptance of its truth calls for justification. Moreover, the justification itself must be a priori if it is to show that the proposition is known a priori; and that is what, on a naturalistic assumption, it cannot be. Mill is prepared to concede the *reliability* of, for example, geometrical “intuition”: but he stresses that its reliability is an empirical fact.

He recognizes only one basic form of inference – both epistemologically and genetically: enumerative induction, simple generalization from experience. But if enumerative inferences are real, must they not also be a posteriori? Mill agrees. “Principles of Evidence and Theories of Method are not to be constructed a priori. The laws of our rational faculty, like those of every other natural agency, are only learnt by seeing the agent

at work.” Like “Happiness is desirable”, “Enumerative induction is rational” is neither verbal nor an a priori intuition. All that Mill will say is that people agree in practice and theory in accepting it.

Logic is “the science of science”. It must accept, contrary to the “well-meant but impracticable precept” of Descartes, that any truly spontaneous form of reasoning has a *prima facie* claim to acceptance. Its task is to describe, codify and systematize. So Mill gives a natural history of the “inductive process” which shows how it starts by establishing local regularities, how induction on these established regularities leads to the conclusion that all events are subject to regularity, and how that conclusion in turn sustains his eliminative “Methods of Experimental Inquiry” – whose success stabilizes the whole structure. The problem of induction is to exhibit this cumulative and interactive process perspicuously and to explain why some inductions have greater weight than do other, formally similar, ones. “We have no ulterior test to which we subject experience in general, but we make experience its own test” – the coherentist element in this is often explicit (*see* COHERENTISM). The sceptical problem, as posed by Hume (*see* HUME), he totally ignores. It is a feature of his philosophy that he takes neither Cartesian nor Humean formulations of scepticism at all seriously (*see* PROBLEMS OF INDUCTION).

Unlike Whewell or Peirce (*see* PEIRCE), Mill rejects hypothetical reasoning as a means, in its own right, of achieving attested knowledge; though he affirms its heuristic value. He reasons that there can always be more than one hypothesis consistent with the data. The point is powerful, but this rejection substantially weakens his empiricist account of logic, arithmetic and geometry. It is also in tension, if not in contradiction, with his principle of accepting as *prima facie* sound what is spontaneously accepted in theory and practice. For hypotheses are central to commonsense and scientific reasoning, as Whewell (with whom Mill had a famous controversy on the subject) was to show.

The rejection of hypotheses produces a further tension in Mill’s naturalism when

combined, as Mill combined it, with the thesis that our immediate consciousness is of our own experience alone. For while enumerative induction can establish correlations within subjective experience (granting the epistemic credentials of memory – a point which troubled Mill) it cannot justify inferences beyond it. Thus Mill arrived at the conclusion that physical objects are knowable only as “Permanent [i.e. ‘certified’ or ‘guaranteed’] Possibilities of Sensation”. The reconcilability of this view with his overall naturalism is moot – though Mill saw no tension, arguing that any finding of natural science could be expressed in phenomenalist terms (see PHENOMENALISM). A conception of meaning and inference which might have removed the tension, by legitimizing hypotheses as proper methods of arriving at truth, is latent in Mill’s functional treatment of reasoning with general propositions, but it is not developed. Nevertheless, in this respect, and in others – such as its fallibilism (see FALLIBILISM) – the *System of Logic* represents a vital step on the road from an eighteenth-century to a pragmatist naturalism (see PRAGMATISM).

See also A PRIORI KNOWLEDGE in Part I; INDUCTION, ENUMERATIVE AND HYPOTHETICAL; INTUITION AND DEDUCTION; MATHEMATICAL KNOWLEDGE; PRAGMATISM.

#### WRITINGS

*An Examination of Sir William Hamilton’s Philosophy* (1865), in *Collected Works of J. S. Mill*, ed. J. M. Robson (London: Routledge, 1965–), vol. 9.  
*System of Logic* (1843), in his *Collected Works*, vols 7–8.

#### BIBLIOGRAPHY

Gillies, D.: *Frege, Dedekind and Peano on the Foundations of Arithmetic* (Assen: Van Gorcum, 1982).  
 Kessler, G.: “Frege, Mill, and the Foundations of Arithmetic,” *Journal of Philosophy* 76 (1980), 65–74.  
 Kitcher, P. S.: “Arithmetic for the Millian,” *Philosophical Studies* 37 (1980), 215–36.

Mackie, J. L.: *The Cement of the Universe* (Oxford: Clarendon Press, 1973), appendix.

Skorupski, J.: *J. S. Mill* (London: Routledge, 1979).

JOHN SKORUPSKI

**Molyneux’s problem** William Molyneux (1656–98) raised the following query in a letter to Locke:

Suppose a Man born blind, and now adult, and taught by his touch to distinguish between a Cube, and a Sphere of the same metal, and nighly of the same bigness, so as to tell, when he felt one and t’other, which is the Cube, which the Sphere. Suppose then the Cube and Sphere placed on a Table, and the Blind Man to be made to see. *Quære*, Whether by his sight, before he touch’d them, he could now distinguish, and tell, which is the Globe, which the Cube.

(quoted in Locke, *Essay* 2.9.8)

Locke agrees with Molyneux that the man would not be able to do this, since “he has not yet attained the Experience, that what affects his touch so or so, must affect his sight so or so; Or that a protuberant angle in the Cube, that pressed his hand unequally, shall appear to his eye, as it does in the Cube”.

At face value, the problem concerns the correspondence between the perceptual modalities. Among the questions raised are whether primary qualities (such as shape, size, weight, and motion) are perceived by more than one sense, and whether the visual and tactile perceptions “resemble” each other. But, as Molyneux expressed it, the main issue is whether we have or could have any *innate* visual capacity for distinguishing form, or whether perception presupposes an acquired (unconscious) interpretative skill on the part of the subject.

A different version of the problem (due to Diderot (*Lettre sur les aveugles*, 1749; as translated in Morgan, 1977, p. 108; and recently revived by Evans, 1985)) asks whether or not the blind man, upon acquiring vision, would be able to distinguish two-dimensional shapes, such as a *square* and *circle* (rather than three-dimensional shapes, such as *sphere* and *cube*). By modifying the terms in which the problem was originally



posed, the issue whether previous experience would be sufficient to trigger in Molyneux's man an innate visual capacity to appreciate the depth cues available in perception (including Necker cubes and similar line drawings) appears somewhat immaterial, since the disputants need not disagree about this.

Despite the different versions, the main question concerns the epistemological relation between perceptual modalities – in particular, the connection between a tangible perceptual representation of a primary quality and a visual perceptual representation of a primary quality. The problem is not whether the newly sighted man would be able to recognize visual shapes as distinct features of objects, but whether he would be able to extend the sortal or primary quality concepts previously acquired in order to identify tactually particular shapes, such as a square or a circle, to an application based on visual evidence alone; or, conversely, whether Molyneux's man would be able to extend the sortal or primary quality concepts acquired in the visual identification of particular shapes, to an application grounded exclusively in tactual experience. The crux of the dispute, then, is whether sortal or primary quality concepts have their "home" in any single one of the five sensory modalities, or association thereof; in effect, whether sortal or primary quality concepts are particular to the bearers of properties presented in any single sensory modality, or whether those concepts are used to discriminate objects which can be the bearers of properties presented in all or in some combination of the five sensory modalities.

There is a question whether, and to what degree, Molyneux's problem can be discussed independently of experimental evidence; and whether it can properly be said to admit of a philosophical solution. Although there is some experimental evidence in favour of Locke's negative reply (see Morgan, 1977; Gregory, 1974), the results are inconclusive, and may not bear unequivocally upon the issue.

See also BERKELEY; CONCEPTS; EMPIRICISM; LOCKE; PERCEPTUAL KNOWLEDGE; PRIMARY AND SECONDARY QUALITIES.

## BIBLIOGRAPHY

- Berkeley, G.: *An Essay towards a New Theory of Vision* (1709); in *Berkeley: Philosophical Works* ed. M. R. Ayers (London: Dent, 1975). esp. §§132–6.
- Evans, G.: "Molyneux's Question," in *Collected Papers* (Oxford: Oxford University Press, 1985), 364–99.
- Gregory, R. L. and Wallace, J. C.: "Recovery from Early Blindness – a Case Study," in R. L. Gregory, *Concepts and Mechanisms of Perception* (London: Duckworth, 1974), 65–129.
- Leibniz, G. W.: *New Essays on Human Understanding* (1704); trans. P. Remnant and J. Bennett (Cambridge: Cambridge University Press, 1981), esp. 2.10.
- Locke, J.: *An Essay Concerning Human Understanding* (1690), ed. P. M. Nidditch (Oxford: Clarendon Press, 1975), esp. 2.9.8.
- Mackie, J. L.: *Problems from Locke* (Oxford: Clarendon Press, 1976), esp. 29–32, 210–11.
- Mérian, J.-B.: *Sur le problème de Molyneux*, ed. F. Markovits (Paris: Flammarion, 1984).
- Morgan, M. J.: *Molyneux's Question* (Cambridge: Cambridge University Press, 1977).

STEVE SMITH

**Montaigne, Michel de (1533–92)**

French essayist. Montaigne was born near Bordeaux to a Catholic father and Spanish Jewish mother, who became a Calvinist. He studied at the College de Guyenne, and perhaps at the University of Toulouse. He held some political positions, including being Mayor of Bordeaux. He was a friend of the Reformation and the Counter-Reformation leaders, including Henri of Navarre who became King Henri IV.

Montaigne translated the rationalist theological treatise by Raimond Sebond. His "Apology for Raimond Sebond" (1576) is the most extended statement of his philosophy. It was composed after he had read the Greek sceptic Sextus Empiricus (see SEXTUS EMPIRICUS), and while he was undergoing a personal sceptical crisis, in which he found everything in doubt. In the "Apology" Montaigne presented and modernized the ancient sceptical arguments about the unreliability of information gained by the senses or by reason, about the inability of human beings to find a satisfactory criterion

of knowledge, and about the relativity of moral opinions. He suggested people should suspend judgement on all matters, and wait until God reveals principles to them. One should follow customs, traditions and social rules undogmatically, and should be tolerant of other views. Religious beliefs should be based on faith rather than doubtful evidence. Montaigne's rambling presentation of Pyrrhonian scepticism (see PYRRHONISM) quickly became the best-known statement of this view. Bacon, Descartes and Pascal, among others, were greatly influenced by Montaigne's scepticism. His advocacy of accepting customary views because there was no adequate reason to change them became a defence of Catholicism against the Reformation.

#### WRITINGS

"Apology for Raimond Sebond," in *The Complete Works of Montaigne* trans. D. M. Frame (Stanford, CA: Stanford University Press, 1958).

#### BIBLIOGRAPHY

- Frame, D. M.: *Montaigne: A Biography* (New York: Harcourt Brace, 1965).  
 Frame, D. M.: *Montaigne's Discovery of Man: The Humanization of a Humanist* (New York: Columbia University Press, 1955).

RICHARD H. POPKIN

#### **Moore, G(eorge) E(dward) (1873–1958)**

Moore was one of the most influential British philosophers of the first half of this century. His reputation in epistemology rests on his defence of commonsense and his point of view, as standardly expounded (cf. Stroud, 1984), is that of the plain man who insists that the validity of our everyday claims to knowledge is not open to serious question, so that philosophers who maintain the contrary can be confidently dismissed.

In truth, Moore's position was rather more complex. Where the standard account suggests

that he simply affirmed the existence of the types of knowledge denied by sceptics, Moore actually developed an intricate anti-sceptical dialectic. One strand of this is his argument from differential certainty (1953, VI, VII). Moore argues that we are entitled to be more certain of the falsity of the conclusions of sceptical arguments than we are of the truth of their premisses. For the premisses will always include general claims about the nature and limits of human knowledge, and the acceptability of such general claims is answerable to their implications concerning particular cases of knowledge; so the conclusions of sceptical arguments undermine our reasons for accepting their premisses. In presenting this argument Moore oversimplifies by suggesting that it just rests on a point about empirical induction; none the less, the argument can be reformulated in the context of an attempt to attain a reflective equilibrium concerning the limits of knowledge, and it brings out a real difficulty sceptics face in arguing for their own position. In my opinion the difficulty is not as decisive as Moore maintains, for sceptical arguments often combine apparently uncontentious theses concerning knowledge with theses from metaphysics and the philosophy of mind that we do not find it easy to abandon. None the less, even though Moore's argument only establishes a defeasible presumption against scepticism, it is not just the question-begging contraposition that the standard account of his position represents him as propounding.

Another strand in Moore's anti-sceptical dialectic is the charge that the sceptic's presentation of his position is incoherent since the sceptic cannot argue for his conclusion without representing himself as having knowledge of his premisses and of their implications. Moore's charge rests on the thesis that assertion essentially involves a claim to knowledge (1959, p. 248), which is the epistemic analogue of the principle underlying Moore's paradox (see MOORE'S PARADOX) and is suggested by the problematic nature of statements of the form "*p* but I do not know whether *p*". In my view, however, the possibility of making hedged assertions, of the form "*p* I think", which manifestly do

not involve a claim to knowledge, undermines this argument. For this possibility strongly suggests that the claims to knowledge implied by normal assertions are just conversational implicatures, in Grice's sense (*see* GRICE), which can be cancelled by hedging when the situation requires. Furthermore, even if Moore's thesis about assertion is accepted, the sceptic can just present his argument as a series of hedged assertions and thereby avoid the incoherence Moore imputes to him.

The writings in which Moore seems to live up to the standard account of him are his 1925 paper "A Defence of Common Sense" and his 1939 lecture "Proof of an External World" (both in Moore, 1959). For in the first paper he sets out, without any apparent argument, a list of commonsense truisms of which he maintains that he has certain knowledge; and in the second he famously maintains that it would be absurd to question his knowledge of such things as that his hands are before him. Yet I think that in both cases the standard account misunderstands Moore's purport. The "Defence" was originally written in response to an invitation to provide a "personal statement" of his philosophical position; so Moore's initial affirmations of common sense knowledge are just statements of his position, not attempts to establish its validity, though Moore does also attempt here to refute sceptical theses, which he attacks as incoherent. In the "Proof" Moore aimed to prove only the existence of an external world, not the existence of knowledge of such a world – i.e. to refute idealism (*see* IDEALISM), not scepticism. This distinction may seem tenuous: if Moore's proof succeeds, then, in giving it, does he not also prove the existence of his own knowledge of an external world? Yet that was not what Moore felt; in his view, a proof of knowledge that *p* requires the refutation of sceptical arguments, but such a refutation is neither required for, nor accomplished by, a proof that *p* itself. Admittedly this latter proof requires premisses that are in fact known, but Moore does not think he needs to prove the existence of this knowledge – it will suffice to remind his audience of the kinds of thing we all take

it for granted that we know (cf. Moore, 1942, pp. 668–9, for an unequivocal repudiation of the thought that sceptical arguments can be refuted by the dialectical strategy employed in the "Proof").

What, then, is Moore's contribution to epistemology? If he was not the philosopher's idealized plain man but just another plain philosopher unsuccessfully defending commonsense against sceptical arguments (cf. "Certainty" in Moore, 1959), wherein lies the value of his writings? In my view it does lie in his attempted defence of common sense, but that defence needs to be set in the context of a naturalistic epistemology (*see* NATURALIZED EPISTEMOLOGY) which Moore himself never adumbrated. In this context Moore's affirmations of certainty concerning particular matters of fact ("Here is one hand", etc.) signal that involuntary commitment to the existence of our perceptible environment on which a naturalistic epistemology relies in order to get itself started. And Moore's remarks about the "strangeness" of our epistemological situation vis-à-vis his commonsense truisms (1959, p. 44) reflect the fact that a naturalistic epistemology just incorporates our general presumptions about the structure of the world and our relationship to it rather than providing us with an independent method for verifying them. This last line of thought was famously developed by Wittgenstein in *On Certainty*, particularly in his remarks concerning "Moorean propositions" (which correspond to Moore's truisms) (*see* WITTGENSTEIN). So my judgement is that it was Wittgenstein who brought to fruition the potential of Moore's epistemology.

*See also* COMMONSENSISM AND CRITICAL COGNITIVISM; SCEPTICISM, CONTEMPORARY.

#### WRITINGS

*Philosophical Studies* (London: Routledge, 1922).  
 "A Reply to My Critics," in *The Philosophy of G. E. Moore*, ed. P. A., Schilpp, 3rd edn (LaSalle, IL: Open Court, 1968), 533–687.  
*Some Main Problems of Philosophy* (London: Allen and Unwin, 1953).

*Philosophical Papers* (London: Allen and Unwin, 1959).

## BIBLIOGRAPHY

- Baldwin, T.: *G. E. Moore* (London: Routledge, 1990).  
 Nuccetelli, S. and Seay, G., eds: *Themes from G. E. Moore. New Essays in Epistemology and Ethics* (Oxford: Oxford University Press, 2007).  
 Stroud, B.: *The Significance of Philosophical Scepticism* (Oxford: Clarendon Press, 1984).  
 Stroll, A.: *Moore and Wittgenstein* (New York: Oxford University Press, 1994).  
 Wittgenstein, L.: *On Certainty* (Oxford: Blackwell, 1969).

THOMAS BALDWIN

**Moore's paradox** Moore's problem is that of explaining the oddity of sentences such as

- (1) Uranus spins sideways but it is not the case that I believe it.
- (2) The Mediterranean Sea once dried up but I believe it did not.

Although (2) describes a commissive error and (1) only an error of omission, they both seem self-contradictory. Sensitivity to this aura of irrationality dates back to Spinoza (*see* SPINOZA):

If anyone says, then, that he has a clear and distinct, that is, a true idea of substance and nevertheless doubts whether such substance exists, he is like one who says he has a true idea and yet doubts whether it may not be false.

(*Ethics* 1p8s2)

Aesthetic appreciation of the tension is reflected in Samuel Beckett's *Waiting for Godot*. For it depicts two tramps as waiting for a man who they believe will never show.

Only within the last fifty years has the characterization of this tension come to be regarded a philosophical problem. Moore thought the absurdity of sentences such as (1) could be explained as a clash between what the speaker asserts and what he implies. The speaker *implies* that he believes Uranus spins

sideways because speakers almost always believe what they assert. Yet contrary to the expectation based on this regularity, the speaker *asserts* that he does not believe Uranus spins sideways. So we are pulled in one direction by the statistical syllogism and in another by the speaker's authority.

Moore's explanation fails to show why (1) is more paradoxical than other sentences that send mixed signals. When the village atheist sneers "Prayer works!", we readily resolve the conflict between what is said and how it was said; there is no paradox here. It is also doubtful whether one can *assert* (1). Of course, one can merely utter the words. But to perform the speech act of assertion, one must give the appearance of believing what one has said.

Some philosophers have tried to amplify this objection into a solution. They say that (1) is queer because it is a consistent but unassertible sentence. This invites a generalization to other speech acts. For example, "Never follow advice" is paradoxical because it cannot be used to advise people. Although this analysis picks out an interesting class of sentences, there are counter examples to the claim that Moorean sentences are unassertible. A Christian might assert,

- (3) The atheism of my mother's nieceless brother's only nephew angers God

because his favourite authority said so. Since this implies "God exists but I believe God does not exist", it is a hidden Moorean sentence.

This example also causes trouble for those who characterise Moorean sentences as consistent but *unbelievable*. For it forces them into a normative rather than a descriptive analysis of credibility. Jaakko Hintikka (*see* HINTIKKA) developed the normative approach by working out a "doxastic logic" that holds for ideal thinkers. Unlike ordinary people, they are perfectly consistent and believe all the logical consequences of what they believe. (More precisely, they believe in accordance with a doxastic interpretation of the modal system deontic S4.)

Critics have complained that Hintikka's system fails even as a normative model. For

in addition to being impossible to satisfy, it cannot even be rationally approximated. A person who works out the trivial consequences of his beliefs is squandering intellectual resources. Further, a certain amount of inconsistency seems mandatory. For any reasonable person believes:

- (4) I have at least one false belief.

My belief in (4) makes it impossible for all of my beliefs to be true, hence it ensures that my beliefs are indirectly inconsistent.

Our survey suggests that an adequate solution to Moore's problem must satisfy two main conditions. The first is a sorting task. All Moorean sentences must be counted as such even if they do not look absurd at first glance. (Surprisingly many proposals only manage to include the commissive variation of the paradox or only the omissive.) Conversely, phony Moorean sentences must be systematically excluded. Thus, (4) and the following must be ruled out.

- (5) Michael Jackson owned a chimp but I did not believe it.  
 (6) I believe octopuses attack people and I believe they don't.

The next requirement is a diagnosis of the air of personal inconsistency. Why is the self-ascription of a specific current error absurd? After all, it is all right to ascribe errors to others or even to a past or future self.

The explanation would gain extra merit if it also accounted for the resemblance between Moorean sentences and close look-alikes such as:

- (7) Nero wore a concave emerald in one eye but I do not know it.  
 (8) G. E. Moore was born in 1873 but I guess he was born in 1874.  
 (9) Pentagonal crystals are impossible but we intend to grow one.

Commentators have connected Moore's problem with the lottery paradox, the preface paradox, the surprise examination paradox, offshoots of Newcomb's problem, certain issues

of freewill, and epistemic semantics (see LOTTERY PARADOX; PREFACE PARADOX; SURPRISE EXAMINATION PARADOX). So it is likely that progress on Moore's problem will lead to advances on other puzzles.

See also EPISTEMIC VIRTUE; PARADOX OF THE KNOWER; PRINCIPLE OF CHARITY.

#### BIBLIOGRAPHY

- Hintikka, J.: *Knowledge and Belief* (Ithaca, NY: Cornell University Press, 1962).  
 Linsky, L.: "On Interpreting Doxastic Logic," *Journal of Philosophy* 65 (1968), 500–2.  
 Martinich, A. P.: "Conversational Maxims," *Philosophical Quarterly* 30 (1980), 215–28.  
 Sorensen, R.: *Blindspots* (Oxford: Clarendon Press, 1988).  
 Williams, J.: "Moore's Problem: One or Two?," *Analysis* 39 (1979), 141–2.

ROY A. SORENSEN

**moral epistemology** A moral realist takes it that it is a matter of fact whether an action is right or wrong, and a moral cognitivist takes it that there are moral beliefs about such facts. On these assumptions, all the ordinary epistemological questions arise, such as: (1) Is there a distinction between basic and non-basic moral belief? If so, which are which? (2) Is it ever possible to *know* a moral fact? If so, is such knowledge a priori or a posteriori? And, whichever it is, what more detailed story are we to tell about the way in which such knowledge is gained? (3) What is the structure (if any) of moral justification? (4) How are we to deal with the moral sceptic?

Within the cognitivist tradition, there are two main styles of approach to our first question. A *particularist* epistemology takes it that our basic moral beliefs are restricted to facts about the particular case before us; we may hope to move on from these later and come to grasp more general truths, laws or principles, but we have to start with the particular case. Philosophies of science are commonly *particularist* in this sense, for good reason. A



*generalist* epistemology holds that we first become aware of general truths which we are then able to apply to particular cases as they come along. We learn moral principles first, there being nothing in a particular case that one could extract a principle from; one then moves from the principles so learned to beliefs, or knowledge, about particular cases. This approach may be unconvincing elsewhere, but it has gained many adherents in moral epistemology.

This contrast between generalist and particularist epistemologies is only a contrast in direction. As stated, both admit that we are able to know both general principles and particular moral truths about what to do here; the only question is one of priority. There are, however, more trenchant versions of both approaches. There is a form of *ethical* (not epistemological) particularism which holds that moral principles cannot be known because there is none to know, and there are many types of view which hold that, even if we can know the general principles, our ability to find out what is right in a particular case is never reliable enough to count as knowledge. An example here is Moore's form of utilitarianism. Moore (1903) held that an action's rightness is a function of the goodness of its consequences, but that since even a simple action's consequences reverberate to the end of time we could never know enough about them to know that the balance will be of the right sort.

Rationalist moral theories of a traditional sort are generalist; they attempt to establish certain highly general moral rules as deliverances of reason and thus knowable a priori (see RATIONALISM). The classic example is Kant (1785). Kant's view is that a moral law is one that we can consistently will that all rational beings should use as a principle of action. More recently, Gewirth (1978) tries to show that the denial of a single central Principle of Generic Consistency is contradictory, and to derive other principles from that. Kant's rules are much more specific than Gewirth's. They are still general, but since they are more specific it is easier to use them to derive answers to the question what to do in a particular situation.

There are other, non-rationalist, ways of being a generalist in ethics. Sidgwick held that basic moral laws are not discovered by reason (e.g. by deriving contradictions from their falsehood, or by some other inference) but directly intuited (see INTUITION AND DEDUCTION). Robert Audi helpfully characterises intuitive knowledge of this sort as knowledge that is non-inferential, firm, pre-theoretical and based on a full understanding of the thing known. But this does little to tell us *how* we know the things we know intuitively; indeed, it does not appear that the intuitionists are going to be able to say anything positive on this topic at all. The rationalist programme, therefore, which speaks of reason rather than of intuition, has two considerable attractions: it gives a concrete account of how we know moral principles, and also if successful would produce principles which it would be irrational to deny. This would enable us to find an answer to the (mythical?) amoralist who wants to know what reason there is to do one's duty, and to the moral sceptic who doubts there are any duties in the first place.

I turn now to particularist approaches to moral epistemology. I start with the views of W. D. Ross. Ross held that what we learn first is that features of an act that we might perform make a difference to whether we should perform it or not. We learn by experience that tactfulness can be such a difference, in the sense that we simply recognise on some occasion that the fact that this act would not be tactful calls for us not to do it; and the same applies to every other morally relevant property; particular cases reveal to us that the fact that we would be letting someone down, or failing to meet some commitment, counts against the action concerned. Given this sort of entirely particular knowledge, however, we can immediately move to something general. For we recognize by a process called *intuitive induction* that what makes a difference here must make the same difference wherever it occurs. So we can learn the truth of a moral principle from what we can see in the particular case; the principle is self-evident to us, given what the present case has taught us, since nothing more is necessary to reveal it to us than what the present case contains.

Intuitive induction is perhaps an unfamiliar process (*but see* Johnson, 1922). Its use is not, however, confined to ethics. An example Ross gives is that of discovering the validity of a principle of inference (*modus ponens*, say) by seeing it in the validity of one instance. In seeing the validity of the inference “If he is here he’ll see her: he is here: so he will see her” we see (or at least *can* see) the validity of its general principle “If *p* then *q*: *p*: so *q*”.

But for Ross the general principle that we learn is not of the form “All tactless actions are wrong”; it says, rather, that tactless actions are *prima facie* wrong – wrong so far as that goes. There might be several such principles applying to the same case. How, then, do we come to decide what we actually ought to do in the case before us? All that we have noticed so far is that certain features count in favour of this course of action and others count against. How are we to judge how we ought to act, all things considered? Ross’s answer here is just that we should consider the various pros and cons and come to the best view we can. But he holds that the result of this process is never knowledge, since the whole thing is far too insecure ever to deserve that title. We can know, perhaps, that this action is *prima facie* wrong in virtue of its deceitfulness, and so by intuitive induction that all deceitful actions are *prima facie* wrong. But we can never know that overall we should not do this deceitful action; our decision on this point will never be more than “probable opinion”.

For Ross, therefore, the starting point is recognition, of some feature, that it goes to make this action wrong (or right, of course). How do we know this? Ross tells us that we know the various *prima facie* principles *a priori*, and he tells us that we know them by extracting them from what is given in particular cases; but he does not tell us how we know the things he takes to be morally basic.

Christopher Peacocke does better on this front. He offers at least some explanation of how we acquire knowledge of general principles, though in doing so he abandons the particularist elements of Ross’s account. *Prima facie* moral principles are *a priori*, he

says, because they are the products of understanding.

Consider your belief that *prima facie* it is good if the institutions in a society are just; or your belief that *prima facie* it is wrong to cause avoidable suffering; or that *prima facie* trials should be governed by fair procedures. . . . Understanding of what justice is, of what suffering is, of what a trial and what fairness is makes these several beliefs rational without justificational reliance on empirical experience. (2004: 200)

Peacocke’s more official accounts of the *a priori* would lead us to say that the justification for these beliefs does not appeal to any perceptual or other cognitive states of the knower. So here we have a Ross-style ethic of *prima facie* principles *plus* a more rationalist account of how we gain knowledge of those principles. We know them by the use of reason (and understanding) rather than by extracting them from what is given in a particular case. So Peacocke’s account is a Rossian form of rationalist generalism.

Those still attracted to particularism might look to the “moral sense” theorists (*see* Hutcheson, 1725; and Broad, 1971). This tradition claims that singular moral judgements are analogous in certain important respects to judgements of sense perception. The analogy depends upon taking a dispositional view of colour, so that the yellowness of this banana is a disposition it has to cause certain experiences in suitably placed observers. Similarly, it is held that the wrongness of an action is a disposition the action has to cause feelings or emotions of a distinctive sort in a suitable observer. In both cases, this analysis is supposed to be compatible with the view that the relevant property can be directly observed. We can effectively *discern* these dispositions instantiated in the particular case.

This moral sense theory seems initially more promising than Ross’s position, but it does of course have some disadvantages. First, one has to accept a dispositional analysis both of colour and of moral qualities. Second, something has to be said about how general moral truths can be discovered by this method, it

being admitted in this tradition that there are such things to be discovered at all. It appears to be inevitable that the model will still be that of our knowledge of colour. We know that bananas are yellow by ordinary enumerative induction (*see* INDUCTION, ENUMERATIVE AND HYPOTHETICAL), not by intuitive induction *à la* Ross, and presumably we are to find out that deceitful actions are wrong in the same sort of way. This one is wrong, that one is wrong, and so with increasing probability we conclude that all are wrong. This is a possible view, but it seems to me mistaken. Knowing the truth of the general principle that deceit is wrong, we know not just that all (or most) deceitful acts are wrong, but also that they are wrong partly because of being deceitful. That is, there is a connection revealed in the moral case that is not revealed in colour perception. Ordinary enumerative induction will not be able to reveal such facts about connections unless it has facts of that sort as input, namely that this action is wrong partly because of its deceitfulness, so is that one, so is yet another, and so (probably) all deceitful acts are wrong, and wrong because of their deceitfulness.

The question now is whether the moral sense theory will be able to give a good sense to an initial awareness that this action is wrong partly because of its deceitfulness. There appears to be no analogy between the moral case and the colour case on this point, and this is a definite advantage for Ross. He is able to give a strong sense to the idea that moral properties such as rightness and wrongness stem from other properties or features of acts. So, according to Ross, when acts are right or wrong, they get this moral status *by virtue of* exemplifying some further properties, and very commonly (to say the least) our awareness of right and wrong is mediated by our awareness of these other properties. There is nothing like this in our awareness of colour. However, an analogy may be significant without being perfect; all analogies give out at some point. And we can surely still give a good sense, within the constraints of a dispositionalist approach, to the idea that our moral perception or response is tied to the properties that ground the

rightness/wrongness here. The dispositional story will just be that we experience this action as disposed to cause suitable observers to respond to its deceitfulness (or to it in the light of its deceitfulness) in a certain characteristic way. Here the “grounding property”, the deceit, is explicitly present in the judgement and related to the wrongness, that is, to the disposition to elicit disapproval.

These are the two main forms of particularist epistemology in ethics.

I now turn to the question which of our moral views are justified. One approach here is to debate the rival merits of foundationalism and coherentism in their application to justification in ethics (*see* FOUNDATIONALISM; COHERENTISM). (I don't think that reliabilism (*see* RELIABILISM) recommends itself in this area.) The first thing to say is that what, e.g., a foundationalist position will amount to will vary radically according to the substantive moral epistemology we adopt. Both rationalist and intuitionist forms of generalism will suppose that there are basic, highly general truths from which particular moral prescriptions can be derived. This position is unlike standard forms of foundationalism in the theory of empirical knowledge, since it makes no pretence to ground moral knowledge on particular facts of moral experience. But it is none the worse for that; it still retains the characteristic foundationalist distinctions between basis and superstructure and between two forms of justification, mediate and immediate, and the foundationalist insistence that justification is all one-way. The difficulty I see for these views is that it is going to be impossible to derive from what is allowed at the basis any detailed and specific suggestion about our moral duties in a particular case. This is a complaint often made against Kant, but it seems to infect all forms of generalist foundationalism.

Much more promising is some form of coherentism, which refuses to draw the standard foundationalist distinctions. It allows that moral justification grows as we achieve a good fit between different aspects of our moral views, those more general and those more specific. Coherentism of this sort is hardly distinct from the view associated with John

Rawls, which holds that what we seek on the epistemological front is a form of rational equilibrium in which we have achieved maximal resolution of the tensions between different views with independent appeal to us. To the extent that coherentists accept this aim, however, they also hope to offer an explanation of why it is a reasonable aim.

Particularist approaches, by contrast, appear to lend themselves much more easily to foundationalist structuring, if only because they take the view that some of our moral beliefs are justified in a way that does not depend on the (moral) justification of anything else. This may give us hope that we are able to acquire some justification in the particular case which is solid enough to support the rest, without falling into the usual forms of circularity or regress. The picture that Ross offers, however, falls far short of this, in two ways. The first is that he had really nothing to say about which particular moral judgements at the *prima facie* level are justified and which are not. The second is that he says nothing about which decisions about our overall duties are justified and which are not. In this, we could say that he fails to offer a theory of moral justification, if by that we mean a theory of the justification of moral belief. And at the moment there is nothing better on the table.

A different approach to the nature of moral justification would be to use the ideas of Wittgenstein (*see* WITTGENSTEIN). We could see certain moral beliefs as “frame” beliefs which play in ethics the role played in ordinary perceptual justification by such beliefs as “I have two hands” and “The sun is a very long way away”. These “frame” beliefs are not justified but stand in no need of justification; they are the things we appeal to in the justification of other beliefs. Candidates would be “All have equal rights” and “One should not torture innocent children”. This would give us something of the structure of foundationalism but from a completely different perspective.

Even if we succeed in giving an adequate account of justification in ethics, this will not in itself yield an account of moral knowledge (unless some version of the tripartite

definition is correct (*see* TRIPARTITE DEFINITION OF KNOWLEDGE). Nor would it alone provide an answer to the moral sceptic. There are in fact two forms of moral sceptic to be considered. The first maintains that there are no moral facts to be known (or justifiably believed). The second maintains only that, though there are or at least may be such facts, we are incapable of coming to know them (and perhaps also of having justified beliefs about them). The former should not really be called a sceptic at all; the right term would be “nihilist”; for sceptics about the future do not deny that there will be a future, but only that we are capable of finding out about it in advance. Considering then the moral sceptic proper, our first question should be whether there are special reasons for scepticism about moral knowledge and justification, or whether what we have here is merely the re-application of standard sceptical arguments. If it is the latter, those of us who don’t know how to answer the standard sceptical arguments anyway should feel no less secure in their moral judgements than they do in any other. It may be, however, that there are special reasons for moral scepticism that do not apply to other areas. (These will not include the fact that moral properties are unobservable, of course.) Suppose, for example, that moral disagreements persist in ways that defy our attempts to explain them as the effects of defects in one or the other party; neither side can be shown to be biased, careless, misinformed or hasty in judgement. And suppose that we cannot say the same of other disagreement on other topics. This might be enough to show that moral scepticism has a solid ground that other forms of scepticism cannot claim.

*See also* COMMONSENSISM AND CRITICAL COGNITIVISM; PROBLEM OF THE CRITERION.

#### BIBLIOGRAPHY

- Audi, R.: “Intuitionism, Pluralism and the Foundations of Ethics,” in his *Moral Knowledge and Ethical Character* (Oxford: Oxford University Press, 1997).

- Broad, C. D.: "Some Reflections on Moral-sense Theories in Ethics," in *Broad's Critical Essays in Moral Philosophy*, ed. D. Cheney (London: Allen and Unwin, 1971).
- Gewirth, A.: *Reason and Morality* (Chicago, IL: Chicago University Press, 1978).
- Hutcheson, F.: *An Inquiry Concerning the Original of Our Ideas of Virtue or Moral Good* (1725), in *British Moralists*, ed. L. A. Selby-Bigge (Indianapolis, IN: Bobbs-Merrill, 1964).
- Johnson, W. E.: *Logic* (Cambridge: Cambridge University Press, 1922), pt 2, ch. 7.
- Kant, I.: *Groundwork of the Metaphysic of Morals* (1785); trans. J. Paton as *The Moral Law* (London: Hutchinson, 1972).
- Moore, G. E.: *Principia Ethica* (Cambridge: Cambridge University Press, 1903).
- Peacocke, C.: *The Realm of Reason* (Oxford: Oxford University Press, 2004).
- Rawls, J.: "Outline of a Decision Procedure in Ethics," *The Philosophical Review* 60 (1951), 177–97.
- Ross, W. D.: *The Right and the Good* (Oxford: Clarendon Press, 1930).
- Sidgwick, H.: *The Methods of Ethics* (London: Macmillan, 1874).
- Wittgenstein, L.: *On Certainty* (Oxford: Blackwell, 1969).

JONATHAN DANCY

**myth of the given** see FOUNDATIONALISM; the GIVEN; SELLARS.



**naïve realism** *see* ARGUMENT FROM ILLUSION; DIRECT REALISM; PROBLEM OF THE EXTERNAL WORLD.

**natural science, epistemology of** The natural sciences (physics, chemistry, astronomy, geology, biology) are of special importance to epistemology because collectively they contain the most extensive, systematic, reliable knowledge that we have. Prior to Newton, Euclidean geometry played that role, and its influence on the history of epistemology was vast (*see* GEOMETRY).

Before the middle of the twentieth century, many philosophers and scientists (e.g. Ernst Mach, Bertrand Russell (*see* RUSSELL)) sought a secure foundation for scientific knowledge in sense-data (*see* SENSE-DATA); the more recent fashion is to take perceptions of middle sized physical objects as fundamental data for natural science. Although it had been argued that sense-data furnish a basis in certainty (*see* CERTAINTY), whereas perceptions of material objects may not be veridical, most philosophers despair of logically constructing material objects from sense-data (*see* LOGICAL CONSTRUCTION; PHENOMENALISM). Following Karl Popper (1959) and Hans Reichenbach (1938), they accept a physicalistic basis and acknowledge its corrigibility. The quest for certainty that had motivated phenomenalism has been largely abandoned. Errors of perception are considered, on the whole, detectable and correctable.

Granting the physicalistic basis, several epistemological problems concerning natural science arise. The first of these is David Hume's (*see* HUME) notorious problem of the justification of induction (*see* PROBLEMS OF INDUCTION). Natural science provides knowledge of events that have not yet occurred

(e.g. a future solar eclipse), events that happened in the remote past (e.g. the extinction of dinosaurs), and events that are happening elsewhere, unexperienced by any human observer (e.g. occurrences on the surface of Venus). This raises the problem of inferences from the observed to the unobserved, or more generally, the problem of ampliative inferences – arguments whose conclusions have factual content not present in their premises. Hume argued persuasively that, without some assumption regarding the uniformity of nature, we have no rational basis for concluding that any of our predictive inferences, even though based on true premises, will ever again have true conclusions. Moreover, he argued, it is impossible to establish the uniformity of nature by arguments a priori or a posteriori. His conclusion was not merely that knowledge of the unobserved is fallible – a point that had been recognized in antiquity by the sceptics – but that it cannot be considered even probable. Although Hume seemed to direct his attack chiefly at induction by simple enumeration, his argument is cogent with respect to any form of ampliative inference.

A wide variety of responses have been made to Hume's inductive scepticism. Immanuel Kant (*see* KANT), who claimed to have been awakened from his dogmatic slumbers by Hume's work, offered a transcendental deduction of a principle of universal causation. P. F. Strawson (1952) (*see* STRAWSON) offered an ordinary language dissolution, arguing that Hume's problem of induction is a pseudo-problem. Russell (1948) offered a set of "postulates of scientific inference". Rudolf Carnap (1963) appealed to "inductive intuition" as the ultimate source of justification. In an approach rather similar to Carnap's, Nelson Goodman (1955)

(see GOODMAN) attempted to replace Hume's old problem of induction with his "new problem of induction" (see PROBLEMS OF INDUCTION). Popper (1935) (see POPPER) sought to evade the problem by advocating deductivism, the view that science does not employ induction at all. Reichenbach (1938, 1949) (see REICHENBACH) offered a pragmatic justification. All of these approaches seem beset by fundamental difficulties (Salmon, 1967).

Natural science, in addition to making inferences regarding particular events and facts, is also concerned to establish general laws of nature (e.g. conservation of momentum) that are presumed to hold at all times and places in the entire history of the universe. Clearly, Hume's problem of induction applies to this enterprise; there is no way to establish such laws a priori, and any a posteriori method will involve ampliative inferences. This difficulty notwithstanding, various philosophers have addressed the problem of knowledge of laws. A traditional approach is the hypothetico-deductive (H-D) method. Given a hypothesis *H* that is to be tested, one deduces from *H* (in conjunction with suitable initial conditions) an observational prediction *O*. If *O* turns out to be true, the hypothesis *H* is said to be confirmed to some extent. One of many major problems with the H-D method is that if *H* is thus confirmed, so is *H*&*X*, where *X* is any arbitrary statement (see also HEMPEL'S PARADOX).

If, in the foregoing H-D schema, *O* turns out to be false, one can immediately conclude by *modus tollens* that at least one of the premises is false. If one has sufficient confidence in the truth of the initial conditions, one can conclude that *H* is refuted. Rejecting the positive form of the H-D method, Popper advocates the method of "conjectures and refutations". The method of science, he claims, is to advance bold explanatory hypotheses (conjectures) and to subject them to the most severe experiential tests in a sincere effort to refute them. Hypotheses that survive such severe tests are said to be "corroborated". It is important to emphasize that "corroboration" is not just another word for "confirmation". Confirmation enhances the probability of the hypothesis

in question; hypotheses that are highly confirmed have high probability. Hypotheses that are highly corroborated are very improbable according to Popper. The fundamental problem with Popper's approach is that, in banishing induction from science, he also bans all predictive power. From the content of our observations alone we cannot deduce anything about future occurrences (Salmon, 1981).

Pierre Duhem (1954) pointed out that, in addition to the hypothesis being tested and statements of initial conditions under which the test is conducted, we also need auxiliary hypotheses to carry out the deduction of observational consequences in the H-D schema. When the observational prediction turns out to be false, we are not entitled to conclude that the test hypothesis is false; at best, we can assert that the conjunction of the test hypothesis and the auxiliaries has been refuted. These and other considerations have led some philosophers, most notably Quine (1951) (see QUINE), to maintain the holistic thesis (see HOLISM) that individual scientific hypotheses are not subject to separate tests, but rather, the entire fabric of scientific knowledge confronts empirical evidence as a whole.

Many philosophers have turned to the theory of probability as a way of understanding the confirmation of scientific hypotheses (see PROBABILITY). The most extensive and systematic example is Carnap's inductive logic (see CARNAP), the major features of which are given in Carnap (1950, 1952). Ironically, in this system, universal generalizations that range over infinite domains always have degree of confirmation zero on any finite body of evidence. One has to be satisfied with "qualified instance confirmation", the non-zero confirmation of the statement that the generalization will hold in the next instance encountered. This is, however, a technical difficulty that can be overcome (Hintikka, 1966). A far more serious difficulty – one of principle – is the fact that a priori measures of prior probabilities, which appear to be altogether arbitrary, are required (Salmon, 1967).

As many philosophers view the situation, Bayes's theorem holds the key to the

confirmation of scientific hypotheses. It may be written

$$\frac{\Pr(H/E\&B)}{\Pr(H/B) \Pr(E/H\&B) / \Pr(H/B) \Pr(E/H\&B)} \\ + \Pr(-H/B) \Pr(E/-H\&B)$$

where *H* is the hypothesis, *B* our background knowledge (including initial conditions and auxiliary hypotheses, if any), *E* is the specific evidence being brought to bear on *H* and  $\Pr(H/E\&B)$  is the probability of *H* on evidence *E* given background *B*. If a hypothetico-deductive test yielding a positive outcome has been conducted,  $\Pr(E/H\&B) = 1$ .  $\Pr(H/B)$  and  $\Pr(-H/B)$  are known as prior probabilities; their status has been the subject of much controversy. Personalists, who are often referred to as “Bayesians”, take them to be subjective degrees of belief. Their critics object to such an infusion of subjectivity into the logic of science. Carnap’s above-mentioned inductive logic also rests on Bayes’s theorem; in that system the prior probabilities are established a priori. Holders of frequency or propensity interpretations of probability are hard put to provide a reasonable interpretation of prior probabilities (Salmon, 1967).

Up to this point, we have considered problems associated with knowledge of unobserved matters of fact, but natural science also poses problems regarding knowledge of unobservable entities. At the turn of the twentieth century, many philosophers and scientists – e.g. Mach, Karl Pearson, and some early logical positivists (*see* LOGICAL POSITIVISM) – denied the existence of such objects as molecules and atoms. Taking an instrumentalist position, they held, for example, that the molecular kinetic theory of gases, which seems to make reference to molecules, is merely a useful tool for the organization of experience, but does not establish the existence of such things. They are nothing more than useful fictions. In recent years, Bas van Fraassen (1980) has advocated a position he calls “constructive empiricism”, which is not a form of instrumentalism, but takes an agnostic attitude toward unobservable entities.

A good deal of the philosophical discussion of this issue has focused on the meaning of theoretical terms and statements. Operationists have maintained that all meaningful scientific concepts are operationally definable in terms of physical operations that can be carried out in the laboratory or in the field and “pencil and paper” operations (calculations or mathematical derivations), while logical positivists have identified the meaning of a statement with the means of its conclusive verification. Operationists and logical positivists thus deny that we can have knowledge of unobservable entities, for discourse putatively about them is strictly meaningless. Carnap (1956) makes a distinction between internal and external questions of existence (*see* EXISTENCE). If one has adopted the standard language of physics and chemistry, one can establish the internal existence of atoms and molecules by appeal to the standard theories. The question of whether to adopt a linguistic framework that incorporates terms that refer to atoms and molecules is an external question; it should be answered with regard to the utility of the theory, not by appeal to metaphysical arguments about the “real existence” of these entities. An excellent survey of the issues regarding the meaning of scientific theories can be found in Carl G. Hempel (1958).

The emphasis on meaning of theoretical terms seems, however, somewhat misplaced. The basic epistemological question is whether we can have knowledge of unobservable entities. We can, after all, make statements about unobservable entities without invoking a special theoretical vocabulary. In his work on Brownian movement, Jean Perrin created large numbers of tiny spheres of gamboge (a yellow resinous substance). He suspended them in water, observed their motions with a microscope, and inferred that many smaller particles were colliding with them. Without using any non-observational terms, I have just described the essentials of an epoch-making experiment on the reality of molecules, namely, the ascertainment of Avogadro’s number (the number of molecules in a mole of any substance).

The foregoing example suggests that we can make at least a rough and ready tripartite division among objects that are directly perceivable by normal human senses, those that are indirectly perceivable through the use of such instrumental extensions of our senses as microscopes and telescopes, and those whose existence and properties can be inferred on the basis of direct and indirect observation. To avoid begging the epistemological questions, we must allow for the possibility that either the second or the third, or both categories, are empty. In adopting a physicalist approach from the outset, we assured the nonemptiness of the first category.

Ian Hacking (1981) has addressed the question of microscopic observations with unprecedented sophistication, and has offered powerful arguments for their veridicality. The veridicality of telescopic observation can be supported by considerations at least as strong. It would seem that the category of indirectly observable entities is not empty. The problem of the third category still remains. It would be unwarranted, at this stage in the history of science at least, to claim that we see the quarks within a proton by means of accelerator experiments or the interior of the sun by means of neutrino detectors.

The problem of the existence of entities not even indirectly observable was essentially settled for natural science as a result of the work of Perrin and others in roughly the first decade of the twentieth century. It is well recounted in Perrin's semi-popular (1923) and Mary Jo Nye elaborates the development in fine historical detail in her (1972). The crux of the argument is this. From a series of physical experiments that are superficially extremely diverse, it is possible to infer the value of Avogadro's number, and the values obtained in all of these types of experiment agree with one another remarkably well. If matter were not actually composed of such micro-entities as molecules, atoms, ions, electrons, etc., this agreement would be an unbelievably improbable coincidence. My analysis of this argument is in terms of common causes (Salmon, 1985), but even if it is analysed in other ways, the argument is exceedingly robust.

Recent thought about the epistemology of natural science has been strongly influenced by the work of Thomas S. Kuhn (1962), who has emphasized the importance of historical considerations for purposes of philosophical analysis. Instead of viewing the development of the natural sciences as a process of gradual accumulation of objective knowledge of the physical world, he sees it as a series of episodes of normal science alternating with periods of scientific revolution. Each episode of normal science is characterized by a paradigm, consisting of a problem domain, a standard set of problem-solving techniques, general principles, and theories. When a problem in that domain is recalcitrant with respect to the available techniques it constitutes an anomaly. Too many anomalies constitute a crisis for that paradigm, and a revolution may occur, in which a new paradigm replaces the older one. When this replacement has taken place, a new period of normal science ensues. The new paradigm, according to Kuhn, does not incorporate the old, but rather, is incommensurable with it. For example, instead of solving the problems that were anomalies for the old paradigm, the new paradigm may simply dismiss them as inconsequential. To change from one paradigm to another requires something like a Gestalt-shift.

According to Kuhn, the choice between theories represented by the old and new paradigms is not determined solely by observed facts and logic; instead, it involves persuasion and judgment as well. The transition from one to another is a social fact accomplished by a community of scientists. There is no single point in the transition at which it is irrational to hold on to the old theory rather than switching to the new. When, on the basis of such comments, various philosophers interpreted Kuhn as denying the objectivity of science, he vehemently denied the charge. On his view, mature physical science is the best example of objective knowledge we have. To understand what objective knowledge consists in we should not lay down formal criteria *a priori*, but rather, we should examine the methodology of physical science. When we do, he says, we see that theories are evaluated on the basis of such criteria as

simplicity, consistency, scope, accuracy and fertility. This list is neither exclusive nor exhaustive, and there cannot be precise rules for application of its members.

I am inclined to think that Kuhn's view that theory choice goes beyond empirical data and logic is based on an overly narrow conception of the logic of science. If we think of scientific confirmation in terms of Bayes's theorem, we recognize that it not only allows for, but actually demands, prior probabilities of the theories under consideration. Such prior probabilities are plausibility judgements. At least some of the criteria Kuhn mentions seem clearly to be qualitative evaluations of prior probabilities. A Bayesian treatment of scientific confirmation appears to bridge the gap, at least to a significant extent, between historically oriented analyses of knowledge in the natural sciences and more traditional approaches (Salmon, 1990).

See also BAYESIAN EPISTEMOLOGY in Part I; EXPLANATION; PROBABILITY; PROBLEMS OF INDUCTION; SOCIAL SCIENCES; SOCIOLOGY OF KNOWLEDGE.

## BIBLIOGRAPHY

- Carnap, R.: *Logical Foundations of Probability* (Chicago, IL: University of Chicago Press, 1950).
- Carnap, R.: "Empiricism, Semantics, and Ontology," in his *Meaning and Necessity*, 2nd edn (Chicago, IL: University of Chicago Press, 1956), 205–21.
- Carnap, R.: *The Continuum of Inductive Methods* (Chicago, IL: University of Chicago Press, 1952).
- Carnap, R.: "Replies and Systematic Expositions," in *The Philosophy of Rudolf Carnap*, ed. P. A. Schilpp (LaSalle, IL: Open Court, 1963), 859–1016.
- Duhem, P.: *The Aim and Structure of Physical Theory*, trans. P. Wiener (Princeton, NJ: Princeton University Press, 1954).
- Goodman, N.: *Fact, Fiction, and Forecast* (Cambridge, MA: Harvard University Press, 1955).
- Hacking, I.: "Do We See through a Microscope?," *Pacific Philosophical Quarterly* 62 (1981), 305–22.
- Hempel, C. G.: "The Theoretician's Dilemma," *Minnesota Studies in the Philosophy of Science*, vol. 2, ed. H. Feigl, M. Scriven and G. Maxwell (Minneapolis, MN: University of Minnesota Press, 1958), 37–98.
- Hintikka, J.: "A Two-dimensional Continuum of Inductive Methods," in *Aspects of Inductive Logic*, ed. J. Hintikka and P. Suppes (Amsterdam: North-Holland, 1966), 113–32.
- Kuhn, T. S.: *The Structure of Scientific Revolutions* (Chicago, IL: University of Chicago Press, 1962).
- Nye, M. J.: *Molecular Reality* (London: Macdonald, 1972).
- Perrin, J.: *Les Atomes* (Paris: 1913); trans. D. L. Hammick as *Atoms* (New York: Van Nostrand, 1923).
- Popper, K. R.: *Logik der Forschung* (Vienna: 1935); trans. the author as *The Logic of Scientific Discovery* (New York: Basic Books, 1959).
- Quine, W. V.: "Two Dogmas of Empiricism," *Philosophical Review* 60 (1951), 20–43.
- Reichenbach, H.: *Experience and Prediction* (Chicago, IL: University of Chicago Press, 1938).
- Reichenbach, H.: *The Theory of Probability* (Berkeley, CA: University of California Press, 1949).
- Russell, B.: *Human Knowledge, Its Scope and Limits* (New York: Simon and Schuster, 1948).
- Salmon, W. C.: *The Foundations of Scientific Inference* (Pittsburgh, PA: University of Pittsburgh Press, 1967).
- Salmon, W. C.: "Rational prediction," *British Journal for the Philosophy of Science* 32 (1981), 115–25.
- Salmon, W. C.: "Empiricism: The Key Question," in *The Heritage of Logical Positivism*, ed. N. Rescher (Lanham, MD: University Press of America, 1985), 1–21.
- Salmon, W. C.: "Rationality and Objectivity in Science, or Tom Kuhn Meets Tom Bayes," *Minnesota Studies in the Philosophy of Science*, vol. 14, ed. C. Wade Savage (Minneapolis, MN: University of Minnesota Press, 1990), 175–204.
- Strawson, P. F.: *Introduction to Logical Theory* (New York: John Wiley, 1952).
- Van Fraassen, B. C.: *The Scientific Image* (Oxford: Clarendon Press, 1980).

WESLEY C. SALMON

**naturalism** This is the doctrine that there are only natural things: only natural particulars and only natural properties. It is a close relative of the doctrine of materialism or physicalism according to which there are only material or physical things. If there is a difference between materialism and physicalism, it is that the materialist takes the category of the material to be given intuitively, the physicalist takes it to be given by an idealization from contemporary physics. A similar



difference affects naturalists. Some take the category of the natural to be given intuitively whereas others – in particular, most contemporary proponents – take it to be given by an idealisation from the natural sciences.

As with physicalism, the science-based version of naturalism is beset with two ambiguities. First, an ontological ambiguity. In holding that there are only natural things, it may take natural things to be, or to be definable in terms of, just the sorts of particulars and properties countenanced in the (idealized) natural sciences; this doctrine is often described as reductionism. Alternatively, it may take natural things to be a broader class: say, to be the things that supervene appropriately on what is countenanced in the natural sciences (see EPISTEMIC SUPERVENIENCE). Second, a methodological ambiguity. In holding that there are only natural things, naturalism may suggest that the natural sciences offer the unique model for discovering what there is – this view is sometimes described as scientism – or it may be open to the possibility that other approaches – say, those of commonsense or those of the humanities – may also tell us about what there is.

Reductionism and scientism often go together. A reductionist, scientific naturalism will tend to deny reality to many commonsense posits: say, to colours and secondary properties, to free will and related psychological capacities, or to moral and other values. It will constitute a hard or strict naturalism. A non-reductionist, non-scientistic view will tend on the other hand to be much more liberal in what it can countenance. It will constitute a soft or tolerant naturalism.

See also SUPERVENIENCE.

PHILIP PETTIT

**naturalized epistemology** This term denotes a family of views which closely tie epistemological theorizing to theorizing in the sciences. The *locus classicus* here is Quine (1969). Quine (see QUINE) argued that the classical foundationalist project (see FOUNDATIONALISM) was a failure, both in its details and in its conception. On the classical view,

an epistemological theory would tell us how we ought to arrive at our beliefs; only by developing such a theory and then applying it could we reasonably come to believe anything about the world around us. Thus, on this classical view, an epistemological theory must be developed independently of, and prior to, any scientific theorizing; proper scientific theorizing could only occur after such a theory was developed and deployed. This was Descartes' view of how an epistemological theory ought to proceed; it was what he called first philosophy (see FIRST PHILOSOPHY). Moreover, it is this approach to epistemological issues which motivated not only foundationalism, but virtually all epistemological theorizing for the next 300 years.

Quine urged a rejection of this approach to epistemological questions. Epistemology, on Quine's view, is a branch of natural science. It studies the relationship between human beings and their environment; in particular, it asks how it is that human beings can arrive at beliefs about the world around them on the basis of sensory stimulation, the only source of belief there is. Thus, Quine commented:

The relation between the meager input [sensory stimulation] and the torrential output [our total science] is a relation 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.

(Quine, 1969, p. 83)

Quine spoke of this projected study as "epistemology naturalized".

## SCEPTICISM

One important difference between this approach and more traditional ones becomes plain when the two are applied to sceptical questions. On the classical view, if we are to explain how knowledge is possible, it is illegitimate to make use of the resources of science; this would simply beg the question against the sceptic by making use of the very

knowledge which he calls into question. Thus, Descartes' attempt (*see* DESCARTES) to answer the sceptic begins by rejecting all those beliefs about which any doubt is possible. Descartes must respond to the sceptic from a starting place which includes no beliefs at all. Naturalistic epistemologists, however, understand the demand to explain the possibility of knowledge differently. As Quine argues, sceptical questions arise from within science. It is precisely our success in understanding the world, and thus in seeing that appearance and reality may differ, that raises the sceptical question in the first place. We may thus legitimately use the resources of science to answer the question which science itself has raised. The question about how knowledge is possible should thus be construed as an empirical question: it is a question about how creatures such as we (given what our best current scientific theories tell us we are like) may come to have accurate beliefs about the world (given what our best current scientific theories tell us the world is like). Quine suggests that the Darwinian account of the origin of species gives a very general explanation of why it is that we should be well adapted to getting true beliefs about our environment (*see* Stich, 1990, ch. 3, for a useful discussion of this suggestion), while an examination of human psychology will fill in the details of such an account. Although Quine himself does not suggest it, investigations in the sociology of knowledge (*see* SOCIOLOGY OF KNOWLEDGE) are obviously relevant here as well. (A detailed discussion of the importance of social considerations, and the many epistemological issues on which they bear, may be found in Goldman, 1999.)

This approach to sceptical questions clearly makes them quite tractable, and its proponents see this, understandably, as an important advantage of the naturalistic approach. It is in part for this reason that current work in psychology and sociology is under such close scrutiny by many epistemologists. Others have stressed the importance of work in biology and cognitive ethology (*see* Godfrey-Smith, 1996; Kornblith, 2002; Millikan, 1993; and Sterelny, 2003). By the same token, the detractors of the naturalistic approach

argue that this way of dealing with sceptical questions simply bypasses the very questions which philosophers have long dealt with. Far from answering the traditional sceptical question, it is argued, the naturalistic approach merely changes the topic (*see, e.g.,* Stroud, 1981). Debates between naturalistic epistemologists and their critics thus frequently focus on whether this new way of doing epistemology adequately answers, transforms, or simply ignores the questions which others see as central to epistemological inquiry. Some see the naturalistic approach as an attempt to abandon the philosophical study of knowledge entirely (*see* DEATH OF EPISTEMOLOGY).

#### NORMATIVITY

Precisely what the Quinean project amounts to is also a subject of some controversy. Both those who see themselves as opponents of naturalized epistemology and those who are eager to sign on to the project frequently disagree about what the project is. The essay of Quine's which prompted this controversy (Quine, 1969) leaves a great deal of room for interpretation.

At the centre of this controversy is the issue of the normative dimension of epistemological inquiry (*see* FACT/VALUE). Perhaps the central role which epistemological theories have traditionally played is normative. Such theories were meant not merely to describe the various processes of belief acquisition and retention, but rather to tell us which of these processes we ought to be using. By describing his preferred epistemological approach as "a chapter of psychology and hence of natural science" (Quine, 1969, p. 82), Quine has encouraged many to interpret his view as a rejection of the normative dimension of epistemological theorizing (*see, e.g.,* Goldman, 1986, p. 2; Kim, 1988). Quine has, however, since repudiated this reading: "Naturalization of epistemology does not jettison the normative and settle for the indiscriminate description of ongoing procedures" (Quine, 1986, p. 664; *see also* Quine, 1990, pp. 19–21).

Unfortunately, matters are not quite as simple as this quotation makes things seem. Quine goes on to say:

For me normative epistemology is a branch of engineering. It is the technology of truth-seeking. . . . There is no question here of ultimate value, as in morals; it is a matter of efficacy for an ulterior end, truth or prediction. The normative here, as elsewhere in engineering, becomes descriptive when the terminal parameter is expressed.

(Quine, 1986, pp. 664–5)

But this suggestion, brief as it is, is compatible with a number of different approaches.

On one approach, championed by Alvin Goldman (Goldman, 1986), knowledge is just true belief which is produced by a reliable process, that is, a process which tends to produce true beliefs (*see* RELIABILISM). Here the “technological” question arises in asking which processes tend to produce true belief. Questions of this sort are clearly part of natural science. But there is also the account of knowledge itself. On Goldman’s view, the claim that knowledge is reliably produced true belief is arrived at independent of, and prior to, scientific investigation; it is a product of conceptual analysis. Given Quine’s rejection of appeals to meaning, the analytic/synthetic distinction, and thus the very enterprise of conceptual analysis, this position is not open to him. Nevertheless, it is for many an attractive way of allowing scientific theorizing to play a larger role in epistemology than it traditionally has, and thus one important approach which might reasonably be thought of as a naturalistic epistemology.

Those who eschew conceptual analysis will need another way of explaining how the normative dimension of epistemology arises within the context of empirical inquiry. Quine says that this normativity is not mysterious once we recognize that it “becomes descriptive when the terminal parameter is expressed”. But why is it conduciveness to truth, rather than something else, such as survival, which is at issue here? Why is it that truth counts as the goal for which we should aim? Is this merely a sociological

point, that people do seem to have this goal? Or is conduciveness to truth itself instrumental to our other goals in some way that makes it of special pragmatic importance (*see* Kornblith, 2002, ch. 5)? It is not that Quine has no way to answer these questions within the confines of the naturalistic position he defines, but rather that there seem to be many different options open here, all of which need further exploration and elaboration.

A number of attempts to fill in the naturalistic account draw a close connection between how people actually reason and how they ought to reason, thereby attempting to illuminate the relation between the normative and the descriptive. One view has it that these two are identical. (For discussion of this view, *see* the introduction to Kornblith, 1985 and 1994; Sober, 1978; Stein, 1996; *see also* PSYCHOLOGISM; RATIONALITY.) Some have argued that these two are, at least, far harder to distinguish than is commonly thought (Harman, 1986). Others hold that, while the two are distinct, any attempt to understand how we ought to reason must proceed in part by an examination of how we do reason (Bishop and Trout, 2005; Gigerenzer et al., 1999; introduction to Kornblith, 1985 and 1994; Stein, 1996). Finally, there are thoroughgoing pragmatic accounts, which prescribe processes of belief acquisition solely on the basis of their conduciveness to whatever we might value (Stich, 1990). In each of these views, the alliance between epistemological theorizing and empirical considerations, especially by way of psychology, is far closer than it is on more traditional views, and it is in virtue of this that these theories are rightly spoken of as naturalistic epistemologies.

*See also* FIRST PHILOSOPHY; PSYCHOLOGY AND EPISTEMOLOGY; QUINE; SCEPTICISM, CONTEMPORARY.

#### BIBLIOGRAPHY

- Bishop, M. and Trout, J. D.: *Epistemology and the Psychology of Human Judgment* (Oxford: Oxford University Press, 2005).

- Cherniak, C.: *Minimal Rationality* (Cambridge, MA: Bradford Books/MIT Press, 1986).
- Gigerenzer, G., Todd, P. and the ABC Research Group: *Simple Heuristics That Make Us Smart* (Oxford: Oxford University Press, 1999).
- Godfrey-Smith, P.: *Complexity and the Function of Mind in Nature* (Cambridge: Cambridge University Press, 1996).
- Goldman, A.: *Epistemology and Cognition* (Cambridge, MA: Harvard University Press, 1986).
- Goldman, A.: *Knowledge in a Social World* (Oxford: Oxford University Press, 1999).
- Harman, G.: *Change in View: Principles of Reasoning* (Cambridge, MA: Bradford Books/MIT Press, 1986).
- Kim, J.: "What Is 'Naturalized Epistemology'?", *Philosophical Perspectives* 2 (1988), 381–405.
- Kornblith, H., ed.: *Naturalizing Epistemology* (Cambridge, MA: Bradford Books/MIT Press, 1985; 2nd edn 1994).
- Kornblith, H.: *Knowledge and Its Place in Nature* (Oxford: Oxford University Press, 2002).
- Millikan, R.: "Naturalist Reflections on Knowledge," in R. Millikan, *White Queen Psychology and Other Essays for Alice* (Cambridge, MA: MIT Press, 1993), 241–64.
- Quine, W. V.: "Epistemology naturalized," in his *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969), 69–90; reprinted in Kornblith, 1985.
- Quine, W. V.: "Reply to Morton White," in *The Philosophy of W. V. Quine*, ed. L. E. Hahn and P. A. Schilpp (LaSalle, IL: Open Court, 1986), 663–5.
- Quine, W. V.: *Pursuit of Truth* (Cambridge, MA: Harvard University Press, 1990).
- Sober, E.: "Psychologism," *Journal for the Theory of Social Behavior* 8 (1978), 165–91.
- Stein, E.: *Without Good Reason: The Rationality Debate in Philosophy and Cognitive Science* (Oxford: Oxford University Press, 1996).
- Sterelny, K.: *Thought in a Hostile World* (Oxford: Blackwell, 2003).
- Stich, S.: *The Fragmentation of Reason* (Cambridge, MA: Bradford Books/MIT Press, 1990).
- Stroud, B.: "The Significance of Naturalized Epistemology," *Midwest Studies in Philosophy* 6 (1981), 455–71; reprinted in Kornblith, 1985.
- White, M.: "Normative Ethics, Normative Epistemology, and Quine's Holism," in *The Philosophy of W. V. Quine*, ed. L. E. Hahn and P. A. Schilpp (LaSalle, IL: Open Court, 1986), 649–62.

HILARY KORNBLITH

**necessary/contingent** Necessary truths are ones which *must* be true, or whose opposite is impossible. Contingent truths are those that are not necessary and whose opposite is therefore possible. 1–3 below are necessary, 4–6, contingent.

1. It is not the case that it is raining and not raining.
2.  $2 + 2 = 4$ .
3. All bachelors are unmarried.
4. It seldom rains in the Sahara.
5. There are more than four states in the USA.
6. Some bachelors drive Maseratis.

Plantinga (1974, p. 2) characterizes the sense of necessity illustrated in 1–3 as "broadly logical". For it includes not only truths of logic, but those of mathematics, set theory, and other quasi-logical ones. Yet it is not so broad as to include matters of causal or natural necessity, such as

7. Nothing travels faster than the speed of light.

One would like an account of the basis of our distinction and a criterion by which to apply it. Some suppose that necessary truths are those we know *a priori*. But we lack a criterion for *a priori* truths, and there are necessary truths we don't know at all (e.g. undiscovered mathematical ones). It won't help to say that necessary truths are ones it is *possible*, in the broadly logical sense, to know *a priori*, for this is circular. Finally, Kripke (1972, p. 253) and Plantinga (1974, p. 8) argue that some contingent truths are knowable *a priori* (see A PRIORI KNOWLEDGE in Part I). Similar problems face the suggestion that necessary truths are the ones we know with certainty (see CERTAINTY): we lack a criterion for certainty; there are necessary truths we don't know; and (barring dubious arguments for scepticism) it is reasonable to suppose that we know some contingent truths with certainty.

Leibniz (see LEIBNIZ) defined a necessary truth as one whose opposite implies a contradiction. Every such proposition, he held, is either an explicit identity (i.e. of the form "A

is A", "AB is B", etc.) or is reducible to an identity by successively substituting equivalent terms. (Thus, 3 above might be so reduced by substituting "unmarried man" for "bachelor".) This has several advantages over the ideas of the previous paragraph. First, it *explicates* the notions of necessity and possibility and seems to provide a criterion we can apply. Second, because explicit identities are self-evident a priori propositions, the theory implies that all necessary truths are knowable a priori, but it does not entail that we actually know all of them, nor does it define "knowable" in a circular way. Third, it implies that necessary truths are knowable with certainty, but does not preclude our having certain knowledge of contingent truths by means other than a reduction.

Nevertheless, this view is also problematic. Leibniz's examples of reductions are too sparse to prove a claim about *all* necessary truths. Some of his reductions, moreover, are deficient: Frege has pointed out, for example, that his proof of " $2 + 2 = 4$ " presupposes the principle of association and so does not depend only on the principle of identity. More generally, it has been shown that arithmetic cannot be reduced to logic, but requires the resources of set theory as well. Finally, there are other necessary propositions (e.g. "Nothing can be red and green all over") which do not *seem* to be reducible to identities and which Leibniz does not show how to reduce.

Leibniz and others have thought of truth as a property of propositions, where the latter are conceived as things which may be expressed by, but are distinct from, linguistic items like statements. On another approach, truth is a property of linguistic entities, and the basis of necessary truth is convention. Thus A. J. Ayer (see AYER), for example, argued that the only necessary truths are analytic statements and that the latter rest entirely on our commitment to use words in certain ways (see LOGICAL POSITIVISM). But, while there have been many attempts to define analyticity, Quine (see QUINE) has criticized the most powerful ones and rendered it uncertain whether a criterion for this notion can be given.

When one predicates necessary truth of a proposition one speaks of modality *de dicto*.

For one ascribes the modal property, *necessary truth*, to a *dictum*, namely, whatever proposition is taken as necessary. A venerable tradition, however, distinguishes this from necessity *de re*, wherein one predicates *necessary* or *essential possession* of some property to an *object*. For example, the statement "4 is necessarily greater than 2" might be used to predicate of the object, 4, the property, *being necessarily greater than 2*. That objects have some of their properties necessarily, or essentially, and others only contingently, or accidentally, is a main part of the doctrine called "essentialism". Thus, an essentialist might say that Socrates had the property of *being bald* accidentally, but that of *being self-identical*, or perhaps of *being human*, essentially. Although essentialism has been vigorously attacked in recent years, most particularly by Quine, it also has able contemporary proponents, such as Plantinga.

See also A PRIORI/A POSTERIORI; A PRIORI KNOWLEDGE in Part I; ANALYTICITY; LOGICAL POSITIVISM; NECESSITY, MODAL KNOWLEDGE; TRUTHS OF REASON/TRUTHS OF FACT.

#### BIBLIOGRAPHY

- Hamlyn, D. W.: "Necessary and Contingent Statements," in *The Encyclopedia of Philosophy*, ed. P. Edwards (New York: Macmillan and the Free Press, 1967), vol. 2, pp. 198–205.
- Kripke, S.: "Naming and Necessity," in *Semantics of Natural Language*, ed. D. Davidson and G. Harman (Dordrecht: Reidel, 1972), 253–355.
- Plantinga, A.: *The Nature of Necessity* (Oxford: Oxford University Press, 1974).
- Quine, W. V.: *Word and Object* (Cambridge, MA: MIT Press, 1960).
- Wilson, M. D.: *Leibniz' Doctrine of Necessary Truth* (New York: Garland, 1990).

DAVID BLUMENFELD

**necessity, modal knowledge** Philosophers have traditionally held that every proposition has a *modal status* as well as a *truth value*. Every proposition is either necessary or contingent as well as either true or false (see NECESSARY/CONTINGENT). The issue of knowledge



of the modal status of propositions has received much attention because of its intimate relationship to the issue of a priori knowledge. For example, many proponents of the a priori contend that all knowledge of necessary propositions is a priori (see A PRIORI KNOWLEDGE in Part I). Others reject this claim by citing Kripke's (1980) alleged cases of necessary a posteriori propositions (see A PRIORI/A POSTERIORI). Such contentions are often inconclusive, for they fail to take into account the following tripartite distinction: S knows the *general modal status* of *p* just in case S knows that *p* is a necessary proposition or S knows that *p* is a contingent proposition. S knows the *truth value* of *p* just in case S knows that *p* is true or S knows that *p* is false. S knows the *specific modal status* of *p* just in case S knows that *p* is necessarily true or S knows that *p* is necessarily false or S knows that *p* is contingently true or S knows that *p* is contingently false. It does not follow from the fact that knowledge of the general modal status of a proposition is a priori that knowledge of its specific modal status is also a priori. Nor does it follow from the fact that knowledge of the specific modal status of a proposition is a posteriori that knowledge of its general modal status is also a posteriori.

## BIBLIOGRAPHY

- Casullo, A.: "Kripke on the A Priori and the Necessary," *Analysis* 37 (1977), 152–9; reprinted in Moser, 1987.
- Kitcher, P.: "Apriority and Necessity," *Australasian Journal of Philosophy* 58 (1980), 89–101; reprinted in Moser, 1987.
- Kripke, S.: *Naming and Necessity* (Cambridge, MA: Harvard University Press, 1980); excerpted in Moser, 1987.
- Moser, P., ed.: *A Priori Knowledge* (Oxford: Oxford University Press, 1987).
- Swinburne, R. G.: "Analyticity, Necessity, and Apriority," *Mind* 84 (1975), 225–43; reprinted in Moser, 1987.

ALBERT CASULLO

**Neoplatonism** This is the conventional designation for the revived form of Platonism inspired by Plotinus (AD 205–70) and

popularised by Porphyry (233–c.305) and Iamblichus (c.245–c.326). Its most important later figures were the Athenian philosophers Syrianus (d. c.437), Proclus (412–85) and Damascius (c.462–after 538), as well as the Aristotelian commentators Ammonius (late fifth–early sixth century), and Simplicius and Philoponus (younger contemporaries of Damascius). The last major figure in Greek Neoplatonism was Olympiodorus of Alexandria (late sixth century), but Neoplatonism continued to exercise a major influence on Arabic and Latin medieval philosophy.

Neoplatonists were largely unconcerned with questions about knowledge of the sensible world and with the debates over justification of knowledge – claims that had been prominent earlier in Greek philosophy. Nevertheless they did have views about the nature and origin of knowledge. They were less interested in the nature of empirical knowledge than in theoretical knowledge: Plotinus' only references to sceptical arguments come in discussions of Intellect's relation to the Ideas (*Enn.* V.5.I; V.3.5). Their most explicit texts on the nature of empirical knowledge are found in Plotinus (V.5.2–3; V.3.4–5; cf. also 1.6.3, 1.3.5 and IV.4.22–3) and Porphyry (*Commentary on Ptolemy* pp. 13–14, Düring), both of whom are following earlier tradition. There are also important discussions of cognition in Proclus' commentary on the *Timaeus*, and we would know considerably more about Neoplatonist epistemology if anything of Proclus' commentary on Plato's *Theaetetus* had survived. Plotinus' account of empirical knowledge, which is presumably typical, is based on Plato's theory of recollection (see ANAMNESIS), and has important connections with the view of recollection found in earlier Platonist texts (Cicero, *Tusculan Disputations* I.57–8; Albinus, *Didaskalikos* §4 and §25; Plutarch *apud* Olympiodorus, *Commentary on the Gorgias* p. 156).

According to Plotinus one judges a sensible thing to have a certain attribute by comparing it, i.e. its abstracted sensible form – here following Aristotle's theory of perception (see ARISTOTLE), except that the soul does not actually receive the form – with an apparently innate standard in the soul (V.3.3),

which is conceived as a trace or reflection of the corresponding Idea. Mistaken judgements can be accounted for, as in the Wax Tablet analogy of the *Theaetetus*, as mismatches due to the trace or standard's initially being confused or obscure. These traces in the soul are identified with the so-called common or natural conceptions, concepts or beliefs about the natures of things that are innate or develop naturally and hence are shared by all or nearly all human beings (e.g. *Enn.* VI.5.1.*init.*; cf. Sallustius, *De diis et mundo* §1).

Theoretical knowledge or understanding, identified with the grasp of Platonic Ideas, is attained by a reflective process of clarifying or “articulating” these initially confused innate traces or conceptions (*Enn.* I.2.4, 18ff.; cf. Porphyry *On Abstinence* I.31.2 and *To Marcella* §10). One who has reached such understanding has come to participate directly in Intellect, an eternal mental act whose content is the Ideas. False theoretical beliefs on such a view are mere opinions, belonging to a lower cognitive level: Intellect by definition is always correct (*Enn.* I.1.9,12). Indeed, Plotinus asserts that Intellect's thinking, which is the Ideas, is identical with Truth itself (V.5.1–2). Correct theoretical judgements are therefore not true of independently existing entities, but rather true thoughts *are* those entities. This is inspired by Aristotle's theory of the nature of Intellect (*De Anima* III.4, 430a 3–4), though Plotinus does not go so far as to identify Intellect as opposed to its acts of thought with their objects. Plotinus thus holds a non-correspondence, direct realist view of theoretical truth. Truth is in turn identified with real Being, conceived as the system of inter-related ideas, which may be investigated by Platonic dialectic (*Enn.* I.3.4; VI.2.19–22). Other Neoplatonists did not disagree with this basic picture; they also agreed with Plotinus that what is prior to Intellect, which Plotinus called the One or the Good, is ineffable and unknowable though like him they persisted in efforts to indicate something of its nature. Such issues, with which they were deeply concerned, involve something higher than mere rational knowledge, though they have implications for its status. It must be

said that the Neoplatonists were no more successful than was Plato in explaining how theoretical knowledge, conceived as having special, separate objects, could apply to the sensible world. Indeed apart from Plotinus (VI.7.1–14) they did not much occupy themselves with this problem.

#### BIBLIOGRAPHY

##### *Texts*

- Alcinous: *Enseignement des Doctrines de Platon [Didaskalikos]*, ed. J. Whittaker and P. Louis (Paris: Bude, 1990).  
 Olympiodorus: *In Platonis Gorgiam Commentaria*, ed. W. Norvin (Leipzig: Teubner, 1936).  
 Plotinus: *Enneads*, 7 vols, trans. A. H. Armstrong (Cambridge, MA: Loeb Classical Library, 1966–88).  
 Porphyry: *De l'Abstinence*, book I, ed. J. Bouffartigue and M. Patillon (Paris: Les Belles Lettres, 1977).  
 Porphyry: *Lettre à Marcella*, ed. E. des Places (Paris: Les Belles Lettres, 1982).  
 Porphyry: *Kommentar zur Harmonielehre des Ptolemaios*, ed. I. Düring (Göteborg: Elanders, 1932); (reprinted New York: 1979).  
 Proclus: *Commentaire sur le Timée*, 5 vols, trans. A. J. Festugière (Paris: Vrin, 1966–8).  
 Sallustius: *Concerning the Gods and the Universe*, ed. A. D. Nock (Cambridge: Cambridge University Press, 1926).

##### *Secondary sources*

- Blumenthal, H. J.: “Plotin and Proclus on the Criterion of Truth,” in *The Criterion of Truth: Essays Written in Honour of George Kerferd*, ed. P. Huby and G. Neal (Liverpool: Liverpool University Press, 1989).  
 Lloyd, A. C.: *The Anatomy of Neoplatonism* (Oxford: Oxford University Press, 1990), ch. 6.  
 Wallis, R. T.: *Neoplatonism* (London: Duckworth, 1972).

S. STRANGE

**Neurath, Otto (1882–1945)** Austrian philosopher. Neurath was one of the leading logical positivists (see LOGICAL POSITIVISM), whose major work was in sociology, economics and scientific method. He wanted to

accord sociology an empirical scientific status, and hoped for the development of a unified, physicalist, scientific language which excluded all imprecision and unverifiable metaphysics. This aspiration led him in the 1930s to plan the *International Encyclopedia of Unified Science*, but only the first two introductory volumes ever appeared. As Passmore (1968) says, “his philosophical position was never worked out in detail: agitation was his *forte*”.

There was a tension in logical positivism between empiricism (see EMPIRICISM) on the one hand, and physicalism and the goal of the “unity of science” on the other. Neurath was never attracted to the empiricist, foundationalist (see FOUNDATIONALISM) strand; from the start, he insisted on the possibility of intersubjective confirmation afforded by a physicalist as opposed to a phenomenalist (see PHENOMENALISM) language. This advocacy of a robust physicalism led him, in “Protocol Sentences” (1932–3), to reject the “methodological” approach to competing metaphysical standpoints advocated in Carnap’s *Der logische Aufbau der Welt* (1928) (see CARNAP).

Neurath’s most important contribution to epistemology, however, lies in his advocacy of coherentism and holism (see COHERENTISM; HOLISM), which seemed to follow from his physicalist rejection of foundationalism. In his (1931–2), he argued (albeit laconically) that “statements are compared with statements, not with ‘experiences’, ‘the world’, or anything else”. And in “Sociology and Physicalism” (1932–3) he argued further that ‘no sentence enjoys the *noli me tangere* which Carnap ordains for protocol sentences (see PROTOCOL SENTENCES). In a celebrated metaphor, he compared science to a boat: “There is no way of taking conclusively established pure protocol sentences as the starting point of the sciences. We are like sailors who must rebuild their ship on the open sea, never able to dismantle it in dry-dock and to reconstruct it there out of the best materials.” These views were criticized by Schlick (1934) (see SCHLICK), who called the “coherence theory of truth” an “astounding error”, and vigorously affirmed the foundational role of an obscure class of unrevisable *Konstatierungen* or quasi-judgements of immediate experience.

This debate between the “right” and “left” wings of the Vienna Circle probably generated more heat than light. Neurath and his supporters ended up by abandoning the distinction between basic and non-basic classes of knowledge-claims; Schlick (1935) issued the “gentle warning of a true empiricist”, criticizing this move as “rationalistic”. Carnap, who always liked to think that everyone was in agreement really, urged a compromise in “Truth and Confirmation” (1935). Rejecting the idea of certain knowledge of basic sentences, Carnap maintained that “the confrontation of statements with observations” was central to logical empiricism. At least partly because of his social science background, Neurath, unlike the other logical positivists, was ready to endorse a naturalistic picture of philosophy as a branch of empirical (perhaps social) science rather than seeing it as the “logic of science”. His views look forward to the naturalized epistemology of Quine, who extended Neurath’s holism to cover logical, mathematical and analytic truths in general (see NATURALIZED EPISTEMOLOGY; QUINE; ANALYTICITY).

See also PROTOCOL SENTENCES.

#### WRITINGS

“Sociology and Physicalism,” *Erkenntnis* 2 (1931–2); in *Logical Positivism*, ed. A. J. Ayer (New York: Free Press, 1959).

“Protokollsätze,” *Erkenntnis* 3 (1932–3); trans. as “Protocol Sentences,” in *Logical Positivism*, ed. A. J. Ayer (New York: Free Press, 1959).

Otto Neurath, *Empiricism and Sociology*, ed. M. Neurath and R. Cohen (Dordrecht: Reidel, 1973).

#### BIBLIOGRAPHY

Carnap, R.: *Der logische Aufbau der Welt* (1928); trans. as *The Logical Structure of the World* (Berkeley, CA: University of California Press, 1967).

Carnap, R.: “Truth and Confirmation” (1935); trans. and adapted in *Readings in Philosophical Analysis*, ed. H. Feigl and W. Sellars (New York: Appleton-Century-Crofts, 1949).

Jacob, P.: "The Neurath–Schlick Controversy," *Fundamenta Scientiae* 5 (1984), 351–66.  
For articles by Schlick, see the bibliography to SCHLICK.

ANDY HAMILTON

### **Nietzsche, Friedrich (1844–1900)**

German philosopher and writer. Nietzsche is openly pessimistic about the possibility of knowledge:

We simply lack any organ for knowledge, for "truth": we "know" (or believe or imagine) just as much as may be *useful* in the interests of the human herd, the species: and even what is here called "utility" is ultimately also a mere belief, something imaginary and perhaps precisely that most calamitous stupidity of which we shall perish some day.

(*The Gay Science*, 354)

This position is very radical. Nietzsche does not simply deny that knowledge, construed as the adequate representation of the world by the intellect, exists. He also refuses the pragmatist (see PRAGMATISM) identification of knowledge and truth with usefulness: he writes that we think we know what we think is useful, and that we can be quite wrong about the latter.

Nietzsche's view, his "perspectivism", depends on his claim that there is no sensible conception of a world independent of human interpretation and to which interpretations would correspond if they were to constitute knowledge. He sums up this highly controversial position in *The Will to Power*: "Facts are precisely what there is not, only interpretations" (481).

It is often claimed that perspectivism is self-undermining. If the thesis that all views are interpretations is true then, it is argued, there is at least one view that is not an interpretation. If, on the other hand, the thesis is itself an interpretation, then there is no reason to believe that it is true, and it follows again that not every view is an interpretation.

But this refutation assumes that if a view (in this case, perspectivism itself) is an interpretation it is *ipso facto* wrong. This is not the case. To call any view, including perspectivism, an interpretation is to say that it can be wrong,

which is true of all views, and that is not a sufficient refutation. To show the perspectivism is actually false it is necessary to produce another view superior to it on specific epistemological grounds.

Perspectivism does not deny that particular views can be true. Like some versions of contemporary anti-realism, it attributes to specific approaches truth in relation to facts specified internally by those approaches themselves. But it refuses to envisage a single independent set of facts, to be accounted for by all theories. Thus Nietzsche grants the truth of specific scientific theories; he does, however, deny that a scientific interpretation can possibly be "the only justifiable interpretation of the world" (*The Gay Science*, 354): neither the facts science addresses nor the methods it employs are privileged. Scientific theories serve the purposes for which they have been devised: but these have no priority over the many other purposes of human life.

The existence of many purposes and needs relative to which the value of theories is established – another crucial element of perspectivism – is sometimes thought to imply a rampant relativism (see RELATIVISM), according to which no standards for evaluating purposes and theories can be devised. This is correct only in that Nietzsche denies the existence of a single set of standards for determining epistemic value once and for all. But he holds that *specific* views can be compared with and evaluated in relation to one another. The ability to use criteria acceptable in particular circumstances does not presuppose the existence of criteria applicable in all. Agreement is therefore not always possible, since individuals may sometimes differ over the most fundamental issues dividing them.

But Nietzsche would not be troubled by this fact, which his opponents too also have to confront only, as he would argue, to suppress it by insisting on the hope that all disagreements are in principle eliminable even if our practice falls woefully short of the ideal. Nietzsche abandons that ideal. He considers irresolvable disagreement an essential part of human life.

*See also* CONTINENTAL EPISTEMOLOGY.



- Jenseits von Gut und Böse* (Leipzig, 1886); trans. W. Kaufmann as *Beyond Good and Evil* (New York: Random House, 1966).
- Die fröhliche Wissenschaft* (Leipzig, 1887); trans. W. Kaufmann as *The Gay Science* (New York: Random House, 1974).
- Zur Genealogie der Moral* (Leipzig, 1887); trans. W. Kaufmann and R. J. Hollingdale as *On the Genealogy of Morals* (New York: Random House, 1969).
- Der Wille zur Macht*, ed. E. Förster-Nietzsche (Leipzig, 1906); trans. W. Kaufmann and R. J. Hollingdale as *The Will to Power* (New York: Random House, 1968).

# BIBLIOGRAPHY

- Clark, M.: *Nietzsche on Truth and Philosophy* (Cambridge: Cambridge University Press, 1990).
- Danto, A. C.: *Nietzsche as Philosopher* (New York: Macmillan, 1965).
- Kaufmann, W.: *Nietzsche: Philosopher, Psychologist, Antichrist* (Princeton, NJ: Princeton University Press, 1974).
- Nehamas, A.: *Nietzsche: Life as Literature* (Cambridge, MA: Harvard University Press, 1985).

ALEXANDER NEHAMAS

**nihilism** *see* DEATH OF EPISTEMOLOGY.

**noumenal/phenomenal** In the philosophy of Kant (*see* KANT), the terms “noumena” and “phenomena” are coextensive (though not exactly synonymous) with the terms “things in themselves” and “appearances”. Things in themselves are things as they exist independently of human cognition; appearances are things that exist only as contents or objects of sensible representations. (*See* A491/B519–A494/B522. Numbers preceded by “A” and “B” are references to pages in the first and second editions respectively of Kant’s *Critique of Pure Reason*.) The adjectives “noumenal” and “phenomenal” have therefore come to mark the distinction between those entities or aspects of reality that exist independently of human cognition and those that exist only in relation to it. How much of the

world is noumenal, how much phenomenal, and the extent to which the noumenal portion can be known are, of course, central questions of philosophy.

Some of Kant’s major views about noumena and phenomena may be listed as follows, (1) There certainly is a noumenal element in the world; otherwise “we should be landed in the absurd conclusion that there can be appearance without anything that appears” (Bxxvii). (2) The noumenal element in some way affects us, giving rise to the representations through which phenomena are presented and constituted. (*See e.g.* AI90/B235 and *Prolegomena* pp. 61–2.) (3) More of the world’s features fall on the phenomenal side of the line than either Descartes or Locke believed. In particular, space and time – and thus nearly all of Locke’s primary qualities (*see* PRIMARY AND SECONDARY QUALITIES) – are merely phenomenal, space and time being “forms of intuition” rather than features of things in themselves. Kant called this view “transcendental idealism”. (A26/B42–A28/B44; *Prolegomena* pp. 36–7.) (4) Things in themselves are unknowable by human beings (A30/B45, A44/B62). We can have knowledge only concerning what is given in intuition (roughly, perception or introspection), and things in themselves are not thus given. So for human beings, things in themselves are not noumena in the positive (and etymological) sense of entities knowable by nous, or pure intelligence; they are noumena only in the negative sense of entities not knowable by sensible means. (For the distinction between the positive and the negative senses of “noumenon”, *see* B307–9.)

Kant’s combination of views has struck many of his readers as highly problematic. Two notorious problems are the following: (1) If things in themselves are unknowable, how can Kant be in a position to affirm that they are outside of space and time, or indeed, that they exist at all? (2) If things in themselves are non-spatial and non-temporal, what sense can we make of the supposition that they cause our representations? To answer (1), Kant would have to qualify the claim that there can be no knowledge concerning things in themselves. He did certainly think



that the non-spatiality and non-temporality of things in themselves could be proved, by the following considerations among others: if noumena were spatial and temporal, there would be no accounting for our a priori knowledge of arithmetical and geometrical truths (A46/B64–A49/B66); nor would there be any way of avoiding certain antinomies, e.g. that the world both does and does not have a beginning in time. (See the chapter of the *Critique of Pure Reason* entitled “The Antinomy of Pure Reason”.) As for (2), Kant himself insisted that things in themselves cannot *cause* appearances, causation being a concept applicable only to phenomena in time, but he maintained none the less that things in themselves are in some mysterious sense the *ground* of appearances (A380).

Not satisfied with these answers, Post-Kantian German philosophers tended to jettison the thing in itself, moving closer to pure idealism (see IDEALISM). Others, including the British realists beginning with Moore (see MOORE), reinvested the thing in itself with the features that Kant had stripped away.

What has been said so far suggests that noumena and phenomena are for Kant two distinct realms of objects. This is the traditional interpretation of Kant, but it is a view now in disfavour among many Kant scholars. The main opposing view, sometimes called the “one world” or “double aspect” view, holds instead that there is only one set of objects, but two ways of considering them or talking about them (Allison, 1983). To talk of appearances is to talk of objects as we know them; to talk of things in themselves is to talk of these same objects as they are independently of our knowledge. Kant’s transcendental idealism then comes to this: things have certain features (e.g. spatial form) *qua* appearances that they lack in *themselves*.

But how can the same things be both spatial and non-spatial? If we omit the qualifying phrases “*qua* appearances” and “in themselves”, we get a contradiction; how does retaining the qualifiers avoid the contradiction? This is a question that proponents of the one-world view have not adequately addressed (Van Cleve, 1992). Another problem for the view is that Kant believes that there are

certain principles, for example, Leibniz’s principle of the identity of indiscernibles, that hold for noumena but not for phenomena (A264/B320; A272/B328). For a one-worlder, this apparently has the following implication: A and B, if indiscernible, will be identical *qua* noumena, but distinct *qua* phenomena. How can that be?

The traditional two-worlds interpretation (which treats appearances as existents distinct from mental representings, though dependent on them) also gives rise to problems. For one thing, some real existents, namely appearances, would be in space and time, a result apparently ruled out by Kant’s argument in the antinomies (Moore, 1953, ch. 9). For another, if appearances are entities genuinely distinct from the acts that apprehend them, it becomes difficult to see why they should have to “conform to our knowledge” in the manner required by the Copernican Revolution (for which, see KANT).

There is a third interpretation that seeks to avoid the problems besetting the first two. According to it, phenomena are not the same objects as noumena, considered from a specially human viewpoint; nor are they a second class of objects existing alongside noumena. Instead, they are what we might call *virtual objects*. A virtual object is similar to what Brentano (see BRENTANO) called an “intentional object”, except that it is not to be conceived as having its own special kind of being. Instead, to say that a virtual object of a certain sort (e.g. a patch of red) exists is shorthand for saying that a certain kind of representation occurs; in the case of a more complex virtual object (e.g. a house or a ship), it is to say that an entire rule-governed sequence of representations occurs or is in the offing. According to this suggestion, Kant’s transcendental idealism is a form of phenomenalism (see PHENOMENALISM). Unlike some phenomenalists, however, Kant was also a noumenalist: he believed there are some objects, the things in themselves, that resist phenomenalist reduction. (If nothing else, there are the cognitive acts and agents to which phenomena are reduced, for these can hardly be supposed to exist only as the virtual objects of further acts.)

The third view gets between the alternatives of one world versus two. In a sense there are two worlds, because noumena and phenomena constitute separate domains of discourse rather than two ways of discoursing about the same entities; but in a sense there is only one, since phenomena are logical constructions out of noumenal beings and their states rather than things existing in their own right.

Interesting comparisons of Kant's views about noumena and phenomena with recent developments in anti-realist philosophy may be found in Putnam (1981, ch. 3) and Posy (1983).

See IN ITSELF/FOR ITSELF; PRIMARY AND SECONDARY QUALITIES.

#### BIBLIOGRAPHY

- Kant, I.: *Critique of Pure Reason* (1781), trans. N. Kemp Smith (London: Macmillan, 1964).
- Kant, I.: *Prolegomena to Any Future Metaphysics* (1783), trans. L. W. Beck (Indianapolis, IN: Bobbs-Merrill, 1950).
- Allison, H.: *Kant's Transcendental Idealism* (New Haven, CT: Yale University Press, 1973).
- Ameriks, K.: "Recent Work on Kant's Theoretical Philosophy," *American Philosophical Quarterly* 19 (1982), 1–24.
- Aquila, R.: "Things in Themselves and Appearances: Intentionality and Reality in Kant," *Archiv für Geschichte der Philosophie* 61 (1979), 293–307.
- Moore, G. E.: *Some Main Problems of Philosophy* (London: Macmillan, 1953).
- Posy, C.: "Dancing to the Antinomy," *American Philosophical Quarterly* 20 (1983), 81–94.
- Prauss, G.: *Kant und das Problem der Dinge an sich* (Bonn: Bouvier, 1974).
- Putnam, H.: *Reason, Truth, and History* (Cambridge: Cambridge University Press, 1981).
- Smith, N. Kemp: *A Commentary on Kant's Critique of Pure Reason*, 2nd edn (London: Macmillan, 1923).
- Strawson, P. F.: *The Bounds of Sense* (London: Methuen, 1966).
- Van Cleve, J.: "Geometry, Transcendental Idealism, and Kant's Two Worlds," forthcoming in the North American Kant Society Studies in Philosophy Series.

JAMES VAN CLEVE

**Nozick, Robert (1940–2002)** Robert Nozick, most famously the author of *Anarchy, State, and Utopia* (1974), made well-known contributions in many areas of philosophy. After receiving his PhD. at Princeton in 1963, he became a professor at Harvard in 1969 and died the Pellegrino University Professor there. In epistemology, he is best-known for a striking reply to skepticism that follows from his tracking theory of knowledge. This theory illustrates a potentially surprising connection between how we answer the question of what knowledge is (the analysis of the concept of knowledge) and how we reply to the skeptic. Nozick came to the tracking view of knowledge via an attempt to make progress on the problem that free will and responsibility seem to be threatened by the fact that our actions are produced by causal processes. Arguably, knowledge is not threatened by the fact that beliefs are formed by causal processes. This point, according to Nozick's approach, might provide a clue to defending free will. His goal, then, was to find a deep analogy between two problems: how knowledge is possible given the threat of skepticism, and how free will is possible given the threat of determinism. The problem of skepticism arises as follows: Since our knowledge of the world can come only through causal intermediaries, we can imagine breaking that sequence in the middle and rendering what is accessible to a person's mind irrelevant to the way she is actually connected to the world. That is, we can imagine being a brain in a vat but fed with all the same impressions as we now have (cf. the impression of free will). How do we know we are not in this situation?

Nozick, then, aimed to explain how knowledge was *possible* given the problem that we might be brains in vats and are apparently unable to rule this out. His explanation appeals to a connection between the world and beliefs about it that is more general than causation. Though beliefs are formed by causal processes, it is not causation per se but rather a set of properties causes often have that explains what makes a true belief knowledge. These properties are modal and are captured by subjunctive conditionals. Taking

fulfillment of these conditionals to be necessary for knowledge shows, for reasons discussed below, that it is possible for us to acquire knowledge, namely by tracking the world, even if we don't know that the skeptical hypothesis is false. Similarly, Nozick argued that, even if determinism is true, there is something that we can do that corresponds to free will, namely tracking bestness, which signals our value as actors. Skepticism and determinism are defused without being refuted.

Nozick's tracking theory of knowledge follows the traditional form of analysis of the concept of knowledge in so far as it takes knowledge of *p* to require true belief that *p* plus fulfillment of other conditions. It departs from the older tradition in not taking justification to be required for knowledge. Importantly, this means that the further conditions he imposes – the “tracking” conditions – are not offered or understood as an analysis of the concept of justification.

S knows that *p*, on the tracking view, if and only if:

- (1) *p* is true;
- (2) S believes *p*;
- (3) If *p* weren't true, then S wouldn't believe *p* (*variation condition*);
- (4) If *p* were true, then S would believe *p* (*adherence condition*).

The variation condition is inspired by the idea that the subject's belief in *p* should vary with the truth value of *p*. Condition (4) completes that idea by requiring that the subject have or stick with the belief in possible worlds where *p* is true despite circumstances that could throw one off. The variation condition has much to recommend it. It is as successful as anything is in resolving Gettier cases, for example. The problem in these cases is that the belief's truth and its being justified are only accidentally related. Condition (3) diagnoses that accidentality as the subject's lack of proper sensitivity to the potential falsity of the belief, a lack that is due precisely to the lack of a proper connection between the truth of the belief and the reason it is held. It is also uniquely suited to resolve

the lottery problem. Though my belief that my ticket will not win the lottery is justified, safe, and formed by a reliable process, still intuition says I don't know it won't win. The reason for my lack of knowledge is failure of the variation condition: Even if my ticket were going to win, I wouldn't believe it was. It may or may not be a problem that the variation condition is undefined for necessary truths, since they can't possibly be false. Nozick thought the adherence condition was sufficient in those cases.

The tracking view is a fallibilist view – it implies it is possible to have knowledge even if your belief might have been false – because of the use of subjunctive conditionals rather than material conditionals or implication. Subjunctives are such that it is not in every possible world that the subject must believe or fail to believe the right thing, but only in the *close* or closest possible world(s). Developing a satisfactory semantics for the subjunctives Nozick's theory needs faces a variety of problems. On the classical Lewisian semantics, for example, the adherence condition is trivially fulfilled if the actual world is the one closest to itself, which it would plausibly seem to be. The semantics of subjunctive conditions is a topic of ongoing research today, on this and other points.

Subjunctive conditionals are also key to understanding the consequences of tracking for skepticism. Simply put, we can fulfill the variation condition for *p* = “I have a hand” while failing to fulfill this condition for *q* = “I am not a brain in a vat”, because the set of worlds closest to the worlds in which *p* is false is different from the set of worlds closest to the worlds in which *q* is false. The closest possible worlds in which I don't have a hand – e.g. I had an accident or have an abnormality – are ones in which I would notice that lack, and therefore also would not believe I had a hand. The closest worlds in which I am a brain in a vat are ones where all my impressions are as they are now. I would have no way of noticing my envattedness, and therefore surely would, and a fortiori might, believe I was not a brain in a vat. These consequences of the variation condition are surprising since my having a hand *implies* that I am not a brain

in a vat, and I know that. What we have is a view of knowledge according to which I may know *p*, and know that *p* implies *q*, and nevertheless not know *q*. We say that a view with this property is not closed under known implication. Many types of cases of closure failure follow from the variation condition, and it is significant that the skeptical situation is one of them.

The challenge from a (particular type of) skeptic has always exploited the presumed closure of knowledge under known implication. After exposing our inability to defend the claim that we know we are not brains in vats, the skeptic and we infer that we also cannot defend the claim that we have common-sense knowledge. This is because if we did know about our hands, then we surely would know we weren't brains in vats. Why? Because the first *implies* the second. Thus, if we don't know we're not brains in vats, then we don't know most of what we think we know. Non-closure severs this link between our ordinary knowledge and knowledge that the skeptical hypothesis is false. We can grant the skeptic's compelling starting point – we don't know we are not victims of misrepresentation on a massive scale – while conceding nothing that threatens our ordinary knowledge.

It may seem that this view of knowledge merely offers a possible explanation of our situation vis-à-vis the skeptic. But exploring the consequences of the variation condition exposes the non-skeptic's stronger dialectical position: To convince us we don't know we're not brains in vats, the skeptic implicitly appeals to the variation condition as necessary for knowledge. However, to draw out bad consequences for our ordinary knowledge, the skeptic must also appeal to closure. As we have just seen, he cannot consistently hold both assumptions. Thus, it appears that it is not merely conditional on the tracking view of knowledge, but unconditionally, that the skeptic is forced to go back to the drawing board.

Nozick also made substantial contributions to epistemology in his theory of rationality, and his account of objectivity. See Nozick, 1993, 2001.

See also EXTERNALISM/INTERNALISM; FALLIBILISM; GETTIER PROBLEM; PROPOSITIONAL KNOWLEDGE; RELIABILISM; SCEPTICISM; SCEPTICISM AND CLOSURE in Part I; SENSITIVITY AND SAFETY.

## WRITINGS

- Anarchy, State, and Utopia* (Cambridge, MA: Harvard University Press, 1974).  
*Philosophical Explanations* (Cambridge, MA: Harvard University Press, 1981).  
*The Examined Life* (Oxford: Oxford University Press, 1990).  
*The Nature of Rationality* (Princeton, NJ: Princeton University Press, 1993).  
*Socratic Puzzles* (Cambridge, MA: Harvard University Press, 1999).  
*Invariances: The Structure of the Objective World* (Cambridge, MA: Harvard University Press, 2001).

SHERRILYN ROUSH

**Nyāya** The Nyāya system of philosophy (traceable to Gotama's *Nyāyasūtras* (AD c.150) and Vatsyāyana's commentary (fourth century AD) on them, developed by a host of brilliant authors such as Uddyotakara (sixth century AD), Vacaspati (ninth century AD) and Udayana (eleventh century AD) developed a theory of *pramānas*, i.e. means of true cognition. The Neo-Nyāya school founded by Gangeśa (fourteenth century AD) developed a technical language for analysing the structure of propositional cognitions whose objects are relational complexes of the form *aRb*. Here *a* is the qualificandum, and *b* is the qualifier; *aRb* is a complex term, not a proposition. For the mode of cognition, the system introduced the idea of "limitor" (*avacchedaka*). The cognition expressed in the term "a brown jar" has as its qualificandum the thing jar and as its qualifier the brown colour. In this case, the qualificierness (*prakāratā*) is limited by brownness and the relation of inherence (*samavāya*), and as so limited it determines the qualificandumness limited by jarness. This pattern of analysis is employed to elaborate/analyse and develop the structures of various kinds of cognitions.

See also INDIAN EPISTEMOLOGY.

BIBLIOGRAPHY

- Annambhatta: *Tarkasamgraha*, trans. with notes by G. Bhattacharya (Calcutta: Progressive Publishers, 1976).
- Bhattacharya, S.: "Some Features of the Technical Language of Navya-Nyaya," *Philosophy East & West* 40 (1990), 129–40.
- Chattopadhyaya, D. and Gangopadhyaya, M.: *Nyaya Philosophy with Vatsyayana Bhasya* (Calcutta: 1975).
- Visvanatha: *Bhasapariccheda* with *Karikavali*, trans. S. Madhavananda (Calcutta: 1951).

J. N. MOHANTY



**objective/subjective** The contrast between the subjective and the objective is made in both the epistemic and the ontological domains. In the former it is often identified with the distinction between the *intrapersonal* and the *interpersonal*, or with that between matters whose resolution depends on the psychology of the person in question and those not thus dependent, or, sometimes, with the distinction between the biased and the impartial. Thus, an objective question might be one answerable by a method usable by any competent investigator, while a subjective question would be answerable only from the questioner's point of view. In the ontological domain, the subjective-objective contrast is often between what is and what is not mind-dependent; secondary qualities, e.g. colours, have been thought subjective owing to their apparent variability with observation conditions. The truth of a proposition, for instance (apart from certain propositions about oneself), would be objective if it is independent of the perspective, especially the beliefs, of those judging it. Truth would be subjective if it lacks such independence, say because it is a construct from justified beliefs, e.g. those well confirmed by observations.

One notion of objectivity might be basic and the other derivative. If the epistemic notion is basic, then the criteria for objectivity in the ontological sense derive from considerations of justification: an objective question is one answerable by a procedure that yields (adequate) *justification* for one's answer; and mind-independence is a matter of amenability to such a method. If, on the other hand, the ontological notion is basic, the criteria for an interpersonal method and its objective use are a matter of its mind-independence and tendency to lead to objective truth, say its applying to external objects and

yielding predictive success. Since the use of these criteria requires employing the methods which, on the epistemic conception, define objectivity – most notably scientific methods – but no similar dependence obtains in the other direction, the epistemic notion is often taken as basic.

In epistemology, the subjective-objective contrast arises above all for the concept of justification and its relatives. Externalism, particularly reliabilism, construes justification objectivistically, since, for reliabilism, truth-conduciveness (non-subjectively conceived) is central for justified belief (*see* EXTERNALISM/INTERNALISM; RELIABILISM). Internalism may or may not construe justification subjectivistically, depending on whether the proposed epistemic standards are interpersonally grounded (say, *a priori*). There are also various kinds of subjectivity; justification may, e.g., be grounded in one's considered standards or simply in what one believes to be sound. On the former view, my justified beliefs accord with my considered standards whether or not I think them justified; on the latter, my thinking them justified makes it so.

Any conception of objectivity may treat one domain as fundamental and the others derivatively. Thus, objectivity for methods (including sensory observation) might be thought basic. Let an objective method be one that is (1) interpersonally usable and tends to yield justification regarding the questions to which it applies (an epistemic conception), or (2) tends to yield truth when properly applied (an ontological conception), or (3) both. Then an objective person is one who appropriately uses objective methods; an objective statement is one appraisable by an objective method; an objective discipline is one whose methods are objective; and so on.

Typically, those who conceive objectivity epistemically tend to take methods as fundamental; those who conceive it ontologically tend to take statements as basic.

*See also* IN ITSELF/FOR ITSELF; NOUMENAL/PHENOMENAL; OBJECTIVITY; PRIMARY AND SECONDARY QUALITIES; SUBJECTIVITY.

#### BIBLIOGRAPHY

- Audi, R.: "Scientific Objectivity and the Evaluation of Hypotheses," in *The Philosophy of Logical Mechanism*, ed. M. H. Salmon (Dordrecht: Kluwer, 1990).
- Chisholm, R. M.: *The Foundations of Knowing* (Minneapolis, MN: University of Minnesota Press, 1982).
- Foley, R.: *The Theory of Epistemic Rationality* (Cambridge, MA: Harvard University Press, 1987).
- Nagel, E.: *The Structure of Science* (New York: Harcourt, Brace and World, 1961).
- Weber, M.: *The Methodology of the Social Sciences* (Glencoe, IL: Free Press, 1949).

ROBERT AUDI

**objectivity** Among the various notions of objectivity that philosophers have investigated and employed, two can claim to be fundamental.

On the one hand, there is a straightforwardly ontological concept: something is objective if it exists, and is the way it is, independently of any knowledge, perception, conception or consciousness there may be of it. Obvious candidates here include plants, rocks, atoms, galaxies and other material denizens of the external world. Less obvious candidates include such things as numbers, sets, propositions, primary qualities, facts, time and space. Subjective entities, conversely, will be those which could not exist or be the way they are if they were not known, perceived or at least conceived by one or more conscious beings. Such things as sensations, dreams, memories, secondary qualities, aesthetic properties, and moral values have been construed as subjective in this sense. (I shall call this ontological notion O-objectivity.)

There is, on the other hand, a notion of objectivity that belongs primarily within epistemology. According to this conception, the objective/subjective distinction is not intended to mark a split in reality between autonomous and dependent entities, but serves rather to distinguish between two grades of cognitive achievement. In this sense only such things as judgements, beliefs, theories, concepts and perceptions can significantly be said to be objective or subjective. Here objectivity can be construed as a property of the contents of mental acts and states. We might say, for example, that a belief that the speed of light is 187,000 miles per second, or that Leeds is to the north of Sheffield, has an objective content; a judgement that rice pudding is disgusting, on the other hand, or that Beethoven is a greater artist than Mozart, will be merely subjective. (I shall call this epistemological concept E-objectivity.)

If E-objectivity is to be a property of contents of mental acts and states, then at this point we clearly need to specify *which* property it is to be. This is a delicate matter; for what we require here is a minimal concept of objectivity, one that will be neutral with respect to the competing and sometimes contentious philosophical theories which attempt to specify what objectivity is. In principle this neutral concept will then be capable of comprising the pre-theoretical datum to which the various competing theories of objectivity are themselves addressed, as attempts to supply an analysis and explanation. Perhaps the best notion is one that exploits Kant's insight that E-objectivity entails what he calls "presumptive universality": for a judgement to be objective it must at least possess a content that "may be presupposed to be valid for all men". (See Kant, 1953, §19).

It is worth noting that an entity that is O-subjective can be the subject of E-objective judgements and beliefs. For example, on most accounts colours are O-subjective: in the analysis of the property of being red, say, there will occur ineliminable appeal to the perceptions and judgements of normal observers under normal conditions. And yet the judgement that a given object is red is an entirely objective one. Rather more bizarrely,

Kant argued that space was nothing more than the form of inner sense, and so was O-subjective. And yet the propositions of geometry, the science of space, are for Kant the very paradigms of E-objectivity; for they are necessarily, universally and objectively true. One of the liveliest debates in recent years (in logic, set theory, the foundations of mathematics, the philosophy of science, semantics and the philosophy of language) concerns precisely this issue: does the E-objectivity of a given class of assertions require the O-objectivity of the entities those assertions apparently invoke or range over? By and large, theories that answer this question in the affirmative can be called *realist*; those that defend a negative answer, *anti-realist*.

One intuition that lies at the heart of the realist's account of objectivity (see *REALISM*) is that, in the last analysis, the objectivity of a belief is to be explained by appeal to the independent existence of the entities it concerns: E-objectivity, that is, is to be analysed in terms of O-objectivity. A judgement or belief is E-objective if and only if it stands in some specified relation to an independently existing, determinate reality. Frege, for example (see *FREGE*), believed that arithmetic could comprise objective knowledge only if the numbers it refers to, the propositions it consists of, the functions it employs, and the truth-values it aims at, are all mind-independent entities. And conversely, within a realist framework, to show that the members of a given class of judgements are merely subjective, it is sufficient to show that there exists no independent reality that those judgements characterize or refer to. Thus J. L. Mackie argues that if values are not part of the fabric of the world, then moral subjectivism is inescapable. For the realist, then, E-objectivity is to be elucidated by appeal to the existence of determinate facts, objects, properties, events, and the like, which exist or obtain independently of any cognitive access we may have to them. And one of the strongest impulses towards platonic realism – the theoretical commitment to the existence of abstract objects like sets, numbers, and propositions – stems from the widespread belief that only if such things

exist in their own right can we allow that logic, arithmetic, and science are indeed objective (see *MATHEMATICAL KNOWLEDGE*).

This picture is rejected by anti-realists. The possibility that our beliefs and theories are objectively true is not, according to them, capable of being rendered intelligible by invoking the nature and existence of reality as it is in and of itself. If our conception of E-objectivity is minimal, requiring only "presumptive universality", then alternative, non-realist analyses of it can seem possible – and even attractive. Such analyses have construed the objectivity of an arbitrary judgement as a function of its coherence with other judgements, of its possession of grounds that warrant it, of its acceptance within a given community, of its conformity to the a priori rules that constitute understanding, of its verifiability (or falsifiability), or of its permanent presence in the mind of God. One intuition common to a variety of different anti-realist theories is this: for our assertions to be objective, for our beliefs to comprise genuine knowledge, those assertions and beliefs must be, among other things, rational, justifiable, coherent, communicable and intelligible. But it is hard, the anti-realist claims, to see how such properties as these can be explained by appeal to entities "as they are in and of themselves"; for it is not on the basis of their relation to any such things as these that our assertions become intelligible, say, or justifiable. On the contrary, according to most forms of anti-realism, it is only on the basis of O-subjective notions like "the way reality seems to us", "the evidence that is available to us", "the criteria we apply", "the experiences we undergo" or "the concepts we have acquired" that the E-objectivity of our beliefs can possibly be explained.

In addition to marking the ontological and epistemic contrasts mentioned already, the objective/subjective distinction has also been put to a third use, namely to differentiate intrinsically perspectival from non-perspectival points of view. An objective, non-perspectival view of the world finds its clearest expression in sentences that are devoid of demonstrative, personal, tensed or other token reflexive

elements. Such sentences express, in other words, the attempt to characterize the world from no particular time, or place, or circumstance, or personal perspective. Nagel calls this “the view from nowhere”. (Anyone might say: “At 15.30 G.M.T., on Tuesday, 6 May 1991, Mr J. Smith gives Mr R. Brown the sum of £5.00. Mr Brown’s immediate emotion is one of delight”.) A subjective point of view, by contrast, is one that possesses characteristics determined by the identity or circumstances of the person whose point of view it is. (In the envisaged situation, only Brown can say: “It is now 15.30, and that man over there just gave me £5.00. I’m delighted!”) The philosophical problems here centre on the question whether there is anything that an exclusively objective description would necessarily fail to reveal about oneself of the world. Can there, for instance, be a language with the same expressive power as our own, but which lacks all token reflexive elements? (Perry, 1979). Or, more metaphysically, are there genuinely and irreducibly subjective aspects to my existence – aspects which belong only to my unique perspective on the world and which must therefore resist capture by any purely objective conception of that world? (Nagel, 1986, *passim*; Wittgenstein, 1922, 5.6–5.641, 6.4–7; Wittgenstein, 1953, §§241–420).

#### HISTORICAL NOTE

The terms “objective” and “subjective” are used, above, in their contemporary sense. This usage is relatively recent, however, appearing in English for the first time in the nineteenth century as the translation of the Kantian terms *objektiv* and *subjektiv*. Kant was in fact the first major philosopher to use these terms in what is recognizably their modern sense. (See Kant, 1781, B.44.) For four centuries before Kant – in the writings of the scholastics, Descartes, Spinoza, Berkeley and others – the contrast between objectivity and subjectivity had a quite different sense: one that is, confusingly, more or less the converse of the modern one. In the pre-Kantian sense, for

something to be objective, or for it to exist objectively, was for it to comprise an idea, a mental representation, a content of consciousness. Conversely, the adjective “subjective” was reserved, in Descartes’s words, for “the reality that philosophers call actual or formal” (Descartes, 1641, III, §II). According to this usage for something to exist actually, formally or subjectively was for it to exist independently of any mental representation there might be of it. Anachronistically, Franz Brentano’s notion of objectivity (*Gegenständlichkeit*) conforms to the older usage. He takes objectivity to be an intrinsic property of mental acts – the property, namely, of possessing an intentional content. His pupils Edmund Husserl and Alexius Meinong followed him in employing a notion of objectivity which owes more to medieval scholasticism than to Kant’s critical philosophy.

See also IN ITSELF/FOR ITSELF; NOUMENAL/PHENOMENAL; OBJECTIVE/SUBJECTIVE; REALISM; SUBJECTIVITY.

#### BIBLIOGRAPHY

- Descartes, R.: *Meditations* (1641), in *The Philosophical Writings of Descartes*, ed. J. Cottingham, R. Stoothoff and D. Murdoch (Cambridge: Cambridge University Press, 1975).
- Frege, G.: “Thoughts,” in *Collected Papers on Mathematics, Logic and Philosophy*, ed. B. McGuinness (Oxford: Blackwell, 1984), 351–72.
- Kant, I.: *Critique of Pure Reason* (1781), trans. N. Kemp Smith (London: Macmillan, 1964).
- Kant, I.: *Prolegomena to Any Future Metaphysics* (1783), trans. L. W. Beck (Indianapolis, IN: Bobbs-Merrill, 1950).
- Kripke, S.: “Wittgenstein on Rules and Private Language: An Elementary Exposition,” in *Perspectives on the Philosophy of Wittgenstein*, ed. I. Block (Oxford: Blackwell, 1981), 238–312.
- Mackie, J. L.: *Ethics: Inventing Right and Wrong* (Harmondsworth: Penguin Books, 1977).
- McGinn, C.: *The Subjective View* (Oxford: Clarendon Press, 1983).
- Nagel, T.: *The View from Nowhere* (New York: Oxford University Press, 1986).
- Perry, J.: “The Problem of the Essential Indexical,” *Nous* 13 (1979), 3–21.

Schlick, M.: "Meaning and Verification," in *Readings in Philosophical Analysis*, ed. H. Feigl and W. Sellars (New York: Appleton-Century-Crofts, 1949), 146–70.

Wittgenstein, L.: *Tractatus Logico-Philosophicus* (1922), trans. D. F. Pears and B. McGuinness (London: Routledge and Kegan Paul, 1974).

Wittgenstein, L.: *Philosophical Investigations*, trans. G. E. M. Anscombe (Oxford: Blackwell, 1953).

DAVID BELL

### **Ockham, William of (c.1285–1347)**

English philosopher. While Ockham intended to follow Aristotle in philosophy of mind, cognitive psychology, and epistemology, he in fact dismantled traditional interpretations in the first two areas. With Scotus against Aquinas (see AQUINAS), Ockham affirms a plurality of substantial forms for living substances; in humans, the forms of corporeity, sensory soul and intellectual soul, all really distinct from one another (where Scotus had regarded the latter two as only formally distinct) (*Rep* II, q.7; *OTH* v 137; cf. *Rep* III, q.3; *OTH* vi 118). Further, he insists, the cognitive and appetitive powers of the sensory and intellectual forms are neither really distinct inherent accidents (as Aquinas claimed; *Rep* II, q.20; *OTH* v 425–9; *Rep* III, q.4; *OTH* vi 131–4) nor formally distinct from one another and the soul (as Scotus said; *Rep* II, 1.20; *OTH* v 433–5), but in every way really the same. Thus "intellect" is a connotative term signifying the substance of the soul able to understand, while "will" signifies the same substance able to will (*Rep* II, q.20; *OTH* v 435). If "sensory power" means "that in the soul which elicits the act as its partial cause", the sensory powers are really the same as the sensory soul: the part that inheres in the eye is the visual power, that in the ear, the auditory power; etc. While all these parts are alike, their different functions are the result of different accidental dispositions in the various sense organs (*Rep* III, q.4; *OTH* vi 138–9).

Likewise, Ockham scraps traditional Aristotelian cognitive psychology. Aquinas had endorsed the tag "like is known by like", and insisted that every knower must have a species – a mental representation distinct

from both the cognitive act and its object – by means of which s/he is assimilated to the object known. In human beings, all knowledge begins with sensation, which involves the reception of sensible species, which in turn gives rise to phantasms, which in turn require to be "illuminated" by the agent intellect to produce the intelligible species received into the possible intellect. Ockham replaces this whole scheme with his theory of intuitive and abstractive cognition and his psychology of habit. If knowledge must still be "by assimilation", it is enough that the act of apprehending itself resemble its object in some relevant way. For both angels and humans, intuitive cognitions are caused by the object and the intellect, together with the general concurrence of God (*Rep* II, qq.12–13; *OTH* v 268–9, 274); no causal role remains for species to play (*ibid.*; *OTH* v 305). For abstractive cognitions (e.g. of memory or imagination), something more is needed – a habit which actively inclines the power to produce acts of apprehending the same objects under the same aspects (*ibid.*; *OTH* v 269–71). If human knowledge begins with sensation, and there are no *ante-mortem* intellectual intuitive cognitions of material things apart from a corresponding sensory cognition, the latter is related to the former only as a partial efficient cause (*Rep* II, qq.12–13; *OTH* v 302–3; cf. *Prologue*, q.1; *OTH* i, 27, 64). Created intellects further act to produce abstract general concepts, to formulate necessary and contingent propositions, and to reason discursively. Because for Ockham particulars are fully intelligible (so that the first intellectual intuitive cognitions are of the same object under the same aspect as sensory ones), there is no need to posit any "agency" to "illuminate" enmattered, particularized mental representations, to abstract intelligible content from them. Hence, the traditional distinction between agent and possible intellect loses much of its point: alike connotative, "agent intellect" signifies the intellect *qua* partial efficient cause of its actions, while "possible intellect" signifies the intellect *qua* subject that receives those actions (*Rep* II, qq.12–13; *OTH* v 304–5). Ockham finds his opponents' case for species stronger where after-images and



sensory imagination are concerned, but still prefers to fill explanatory gaps in other ways (*Rep* III, 1.3; *OTH* vi 105–21).

Despite an undeserved reputation as father of scepticism, Ockham is a fully conventional Aristotelian in epistemology (see RELIABILISM). Taking for granted the general reliability of our cognitive faculties, Ockham assumes that we have certain knowledge of mind-independent material things (*Quodl* 1, q.15; *OTH* ix 83; *Prologue*, q.1; *OTH* i 23) as well as of our own present mental acts (*Prologue*, q.1; *OTH* i 40, 43), and proceeds to trace such evident judgements to a distinctive species of acts of awareness – intuitive cognitions (*Prologue*, q.1; *OTH* i 31–2; *Rep* II, qq.12–13; *OTH* v 256–8) – which (by contrast with abstractive cognitions of the same things) are the power to produce evident judgements concerning their objects. Similarly, where our sensory faculties are concerned, Ockham appeals to the Aristotelian maxim that where the sensory faculty is well disposed and close enough to its object, and where conditions in the medium are suitable, sense does not err with respect to its proper object (*Rep* III, q.2; *OTH* vi 57).

Like other Aristotelian reliabilists, Ockham admits that the operation of such cognitive powers can be *naturally obstructed*. He accepts in a matter of fact way that while we have evident knowledge of sensible things (e.g. that S is white) through intuitive cognitions, sometimes the flow of information gets blocked through imperfections in the object or in intuitive cognition or through obstacles in the medium. Ockham also allows that demons and God can deceive us in various ways. This admission is no daring Cartesian innovation, however, but rather accommodation to his colleagues' interpretation of Divine omnipotence in the wake of the Condemnation of 1277. Indeed, like Aristotle and Scotus, Ockham does not understand certainty (see CERTAINTY) to involve freedom from the natural or metaphysical or logical possibility of error, but absence of actual doubt and error. For him, the arguments of the Academic sceptics no longer merit serious attention, having been sufficiently refuted by Scotus before him.

## WRITINGS

Guillelmi de Ockham: *Opera Philosophica et Theologica* (St Bonaventure, NY: The Franciscan Institute), vols I–VI and IX–X.

## BIBLIOGRAPHY

- Adams, M. McCord: *William Ockham* (Notre Dame, IN: University of Notre Dame Press, 1987), chs 13–14.  
 Day, S.: *Intuitive Cognition: A Key to the Significance of the Later Scholastics* (St Bonaventure, NY: The Franciscan Institute, 1947).  
 Fuchs, O.: *The Psychology of Habit According to William Ockham* (St Bonaventure, NY: The Franciscan Institute, 1952).  
 Tachau, K. H.: *Vision and Certitude in the Age of Ockham: Optics, Epistemology, and the Foundations of Semantics 1250–1345* (Leiden: E. J. Brill, 1988).

MARILYN MCCORD ADAMS

**ontological commitment** A criterion of ontological commitment is a rule determining which things or sorts of things there must be if a given theory is true; these are the things to whose existence the theory is committed. It cannot settle whether there really are such things as numbers or sense data, but only whether such things are part of the ontological baggage of some theory or theorist. Since we understand names which fail to refer, and acknowledge many objects which lack names, we cannot rely upon the use of names for this purpose, but must find a locution which always signals an ontological commitment. The most famous criterion is Quine's (see QUINE). If we formulate our views in a canonical notation based upon first order logic, we are ontologically committed to an object if and only if it must be among the values of the variables if all of the statements we accept are true. This rests upon Quine's controversial view that the existential quantifier captures the notion of existence.

See also EXISTENCE; ONTOLOGICAL RELATIVITY.

BIBLIOGRAPHY

- Alston, W.: "Ontological Commitment," *Philosophical Studies* 6 (1958), 8–17.  
 Church, A.: "Ontological Commitment," *Journal of Philosophy* 55 (1958), 1008–14.  
 Quine, W. V.: *From a Logical Point of View* (Cambridge, MA: Harvard University Press, 1961), ch. 1.  
 Quine, W. V.: *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969), ch. 4.

CHRISTOPHER HOOKWAY

**ontological relativity** If Quine's thesis of the indeterminacy of reference were true (see QUINE; INDETERMINACY OF REFERENCE), then there would be no fact of the matter what the ontological commitments of a given theory were. Depending upon the background language in which we try to state a theory's ontological commitments and, most important, the choice of a translation manual from the language of the theory into this background language, Quine's criterion of ontological commitment would yield different results. The thesis of ontological relativity, which Quine admits is not really distinct from the indeterminacy of reference, holds that claims about ontological commitment are thus doubly relative to background language and to the interpretation of the theory in the background language. In reflecting on our own ontological commitments, such indeterminacy and relativity are disguised since we use the "identity transformation" as our translation manual, translating "rabbit" as "rabbit", and so on.

See also INDETERMINACY OF REFERENCE; ONTOLOGICAL COMMITMENT; PRINCIPLE OF CHARITY; RELATIVISM; SOCIAL SCIENCES.

BIBLIOGRAPHY

- Quine, W. V.: *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969), ch. 2.  
 Quine, W. V.: *Pursuit of Truth* (Cambridge, MA: Harvard University Press, 1990), ch. 2.

CHRISTOPHER HOOKWAY

**ostensive definition** An ostensive definition is an explanation of the meaning of a word typically involving three elements: (1) an ostensive gesture, (2) an object pointed at which functions as a sample, (3) the utterance "This is (a) W". Like other forms of explanation of word-meaning, an ostensive definition functions as a rule or standard of correctness for the application of a word. The utterance "This is W" when employed in giving an ostensive definition does not describe an object (i.e. the thing pointed at) as having the property W, but defines a word. It is most illuminatingly viewed as providing a kind of substitution-rule in accord with which one symbol, e.g. "red", is replaceable by a complex symbol consisting of utterance ("This" or "This colour"), gesture, and sample. Hence instead of "The curtains are red" one can say "The curtains are *this* colour", while pointing at a red sample. The ostensive definition specifies that anything which is *this* is correctly characterized as being W.

Like all definitions, ostensive definitions are misinterpretable. One way of warding off misunderstanding is to specify the "grammatical signpost" by which the definiendum is stationed, i.e. to give the logico-grammatical category to which it belongs, viz. "This C is W", where "C" is a place-holder for, e.g. "colour", "length", "shape", "weight". Like all rules, an ostensive definition does not provide its own method of application. Understanding an ostensive definition involves grasping the "method of projection" from the sample to what it represents or from the ostensive gesture accompanying the definition to the application of the word. Thus in the case of defining a length by reference to a measuring rod, one must grasp the method of laying the measuring rod alongside objects to determine their length before one can be said to grasp the use of the definiendum. Ostensive definitions fulfil a crucial role both in explaining word meaning and in justifying or criticizing the application of that word (e.g. "Those curtains are not ultramarine – *this* colour is ultramarine [pointing at a colour chart] and the curtains are not *this* colour"). An ostensive definition does not give evidential grounds for the application of a word

“W”, but rather specifies what counts as being W.

The boundaries of the notion of ostensive definition are vague. A definition of a smell, taste or sound by reference to a sample typically involves no deictic gesture but a presentation of a sample (by striking a keyboard, for example). Conversely, defining directions (for example, “north”) by a deictic gesture involves no sample. Nor is the form of words “This is (a) W” essential. “This is called ‘W’” or “W is this C” can fulfil the same role.

Whether something functions as a sample (or paradigm) for the correct application of a word is not a matter of its essential nature, but of human choice and convention. Being a sample is a role conferred upon an object momentarily, temporarily or relatively permanently by us – it is a use to which we put the object. Thus we can use the curtains here and now to explain what “ultramarine” means – but perhaps never again, although we may often characterize (describe) them as being ultramarine. Or we can use a standard colour chart to explain what “ultramarine” means, although if it is left in the sun and fades, it will no longer be so used. Or we may establish relatively permanent canonical samples, as was the case with the Standard Metre bar. A sample *represents* that of which it is a sample, and hence must be typical of its kind. It can characteristically be copied or reproduced and has associated with it a method of comparison. It is noteworthy that one and the same object may function now as a sample in an explanation of meaning or evaluation of correct application and now as an item described as having the defined property. But these roles are exclusive in as much as what functions as a norm for description cannot simultaneously be described as falling under that norm. *Qua* sample the object belongs to the means of representation and is properly conceived as belonging to grammar in an extended sense of the term. Hence the Standard Metre bar cannot be said to be (or not to be) one metre long. Furthermore, one and the same object may be used as a defining sample for more than one expression. Thus a black patch on a colour chart may serve both to explain what “black” means

and as part of an explanation of what “darker than” means.

Although the expression “ostensive definition” is modern philosophical jargon (W. E. Johnson, *Logic* 1921) the idea of ostensive definition is venerable. It is a fundamental constituent of what Wittgenstein (see WITTGENSTEIN) called “Augustine’s picture of language” in which it is conceived as the fundamental mechanism whereby language is “connected with reality”. The mainstream philosophical tradition has represented language as having a hierarchical structure, its expressions being either “definables” or “indefinables”, the former constituting a network of lexically definable terms, the latter of simple, unanalysable expressions that link language with reality and that inject “content” into the network. Ostensive definitions thus constitute the “foundations” of language and the terminal point of philosophical analysis, correlating primitive terms with entities which are their meanings. On this conception, ostensive definition is privileged: it is final and unambiguous, settling all aspects of word use – the grammar of the definiendum being conceived to flow from the nature of the entity with which the indefinable expression is associated. In classical empiricism (see EMPIRICISM) definables stand for complex ideas, indefinables for simple ideas that are “given” in experience. Accordingly the “given” is mental in nature, the linking mechanism is private “mental” ostensive definition, and the basic samples, stored in the mind, are ideas which are essentially epistemically private and unshareable (cf. Locke, *Essay* II, XI, 9).

Wittgenstein, who wrote more extensively on ostensive definition than any other philosopher, held this picture of language to be profoundly misleading. Far from samples being “entities in reality” to which indefinables are linked by ostensive definition, they themselves belong to the means of representation. In that sense, there is no “link between language and reality”, for explanations of meaning, including ostensive definitions, remain within language. Ostensive definitions are not privileged but are as misinterpretable as any other form of explanation. The objects pointed at are not “simples” that

constitute the ultimate metaphysical constituents of reality, but samples with a distinctive use in our language-games. They are not the meanings of words, but instruments of our means of representation. The grammar of a word ostensively defined does not flow from the essential nature of the object pointed at, but is constituted by *all* the rules for the use of the word, of which ostensive definition is but one. It is a confusion to suppose that expressions *must* be explained exclusively either by analytic definition (definables) or by ostension (indefinables), for many expressions can be explained in both ways, and there are many other licit forms of explanation of meaning. The idea of "private" or "mental" ostensive definition is wholly misconceived, for there can be no such thing as a rule for the use of a word which cannot logically be understood or followed by more than one person, there can be no such thing as a logically private sample nor any such thing as a mental sample (*see* PRIVATE LANGUAGE ARGUMENT).

Apart from these negative lessons, a correct conception of ostensive definition by reference to samples resolves the venerable puzzles of the alleged synthetic *apriority* of colour exclusion (e.g. that nothing can be simultaneously red and green all over) and of the nature of the necessity of such apparently metaphysical propositions as "black is darker than white". Such "necessary truths" are indeed not derivable from explicit definitions and the laws of logic alone – i.e. are not analytic (*see* ANALYTICITY) – but nor are they descriptions of the essential natures of objects in reality. They are rules for the use of colour words, exhibited in our practices of explaining and applying words defined by reference to samples. What we employ as a sample of red we do not also employ as a sample of green; and a sample of black can, in conjunction with a sample of white, also be used to explain what "darker than" means. What appear to be metaphysical propositions about essential natures are but the shadows cast by grammar.

*See also* DEFINITION; LINGUISTIC UNDERSTANDING; PRIVATE LANGUAGE ARGUMENT.

## BIBLIOGRAPHY

- Arrington, R. L.: "'Mechanism' and 'Calculus': Wittgenstein on Augustine's Theory of Ostension," in *Wittgenstein: Sources and Perspectives*, ed. C. G. Luckhardt (Ithaca, NY: Cornell University Press, 1979), 303–38.
- Baker, G. P. and Hacker, P. M. S.: "Wittgenstein and the Vienna Circle: the Exaltation and Deposition of Ostensive Definition," in *Ludwig Wittgenstein: Critical Assessments*, ed. S. Shanker, 4 vols (London: Croom Helm, 1986), vol. 1, pp. 241–62.
- Baker, G. P. and Hacker, P. M. S.: *Wittgenstein: Understanding and Meaning* (Oxford: Blackwell, 1980), 168–205, 284–96.
- Locke, J.: *An Essay Concerning Human Understanding* (1690), ed. P. H. Niddich (Oxford: Clarendon Press, 1975).
- Wittgenstein, L.: *Philosophical Grammar*, trans. A. Kenny (Oxford: Blackwell, 1974).
- Wittgenstein, L.: *Philosophical Investigations*, trans. G. E. M. Anscombe (Oxford: Blackwell, 1953).

P. M. S. HACKER

**other minds** The interest in the traditional philosophical problem about other minds lies in the question whether, given what would seem a natural way we have of thinking about the meaning of mental terms, we can even make sense of the sentences we use to attribute mental states to others. This distinguishes the problem from a less interesting sceptical problem about other minds, in which it is a purely epistemological question of how we know about other minds, on a par with the question of how we know about the external world.

The distinctive problem arises because there is a natural tendency to assume that our mental terms ("pain", "belief", etc.) get their meaning from our relating them to our own experiences (call this "The Assumption"). If we understand what it is to have a pain or a belief from our own case, a question arises as to whether such a perspective on them will even allow our talk of others having pains and beliefs to make sense. This perspective is a strictly first person perspective. The idea that these terms get their meaning from our own case is meant to emphasize that we have

an irreducibly subjective conception of pain, belief and so on. Descartes (*see* DESCARTES) proposed such a conception of mentality in his "Second Meditation", though he did compromise it somewhat in other places by talking of the mental as a substance and, therefore, as an objective thing. Husserl (1960) (*see* HUSSERL) later embraced a more uncompromising version of the Cartesian first person point of view. As Wittgenstein's (1953) discussion of these issues brought out (*see* WITTGENSTEIN), all such positions are subjectivist in the strong sense that they rest on the claim that terms get their meaning from logically private states or episodes. The stricter versions of phenomenism (*see* PHENOMENALISM), where all things in the empirical domain which are not themselves sense-data (*see* SENSE-DATA) are logical constructs (*see* LOGICAL CONSTRUCTION) out of (and not just entities inferred from) sense-data, also give rise to the distinctive problem of other minds. This doctrine allows one to retain the intelligibility of our concept of material objects by reducing them to one's own sense-data, but it is not clear that it allows one to find intelligible the idea of another *mind* since all we have to work with are our own sense-data and logical constructs out of them. Such sense-data theories are no longer current, but Thomas Nagel (1986) has done much to revive the uncompromising Cartesian and Husserlian version of the problem by stressing the irreducibility of the first person point of view.

There are, as one might expect, two strategies of response to such a problem. One is to reject The Assumption, which underlies it; the other is to accept The Assumption and then struggle to solve the problem.

The latter course is most famously pursued in the argument from analogy (*see* ARGUMENT FROM ANALOGY), which was first formulated by Mill (1867) (*see* MILL). When faced with the problem of other minds, it is tempting to say first: The very same thing that I describe when I say "John is in pain" is what John describes when he says "I am in pain". But it is not clear what gives us the right to say this if words like "pain" are defined in terms of The Assumption. The argument from analogy is

intended to help provide that right. It proceeds from the observation of a causal connection between one's own mental states and one's bodily behaviour, and the observation of similar forms of behaviour in other bodies, to a conclusion about mental states of others similarly causing their behaviour.

The deepest criticism of this argument is due to Wittgenstein (1953) who, in effect, notices that if one takes The Assumption seriously, then the sense of "mind" or "mental states" that is being delivered up for others by the argument is not going to be the same as has been defined in one's own case. Whatever it is in one's own case that confers meaning on them is, *ex hypothesi*, not available when one is applying those terms to others. The argument, thus, does not even as much as address the problem raised by The Assumption; and it is not clear that anything can, for if the states which confer meaning are logically private then nothing could bridge the gap which gives rise to the problem. There are other less interesting criticisms of the argument from analogy (such as that it is an induction on the basis of just one case), but they seem to proceed without recognizing the fact that the *distinctive* problem of other minds is not even addressed by the argument.

Wittgenstein extends his critique to The Assumption itself, and in doing so, takes up the second strategy of response to the problem. In his well-known attack on the idea of a private language (*see* PRIVATE LANGUAGE ARGUMENT), he denies the plausibility of the suggestion that we can define mental terms by relating them to our own thoughts and experiences, since that procedure allows for no possibility of one's ever being able to make a mistake in the application of the terms. Whatever seems to us to be a right application would be a right application. Unless there is a more publicly available source of the meaning of mental terms, one would not have any standard for their incorrect (and therefore also for their correct) use; and without that, it would be quite wrong to say that we have a mastery of the use of these terms.

Strawson (1959) also raises an interesting objection to The Assumption. He denies that anyone possesses a concept or has a mastery



of terms unless he or she can apply the term to more than one thing. Gareth Evans has called this the “generality constraint” which demands that “any thought which . . . has the content ‘a is F’ involves the exercise of an ability which can be exercised in indefinitely many distinct thoughts, and would be exercised in, for instance, the thought that ‘b is F’” (Evans, 1982, p. 101). Strawson specifically explores the effect of imposing this constraint on the concept of pain. The constraint forbids attributing to someone the thought expressed by “I am in pain” unless he or she is also possessed of the idea of someone being in pain, where that someone need not be himself or herself. This does not allow the distinctive problem of other minds to arise since it rejects the coherence of The Assumption (that we can define mental terms on the basis of just our own experiences).

These arguments have not convinced everyone. But, whether one has an effective argument against The Assumption, the fact still remains that it raises an insuperable problem of other minds, a problem which can only be removed if one at least provides some cogent alternative to it.

Wittgenstein’s proposed alternative to the rejected Assumption is that mental terms get their meaning due to a special connection that holds between them and the publicly observable behaviour of agents who possess mental states. He therefore shifts the emphasis to a third person conception. This shift has been favoured by a number of philosophers since Wittgenstein. Though there are very important differences between them, when the direction is described as generally as I have, it is fair to say that philosophers as diverse as Ryle (1949), Quine (1960), Davidson (1984), and a substantial infantry marching under the influence of the doctrine of functionalism have all been following it (e.g. Putnam, 1975).

However problematic the first person perspective on mentality might be, the fact that it has some intuitive appeal should make us expect that simply reversing The Assumption to a third person mode of defining mental terms would raise various difficult problems of its own.

The two problems which have received the most attention in recent philosophy are first, how knowledge of others is related to self-knowledge, and second, how exactly the special connection between behaviour and mentality must be conceived in this third person characterization of the mind.

The first problem arises because with the shift in direction it would now seem that when we know our own mental states we do so on the basis of our behaviour – which is highly implausible and counterintuitive. A basic dilemma thus emerges: if one accepts The Assumption, one has a seemingly insuperable problem of other minds, and if one rejects it and adopts a more third person perspective, that leads to unintuitive consequences about knowledge of our own minds.

Gilbert Ryle (see RYLE) was so fearful of the first person perspective, and of the Cartesian idea that one knows one’s mental states by an inner observation or introspection (see INTROSPECTION) of items in one’s consciousness, that he accepted this highly implausible and unintuitive consequence of the alternative view. But Wittgenstein himself did not. He avoided it by denying that reports made in the first person of states such as pain were genuinely reports, and by denying that assertions such as “I know that I am in pain” were well-formed knowledge claims. My assertion “I am in pain” is rather to be thought of as an expression of my pain, albeit a rather more schooled expression than a cry or a grimace (see AVOWALS).

The dilemma, as we saw, emerged as a result of a thorough-going asymmetry that afflicts the epistemology of mind: we seem to know other minds on the basis of behaviour but not – usually – *our* own minds. Strawson saw this asymmetry as an essential feature of the very idea of creatures with minds, of “persons”. But what is needed is not simply an acknowledgement of the feature but some effort to explain it or to illuminate it by situating it in a wider context. And, moreover, to do so, without relinquishing the third person account of mentality so that we are not landed again with the problem of other minds. Ryle’s response, as we saw, is to deny the feature, Wittgenstein’s to explain it by

claiming that self-knowledge unlike knowledge of others is not a cognitive achievement. But there may be less drastic ways of dealing with the asymmetry within a basically third person point of view of mental terms.

A response due to David Armstrong (1968) is that agents possess a special mechanism of iterated belief-acquisition such that whenever the agent acquires a mental state he or she also acquires a belief that it has that mental state. This is an empirical thesis about how mental states cause one to have beliefs about them. It does not submit to the Rylean idea that we know our own mental states by inferring them from our behaviour, thereby capturing something special about self-knowledge missing in one's knowledge of others.

Some remarks of Gareth Evans' (1982, ch. 7) suggest another explanation of the asymmetry for a more restricted class of mental states, i.e. beliefs. He points out that when one is asked what one believes – as opposed to what another believes – about some subject matter, one does not survey one's own mind for the answer but rather one thinks about the subject matter itself. This seems to describe things accurately for the most part and it shows that the asymmetry need not take the form of making self-knowledge introspective and thereby spoil the basically third person perspective on mental terms. Evans' point also suggests the following further context in which the asymmetry emerges. If we find that someone's answer about another's belief does not square with that person's behaviour, we find ourselves saying that the answer was wrong. On the other hand, if it turns out that someone's answer about himself did not square with his behaviour, we would be just as inclined to say that the person was divided and irrational than that he had given the wrong answer – since his answer was a response to the subject matter rather than derived from an inner survey of his mind. This establishes a satisfying asymmetry between knowledge of others' beliefs and knowledge of one's own.

The second problem created by the shift in direction to the third person is about how the connection between behaviour and mental states ought to be conceived.

Wittgenstein himself introduces the term "criterion" to convey that the connection is not merely that one or other kind of behaviour is caused by one or other kind of mental state. Rather there is (also) a *conceptual* connection. Behavioural circumstances are the "criteria" rather than merely the effects or "symptoms" of mentality. If it were just a contingent connection, the sceptical worry would persist that, for all one knows, the behaviour we observe in others might be a sustained deceit, reflecting nothing mental at all – in much the same way that one might by a comprehensive sensory illusion come to believe falsely that there is an external world.

This idea of a "criterial" connection has been much discussed in the philosophical literature. The most interesting recent discussions may be found in McDowell (1982) and Wright (1984). Wright takes the notion of criterion to support an "anti-realist" view of the meanings of sentences attributing mental states to others. The view replaces the "realist" idea that the meanings of such sentences are given by the conditions in which they are true (*see* LINGUISTIC UNDERSTANDING) with the idea that they are given by conditions which warrant their assertion, i.e. given by their criteria. The truth-conditions for these sentences are undetectable by the person making the attributions, since the mental states being attributed (pain, say) are not themselves directly available to his experience in a way that they are to the subject of the attribution. Thus it cannot be knowledge of the truth-conditions which underlies his linguistic competence with these sentences, as the realist claims. What underlies his competence rather is knowledge of the criteria since the subject's behaviour is within the purview of the attributor's experience. Wright, following Wittgenstein, claims that there is a conceptual rather than a contingent link between the behaviour of the subject and the mental state because it is not on the basis of an inference (based on an empirical theory) that one goes from an observation of the behaviour to an attribution of the mental state. But Wright, following a widely held interpretation of Wittgenstein, also claims that the criteria are defeasible (*see* DEFEASIBILITY). That is,

it is possible that the criteria should be fulfilled but that the attribution of the relevant mental state turn out to be false.

McDowell raises two objections against this. He first points out that defeasibility spoils the non-inferentiality of criteria since the idea that an attribution based on the fulfilment of criteria could turn out to be false suggests that there is a logical gap between the behaviour and the mental state, just what the conceptual connection is supposed to be disallowing. It looks as if the gap suggests that it is an empirical theoretical connection between them, at best. The second objection has to do with certain modal notions that surround the concept of knowledge, in this case knowledge of another's mental states. When one claims to know that James, say, is angry, one cannot accept that, for all one knows, he is not angry. For a claim to know that  $p$  is incompatible with an admission that for all one knows,  $p$  is false. But if the criteria for  $p$  are defeasible, and all we know is whether the criteria are satisfied, it will always be true that, for all we know, not- $p$ . And so, if the relevant criteria are defeasible, we never know anything beyond them.

McDowell proposes instead to take criteria to be non-defeasible and to give up on an idea that seems to be shared by the anti-realist and the realist in Wright's exposition. This is the idea that a subject's mental state (and therefore the truth-condition of an attribution of it to him) is necessarily unavailable to the experience of another who is attributing the state to him. He thinks we may rightly say that one can directly perceive another person's pain (say) in his facial expressions and visible bodily states and behaviour. There is no need to see the behaviour, then, as a "proxy" for the real thing which is undetectable by direct experience. This means that the special connection between behaviour and mental states which defines the notion of a "criterion" is not to be expounded by saying that behaviour is a privileged kind of evidence. It is not evidence at all. "Evidence" suggests that it is better known than what it is evidence for, and that brings in the invidious idea of a "proxy". What makes the connection special is simply that there is no gap between the

two, and therefore there is no theoretical inference to be made from an observation of behaviour to an attribution of mental state. There is no gap because the mental state is taken in in the experience of the behaviour itself. It is not that one can never be wrong in making mental attributions to others, but that when one is wrong the criteria – despite appearances – have not been fulfilled.

McDowell takes this not only to be the right view of our knowledge of other minds but also to have been Wittgenstein's view. He says that Wittgenstein never thought that the behaviour of others was better-known than their mental states. Instead he always viewed "human beings" as basic, which is to say that he saw their behaviour and their mentality as more integrated than the traditional interpretation of his view of "criteria" has taken them to be.

There is an obvious and immediate response to McDowell's second objection to the more traditional interpretation of the notion of "criterion": criteria are only intended to provide the basis for claims to know something, not to knowing it. Such claims may be allowed later as being false. A full analysis of knowledge of a proposition merely requires that the knower has a criterial basis for claiming knowledge and that the proposition be true; it does not require that the knower have a guarantee of its truth. So the claim to knowledge can be withdrawn if it turns out that the proposition is not true.

The trouble with this response is that it seems to have surrendered to the sceptic, who can now claim that if criteria can only get us as far as claims to knowledge of other minds rather than to knowledge, then there is always the sceptical possibility that a claim to know that a person is in a certain mental state is false.

But is it really a surrender? Notice, first, that this sceptical position is no longer the distinctive position that we sketched at the outset. This is merely the less interesting epistemological scepticism. But putting that aside, even this less interesting scepticism (whether about other minds or the external world) is not correctly describable as insisting that any given knowledge claim could turn out

to be false. It claims, much more dramatically, that it is always a conceptual possibility that all our claims in a certain region (the external world, other minds) should turn out to be false. This shows that all that one has to do to oppose the sceptic is not to argue that particular claims to being knowledge should be guaranteed correct, but rather that there should be a way of preventing the slide from admitting that any given belief could be false to allowing that all beliefs in a certain region could be false. One may or may not be able to produce an argument that prevents the slide, but the point for now is that once the sceptic's position has been properly described, the modalities surrounding the concept of knowledge show that even if we conceive of criteria as defeasible we can still find some defence against the sceptic.

In any case, McDowell's own positive alternative view of criteria does not help against the sceptic either. It just pushes the sceptical issue to another place. McDowell does after all have to allow for error in mental attributions to others. The only difference is that he accounts for error not in terms of the defeasibility of criterial evidence but by saying that the criteria have not been fulfilled, even though they appear to have been fulfilled. But then it is always open to the sceptic to say that, *for all we know*, the criteria are never fulfilled, they only appear to be fulfilled.

McDowell's first objection is on stronger ground since it shows a serious tension between the twin insistence on non-inferentiality and defeasibility. But even here the response to the second objection suggests a way out. If the only reason to insist on the non-inferentiality is that allowing an inference would be to allow that any given attribution of a mental state could be false, then one may cease to insist on it since, as we just saw, allowing this need not amount to a surrender to the sceptic.

If, however, the defeasibility of criteria is purchased at the cost of allowing the inferentiality of the connection between criteria and ascription of mental states to others, then what is special about the connection between behaviour and mentality, which the notion of criteria is supposed to convey? Is it not now, like all inference, something based on empirical

theory? An answer – one which captures much of the spirit of Wittgenstein's conception of the mental – might be that what is special is that the theory upon which the inferences are based is necessarily a commonsense theory and not a scientific theory. The idea of an inferential but nevertheless *criterial* connection points to a mediating theory, which, like our theory of the middle-sized objects around us, is highly internalized in our ordinary, unscientific thinking about others' minds.

*See also* ARGUMENT FROM ANALOGY; CRITERIA AND KNOWLEDGE; INTROSPECTION; PRIVATE LANGUAGE ARGUMENT; SELF-KNOWLEDGE AND SELF-IDENTITY; SOLIPISM; TRANSCENDENTAL ARGUMENTS; WITTGENSTEIN.

#### BIBLIOGRAPHY

- Armstrong, D. M.: *A Materialist Theory of Mind* (London: Routledge and Kegan Paul, 1968).  
 Davidson, D.: *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984).  
 Descartes, R.: *Meditations on First Philosophy* (1641), in *The Philosophical Works of Descartes*, ed. J. Cottingham, R. Stoothoff and D. Murdoch (Cambridge: Cambridge University Press, 1975).  
 Evans, G.: *Varieties of Reference* (Oxford: Oxford University Press, 1982).  
 Husserl, E.: *Cartesian Meditations*, trans. D. Cairns (The Hague: Martinus Nijhoff, 1960).  
 McDowell, J.: "Criteria, Defeasibility and Knowledge," *Proceedings of the British Academy* 68 (1982), 455–79.  
 Mill, J. S.: *An Examination of Sir William Hamilton's Philosophy* (London: Longmans, 1867).  
 Nagel, T.: "Subjective and Objective," in his *Mortal Questions* (New York: Oxford University Press, 1979).  
 Quine, W. V.: *Word and Object* (Cambridge, MA: MIT Press, 1960).  
 Ryle, G.: *The Concept of Mind* (London: Hutchinson, 1949).  
 Strawson, P. F.: *Individuals* (London: Methuen, 1959).  
 Wittgenstein, L.: *Philosophical Investigations*, trans. G. E. M. Anscombe (Oxford: Blackwell, 1953).  
 Wright, C.: "Second Thoughts about Criteria," *Synthese* 58 (1984), 383–405.

AKEEL BILGRAMI

# P

---

**paradox** Somewhat loosely, a paradox is a compelling argument from unexceptionable premisses to an unacceptable conclusion; more strictly speaking, a paradox is specified to be a sentence that is true if and only if it is false. An example of the latter would be:

The displayed sentence is false.

It is easy to see that this sentence is false if true, and true if false. A paradox, in either of the senses distinguished, presents an important philosophical challenge. Epistemologists are especially concerned with various paradoxes having to do with knowledge and belief.

See also ANTINOMY; LOTTERY PARADOX; MOORE'S PARADOX; PARADOXES OF ANALYSIS; PREFACE PARADOX; PARADOX OF THE KNOWER.

## BIBLIOGRAPHY

- Quine, W. V.: "The Ways of Paradox," in his *The Ways of Paradox and other Essays* (Cambridge, MA: Harvard University Press, 1966).  
Sainsbury, R. M.: *Paradoxes* (Cambridge: Cambridge University Press, 1988).

JONATHAN VOGEL

**paradox of the knower** The knower paradox is an argument which begins with apparently impeccable premisses about the concepts of knowledge and inference and derives an explicit contradiction. The origin of the reasoning is the surprise examination paradox (see SURPRISE EXAMINATION PARADOX). Suppose your teacher announces that an examination will be given next week but that it will not be possible for you to predict

the day upon which it will occur. Apparently one can prove that it won't be on Friday: it wouldn't then be a surprise. Having ruled out Friday, one can infer that it won't be on Thursday, else again, no surprise. In a similar manner every day of the week can be eliminated, but when the exam is given on Wednesday you are surprised! By analysing this puzzle and focusing on a single day, Kaplan and Montague (1960) distilled the following "self-referential" paradox, the Knower. Consider the sentence:

- (S) The negation of this sentence is known (to be true).

Suppose that S is true. Then its negation is known and hence true. But if its negation is true, then S must be false. So we may conclude that if S is true, then S is false. Therefore S is false. Or, what is the same, the negation of S is true.

This conclusion does not depend on our initial assumption that S is true – that hypothesis has been discharged. So we have here a proof that the negation of S is true. And having thus given a rigorous proof it appears that we know that the negation of S is true. But now consider again what S says. Isn't this what we have just concluded? So S is true after all! It looks as if we have proved a contradiction: S is false and S is true.

This paradox and its accompanying reasoning are strongly reminiscent of the Liar Paradox which (in one version) begins by considering a sentence "This sentence is false" and derives a contradiction. Versions of both arguments using axiomatic formulations of arithmetic and Gödel-numbers to achieve the effect of self-reference yield important metatheorems about what can be



expressed in such systems. (See Montague, 1963, for rigorous and detailed formulations.) Roughly these are to the effect that no predicates definable in the formalized arithmetics can have the properties we demand of truth (Tarski's Theorem) or of knowledge (Montague, 1963).

These metatheorems still leave us with the following problem. Suppose we *add* to these formalized languages predicates intended to express the concepts of knowledge (or truth) and inference – as one might do if a logic of these concepts is desired. Then the sentences expressing the leading principles of the Knower Paradox will apparently be true. Explicitly, the assumptions about knowledge and inference are:

- (K1) If sentence A is known, then A.
- (K2) (K1) is known.
- (K3) If B is correctly inferred from A, and A is known, then B is known.

To give an absolutely explicit derivation of the paradox by applying these principles to S, we must add (contingent) assumptions to the effect that certain inferences have been performed (see Anderson, 1983). But as we go through the argument of the Knower, these inferences *are* performed. And even if we can somehow restrict such principles and construct a consistent formal logic of knowledge and inference, the paradoxical argument as expressed in the natural language still demands some explanation.

The usual proposals for dealing with the Liar often have their analogues for the Knower, e.g. that there is something wrong with self-reference or that knowledge (or truth) is properly a predicate of propositions and not of sentences. And the replies which show that some of these are not adequate are often parallel to those for the Liar paradox (see Kripke, 1975; Mates, 1981; Anderson, 1983). In addition one can try here what seems to be an adequate solution for the Surprise Examination Paradox, namely the observation that "new knowledge can drive out old knowledge" (see Ayer, 1973). But this doesn't seem to work on the Knower (Anderson, 1983).

The two approaches which have some hope of adequately dealing with this paradox are "hierarchy" solutions and "truth-value gap" solutions. According to the first, knowledge is structured into "levels". It is urged that there is no one coherent notion expressed by the verb "knows", but rather a whole series of notions: knows<sub>0</sub>, knows, . . . and so on (perhaps into the transfinite). Stated in terms of predicates expressing such "ramified" concepts and properly restricted, (K1)–(K3) lead to no contradictions. (See Burge (1979) for a version of the idea as applied to truth, Anderson (1983) for an analogous approach to knowledge.) The main objections to this procedure are that the meaning of these levels has not been adequately explained and that the idea of such subscripts, even implicit, in a natural language is highly counter-intuitive. The "truth-value gap" solution takes sentences such as S to lack truth-value, they are neither true nor false – they do not express propositions. This defeats a crucial step in the reasoning used in the derivation of the paradox. Kripke (1975) has developed this approach in connection with the Liar and Asher and Kamp (1986) have worked out some details of a parallel solution to the Knower. The principal objection is that "strengthened" or "super" versions of the paradoxes tend to reappear when the solution itself is stated.

Since the paradoxical deduction uses only the properties (K1)–(K3) and since the argument is formally valid, any notions which satisfy these conditions will lead to paradox. Thus Grim (1988) notes that "K" may be read as "is known by an omniscient God" and concludes that there is no coherent single notion of omniscience. And Thomason (1980) observes that with some slightly different conditions, analogous reasoning about belief can lead to paradoxical consequences.

All things considered it looks as if we should conclude that knowledge and truth are ultimately intrinsically 'stratified' concepts. And it would seem that we must simply accept the fact that these (and similar) concepts cannot be assigned any one fixed level, finite or infinite. But the meaning of this idea certainly needs further clarification.

## BIBLIOGRAPHY

- Anderson, C. A.: "The Paradox of the Knower," *Journal of Philosophy* 80 (1983), 338–55.
- Asher, N. and Kamp, J. A. W.: "The Knower's Paradox and Representational Theories of Attitudes," *Proceedings of the IBM Conference on Theoretical Aspects of Reasoning about Knowledge*, ed. J. Halpern (St Louis, MO: Morgan Kaufman, 1986), 131–48.
- Ayer, A. J.: "On a Supposed Antinomy," *Mind* 83 (1973), 125–6.
- Burge, T.: "Semantical Paradox," *Journal of Philosophy* 76 (1979), 169–98.
- Grim, P.: "Truth, Omniscience, and the Knower," *Philosophical Studies* 54 (1988), 9–41.
- Kaplan, D. and Montague, R.: "A Paradox Regained," *Notre Dame Journal of Formal Logic* (1960), 79–90.
- Kripke, S.: "Outline of a Theory of Truth," *Journal of Philosophy* 72 (1975), 690–716.
- Mates, B.: *Sceptical Essays* (Chicago, IL, and London: University of Chicago Press, 1981).
- Montague, R.: "Syntactical Treatments of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability," *Acta Philosophica Fennica* 16 (1963), 153–67.
- Thomason, R.: "A Note on Syntactical Treatments of Modality," *Synthese* 44 (1980), 391–5.

C. ANTHONY ANDERSON

**paradoxes of analysis** How can analysis be informative? This is the question that gives rise to what philosophers have traditionally called "the" paradox of analysis. Thus, consider the following proposition.

- (1) To be an instance of knowledge is to be an instance of justified true belief not essentially grounded in any falsehood.

(1), if true, illustrates an important type of philosophical analysis. For convenience of exposition, I will assume (1) is a *correct* analysis. The paradox arises from the fact that if the concept of justified true belief not essentially grounded in any falsehood is the *analysans* of the concept of knowledge, it would seem that they are the same concept and hence that

- (2) To be an instance of knowledge is to be an instance of knowledge

would have to be the same proposition as (1). But then how can (1) be informative when (2) is not? This is what I call the first paradox of analysis.

Classical writings on analysis suggest a second paradox of analysis (Moore, 1942). Consider this proposition:

- (3) An analysis of the concept of being a brother is that to be a brother is to be a male sibling.

If (3) is true, it would seem that the concept of being a brother would have to be the same concept as the concept of being a male sibling and that

- (4) An analysis of the concept of being a brother is that to be a brother is to be a brother

would also have to be true and in fact would have to be the same proposition as (3). Yet (3) is true and (4) is false.

Both of these paradoxes rest upon the assumptions that analysis is a relation between concepts, rather than one involving entities of other sorts, such as linguistic expressions, and that in a true analysis, *analysans* and *analysandum* are the same concept. Both these assumptions are explicit in Moore. But some of Moore's remarks hint at a solution – that a statement of an analysis is a statement partly about the concept involved and partly about the verbal expressions used to express it. He says he thinks a solution of this sort is bound to be right, but fails to suggest one because he cannot see a way in which the analysis can be even partly about the expressions (Moore, 1942).

Elsewhere, I have suggested such a way as a solution to the second paradox, which is to explicate (3) as

- (5) An analysis is given by saying that the verbal expression "x is a brother" expresses the same concept as is expressed by the conjunction of the verbal expressions "x

is male" when used to express the concept of being male and "x is a sibling" when used to express the concept of being a sibling (Ackerman, 1990).

An important point about (5) is as follows. Stripped of its philosophical jargon ("analysis", "concept", "x is a . . ."), (5) seems to state the sort of information generally stated in a definition of the verbal expression "brother" in terms of the verbal expressions "male" and "sibling", where this definition is designed to draw upon listeners' antecedent understanding of the verbal expressions "male" and "sibling", and thus to tell listeners what the verbal expression "brother" really means, instead of merely providing the information that two verbal expressions are synonymous without specifying the meaning of either one. Thus, my solution to the second paradox seems to make the sort of analysis that gives rise to this paradox a matter of specifying the meaning of a verbal expression in terms of separate verbal expressions already understood and saying how the meanings of these separate, already-understood verbal expressions are combined. (This corresponds to Moore's intuitive requirement that an analysis should both specify the constituent concepts of the *analysandum* and tell how they are combined; see Moore, 1942.) But is this all there is to philosophical analysis?

To answer this question, we must note that, in addition to there being two paradoxes of analysis, there are two types of analysis that are relevant here. (There are also other types of analysis, such as reformatory analysis, where the *analysans* is intended to improve on and replace the *analysandum*. But since reformatory analysis involves no commitment to conceptual identity between *analysans* and *analysandum*, reformatory analysis does not generate a paradox of analysis and so will not concern us here.) One way to recognize the difference between the two types of analysis concerning us here is to focus on the difference between the two paradoxes. This can be done by means of the Frege-inspired sense-individuation condition, which is the condition that two expressions have the same sense if and only if they can be

interchanged *salva veritate* whenever used in propositional attitude contexts. If the expressions for the *analysans* and the *analysandum* in (1) met this condition, (1) and (2) would not raise the first paradox, but the second paradox arises regardless of whether the expressions for the *analysans* and the *analysandum* meet this condition. The second paradox is a matter of the failure of such expressions to be interchangeable *salva veritate* in sentences involving such contexts as 'an analysis is given by'. Thus, a solution (such as the one I have offered) that is aimed only at such contexts can solve the second paradox. This is clearly false for the first paradox, however, which will apply to all pairs of propositions expressed by sentences in which expressions for pairs of *analysanda* and *analysantia* raising the first paradox are interchanged. For example, consider the following proposition:

(6) Mary knows that some cats lack tails.

It is possible for John to believe (6) without believing

(7) Mary has justified true belief, not essentially grounded in any falsehood, that some cats lack tails.

Yet this possibility clearly does not mean that the proposition that Mary knows that some cats lack tails is partly about language.

One approach to the first paradox is to argue that, despite the apparent epistemic inequivalence of (1) and (2), the concept of justified true belief not essentially grounded in any falsehood is still identical with the concept of knowledge (see Sosa, 1983). Another approach is to argue that in the sort of analysis raising the first paradox, the *analysans* and *analysandum* are concepts that are different but that bear a special epistemic relation to each other. Elsewhere I have developed such an approach and suggested that this *analysans*–*analysandum* relation has the following facets (Ackerman, forthcoming).

(a) The *analysans* and *analysandum* are necessarily coextensive; i.e. necessarily

- every instance of one is an instance of the other.
- (b) The *analysans* and *analysandum* are knowable a priori to be coextensive.
  - (c) The *analysandum* is simpler than the *analysans* (a condition whose necessity is recognized in classical writings on analysis, such as Langford, 1942).
  - (d) The *analysans* does not have the *analysandum* as a constituent.

Condition (d) rules out circularity. But since many valuable quasi-analyses are partly circular (e.g. knowledge is justified true belief supported by *known* reasons not essentially grounded in any falsehood), it seems best to distinguish between full analysis, for which (d) is a necessary condition, and partial analysis, for which it is not.

These conditions, while necessary, are clearly insufficient. The basic problem is that they apply to many pairs of concepts that do not seem closely enough related epistemically to count as *analysans* and *analysandum*, such as the concept of being 6 and the concept of being the fourth root of 1296. Accordingly, my solution finds the fifth condition by drawing upon what actually seems epistemically distinctive about analyses of the sort under consideration, which is a certain way they can be justified. This is by the philosophical example-and-counterexample method, which in general terms goes as follows. J investigates the analysis of K's concept Q (where K can but need not be identical to J) by setting K a series of armchair thought experiments, i.e. presenting K with a series of simple described hypothetical test cases and asking K questions of the form "If such-and-such were the case, would this count as a case of Q?" J then contrasts the descriptions of the cases to which K answers affirmatively with the descriptions of the cases to which K does not, and J generalizes upon these descriptions to arrive at the concepts (if possible not including the *analysandum*) and their mode of combination that constitute the *analysans* of K's concept Q. Since J need not be identical with K, there is no requirement that K himself be able to perform this generalization, to recognize its result as correct, or even to

understand the *analysans* that is its result. This is reminiscent of Walton's observation that one can simply recognize a bird as a swallow without realizing just what features of the bird (beak, wing configuration, etc.) form the basis of this recognition. (The philosophical significance of this way of recognizing is discussed in Walton, 1972.) K answers the questions based solely on whether the described hypothetical cases just strike him as cases of Q. J observes certain strictures in formulating the cases and questions. He makes the cases as simple as possible, to minimize the possibility of confusion and also to minimize the likelihood that K will draw upon his philosophical theories (or quasi-philosophical, rudimentary notions if he is unsophisticated philosophically) in answering the questions. For this reason, if two hypothetical test cases yield conflicting results, the conflict should *ceteris paribus* be resolved in favour of the simpler case. J makes the series of described cases wide-ranging and varied, with the aim of having it be a *complete* series, where I say a series is complete if and only if no case that is omitted is such that, if included, it would change the analysis arrived at. J does not, of course, use as a test-case description anything complicated and general enough to express the *analysans*. There is no requirement that the described hypothetical test cases be formulated only in terms of what can be observed. Moreover, using described hypothetical situations as test cases enables J to frame the questions in such a way as to rule out extraneous background assumptions to a degree. Thus, even if K correctly believes that all and only P's are R's, the question of whether the concepts of P, R, or both enter into the *analysans* of his concept Q can be investigated by asking him such questions as "Suppose (even if it seems preposterous to you) that you were to find out that there was a P that was not an R. Would you still consider it a case of Q?"

Taking all this into account, the fifth necessary condition for this sort of *analysans-analysandum* relation is as follows:

- (e) If S is the *analysans* of Q, the proposition that necessarily all and only instances of

S are instances of Q can be justified by generalizing from intuitions about the correct answers to questions of the sort indicated about a varied and wide-ranging series of simple described hypothetical situations.

Are these five necessary conditions jointly sufficient? For a discussion of qualifications and additional conditions that some objections may be necessitate, see Ackerman, forthcoming.

See also PARADOX.

#### BIBLIOGRAPHY

- Ackerman, D. F.: "The Informativeness of Philosophical Analysis," *Midwest Studies in Philosophy* 6 (1981), 313–20.
- Ackerman, F.: "Analysis, Language, and Concepts: the Second Paradox of Analysis," *Philosophical Perspectives* 4 (1990), 535–43.
- Ackerman, F.: "Analysis and Its Paradoxes," *The Scientific Enterprise: The Israel Colloquium Studies in History, Philosophy, and Sociology of Science* 4, ed. E. Ullman-Margalit (Norwell: Kluwer, forthcoming).
- Langford, C. H.: "The Notion of Analysis in Moore's Philosophy," in *The Philosophy of G. E. Moore*, ed. P. A. Schilpp (Evanston, IL: Northwestern University Press, 1942), 319–43.
- Moore, G. E.: "A Reply to My Critics," in *The Philosophy of G. E. Moore*, ed. P. A. Schilpp (Evanston, IL: Northwestern University Press, 1942), 660–7.
- Pap, A.: *Semantics and Necessary Truth: An Inquiry into the Foundations of Analytic Philosophy* (New Haven, CT: Yale University Press, 1958).
- Sosa, E.: "Classical Analysis," *Journal of Philosophy* 53 (1983), 695–710.
- Walton, K. L.: "Linguistic Relativity," in *Conceptual Change*, ed. G. Pearce and P. Maynard (Dordrecht: Reidel, 1972), 1–29.

FELICIA ACKERMAN

**paranormal knowledge** Among claimed forms of paranormal knowledge are clairvoyance, telepathy and precognition, sometimes grouped together as "ESP" or as "psi-gamma". Intriguing philosophical questions arise regarding both basic conceptions at

issue and the character of the parapsychological data.

#### PROBLEMS IN DEFINING "PARANORMAL"

What does it mean to claim that something is "paranormal"? A tempting beginning, perhaps, is something like (1):

- (1) P is paranormal =<sub>df</sub> P violates the laws of nature

If "the laws of nature" are taken to be the ideal principles which science attempts to discover, however – an assumed set of true and complete generalizations regarding the universe – a definition such as (1) would seem to guarantee that no paranormal phenomena could possibly occur. Were paranormal phenomena by definition those which no complete true description of the universe would include, there *could* be no genuinely paranormal phenomena.

For that reason most philosophers who have addressed the question have preferred a definition more along these lines:

- (2) P is paranormal =<sub>df</sub> P violates the laws of nature *as currently understood*

As an account of common usage, definition (2) may still face objections. So defined, "paranormal" carries an indexical reference to laws of nature as understood at a particular time. In the year 1500, on such a definition, it would thus have been *true* to say that the northern lights or schizophrenia were paranormal, though they are not paranormal now. Such a definition may also be too broad. Because incorrect and incomplete, our science is constantly faced with anomalies. But not even those anomalies that have occasioned the most profound changes in scientific theory have been regarded as *paranormal* (see Braude, 1979).

In what has become a very influential piece, C. D. Broad (1949) defined the (ostensibly) paranormal as that which (*prima facie*) violates one or more of a list of fundamental assumptions or "basic limiting principles". Among these he included:



- (a) general principles of causation, including the principle that a cause must precede its effect, and
- (b) limitations on ways of acquiring knowledge, including the principle that all knowledge is the end product of a causal chain that at some point involves the sense organs of the knower.

Broad's list can be accused of building too much into our fundamental assumptions, however – in particular, an explicit Cartesian dualism and fairly radical empiricism.

There are also questions of definition regarding other terms at issue. Despite the familiarity of the term "ESP" or "extra-sensory perception", for example, it seems to be widely agreed that phenomena at issue – even if genuine – may not properly be characterized on the model of perception (*see* Mundle, 1964). Phenomena of "precognition", despite the term, may not properly be conceived of as phenomena of cognition. Despite the fact that the gamma of "psi-gamma" is taken from the Greek word for knowledge, whether the phenomena at issue can be characterized in terms of knowledge remains an open question.

The notion of precognition faces particularly pointed conceptual questions, since as standardly characterized it would seem to demand that a later "precognized" event at  $t_2$  was in some way causally responsible for an earlier event of precognition at  $t_1$ . In accord with Broad's basic limiting principle (a), it has been argued that the backwards causation so demanded is logically impossible in virtue of the meaning of "cause", and thus that precognition must itself be logically impossible (*see esp.* Flew, 1980). Others have argued, often using the example of tachyons in physics – hypothetical particles moving backward in time – that what such an argument establishes is merely that a new and non-temporal notion of "cause" is required. It is not clear that tachyons don't face their own philosophical problems, however (*see* Craig, 1988) nor that any mere revision of the meaning of "cause" would genuinely address the central difficulty at issue (*see esp.* Flew, 1987).

Attempts have also been made to argue that clairvoyance and telepathy must be impossible on the grounds that they violate Broad's basic limiting principles (*see esp.* Duran, 1990). Here the argument turns on principles that prohibit action at a distance or knowledge by means of other than the standard senses. It's not clear that clairvoyance and telepathy would necessarily violate these, however, nor that the principles at issue should command our allegiance without exception.

Defined as in (2) above or in terms of Broad's "basic limiting principles", paranormal phenomena will at least not be ruled out as impossible by definition. On the basis of such definitions it can still be argued, however, that it would be irrational to believe any evidence offered for a case of the paranormal. Here the basic argument, which appears in one form or another throughout the literature, is essentially Hume's against miracles. A miracle, as Hume characterizes it, is a violation of the laws of nature as we understand them. But

... as a firm and unalterable experience has established these laws, the proof against a miracle ... is as entire as any argument from experience can possibly be imagined.

(*Enquiry*, X)

#### THE CHARACTER OF THE DATA

Here we have concentrated on conceptual questions, but any serious consideration of claims regarding paranormal knowledge must confront the wealth of existing parapsychological data. Questions of basic statistical technique in parapsychological experiment seem to have been settled, though questions may remain regarding selective publication of positive results and experimental repeatability. A particularly sensitive issue is that of fraud, which does unfortunately have a history in parapsychology; at least some of the work of S. G. Soal and W. J. Levy has been shown to be fraudulent, for example. Critics often discount particular experiments on the basis that fraud *could*

have occurred; defenders often object that the mere possibility of fraud should no more discredit results in parapsychology than in other fields. J. B. Rhine attempts to give a summary of fraud-proof evidence in Rhine (1975).

Even if entirely genuine, a question remains as to whether the data would support a claim of paranormal *knowledge*. Necessary conditions for knowledge (even if not sufficient) have standardly been taken to include justified true belief. In cases at issue, truth is granted. But do purported cases of precognition, telepathy and clairvoyance as they appear in the literature support a claim of justified belief?

It can be argued that such cases don't even characteristically involve *belief*. In standard statistical tests, for example, it does not appear that the subject has any feeling of conviction which distinguishes "hits" from "misses" (see Mundle, 1964). Even in a large percentage of spontaneous cases, it appears, there is no feeling of conviction which distinguishes the true experiences from the false (see Ducasse, 1954).

The argument that such cases would not at any rate involve *justified* belief is similarly factual: the success rates simply don't seem good enough to justify belief in any particular instance. In Flew's words,

... it is easy to see how preposterous it is to describe even star performers on their best days as knowing the values of the targets ... they will, if it is a very good day, be averaging 7 or 8 out of 25 where mean chance expectation is 5. ... (Flew, 1987, p. 91)

For spontaneous cases we would have to make similar estimates of success rates, and here too the data seem none too strong. It should also be emphasized that among the strongest parapsychological evidence we find data involving temporal displacement (a pattern of "hits" systematically displaced from the subject's intended targets) and psimissing (a statistically significant pattern of "misses"). Neither of these would seem to fit a pattern of either justified belief or knowledge.

By this stage we are considering not so much whether there *could* be such things as precognition, telepathy, etc., but whether the evidence in their favour is in fact strong enough. Were evidence produced which showed a particularly reliable record of "hits" with a corresponding pattern of conviction these particular objections would seem to be overridden.

# BIBLIOGRAPHY

The *Journal of Parapsychology* is a primary source regarding the parapsychological data.

- Braude, S. E.: *ESP and Psychokinesis* (Philadelphia, PA: Temple University Press, 1979).
- Broad, C. D.: "The Relevance of Psychical Research to Philosophy," *Philosophy* 24 (1949), 291-309; reprinted in Flew (1987), 37-52.
- Craig, W. L.: "Tachyons, Time Travel and Divine Omniscience," *Journal of Philosophy* 85 (1988), 135-50.
- Ducasse, C. J.: "The Philosophical Importance of 'Psychic Phenomena,'" *Journal of Philosophy* 51 (1954), 810-23.
- Duran, J.: "Philosophical Difficulties with Paranormal Knowledge Claims," in Grim (1990), 232-42.
- Flew, A., ed.: *Readings in the Philosophical Problems of Parapsychology* (Buffalo, NY: Prometheus Books, 1987).
- Flew, A.: "Analysing the Concepts of Parapsychology," in Flew (1987), 87-106.
- Flew, A.: "Parapsychology: Science or Pseudoscience?," in *Science, Pseudoscience, and Society*, ed. M. Hanen, M. J. Osler and R. G. Weyant (Waterloo, Ont.: Calgary Institute for the Humanities and Wilfrid Laurier University Press, 1980); reprinted in Grim (1990), 214-31.
- Grim, P., ed.: *Philosophy of Science and the Occult*, 2nd edn (Albany, NY: SUNY Press, 1990).
- Ludwig, J., ed.: *Philosophy and Parapsychology* (Buffalo, NY: Prometheus Books, 1978).
- Mundle, C. W. K.: "Is 'paranormal precognition' a coherent concept?," *Journal of Parapsychology* 6 (1964), 179-94.
- Rhine, J. B.: "Second Report on a Case of Experimenter Fraud," *Journal of Parapsychology* 39 (1975), 306-25; excerpted in Grim (1990), 253-63.

PATRICK GRIM

**Peirce, Charles S. (1839–1914)** Peirce was an American philosopher and logician best-known as the founder of pragmatism (*see* PRAGMATISM) (he called his own version of the doctrine “Pragmatism”). A systematic philosopher influenced by Kant, he attempted to provide a vindication of the scientific method exploiting a sophisticated general theory of representation and a system of categories which, he claimed, improved on Kant’s. Throughout he sought to bring “mathematical exactitude” to philosophy, and was an innovator in logic and mathematics.

From his earliest writings Peirce was critical of Cartesian approaches to epistemology (*see* CARTESIANISM). He charged that the method of doubt encouraged people to pretend to doubt what they did not doubt in their hearts, and criticized its individualist insistence that “the ultimate test of certainty is to be found in the individual consciousness”. We should rather begin from what we cannot in fact doubt, progressing towards the truth as part of a community of inquirers trusting to the multitude and variety of our reasonings rather than to the strength of any one. He claimed to be a contrite fallibilist (*see* FALLIBILISM) and urged that our reasonings “should not form a chain that is no stronger than its weakest link, but a cable whose fibres may be ever so slender, provided they are sufficiently numerous and intimately connected”. In later writings his anti-Cartesianism took the form of “critical commonsensism” (*see* COMMONSENSISM: our inquiries are guided by a slowly evolving body of vague commonsense certainties which are, in principle, fallible; rational self-control requires that we try to doubt these in order to establish that they genuinely form part of commonsense).

Peirce insisted that something can serve as a sign, can represent an object, only if it is interpreted in subsequent thought as standing for that object, and he argued that all thoughts were signs. His extensive writings in “semiotic”, the general theory of signs, provide a sophisticated model of mind and language which was intended to serve as the foundations of his logical doctrines. He held that we investigate the world as members of a scientific commu-

nity of sign interpretation engaged in co-operative investigation, criticizing and challenging opinions, attempting to eliminate error and thus progress towards truth.

In “The Fixation of Belief” (published 1877) Peirce argued that the aim of inquiry must be characterized as fixation or settlement of belief: once doubt is removed we cannot but be satisfied with what we then believe. He compared different methods for fixing belief, concluding that the only method which can be sustained, the only one which is consistent with the presuppositions of inquiry, is the method of science. Its “fundamental principle” is that there are real things entirely independent of our opinions about them which affect our senses in regular ways; any inquirers with sufficient experience who reason hard enough on the matter will be fated to participate in a consensus about how these realities are. Much of his work was concerned with describing the method of science in more detail and with attempting to prove that it will take us to the truth.

For Peirce, the method of science has three components: abduction (or retroduction), deduction and induction (*see* ABDUCTION; INTUITION AND DEDUCTION; INDUCTION). His vindication of induction rests upon the claims that all induction resembles statistical sampling and that such reasoning has a self-correcting character. Although there is no logical basis for relying on induction in the short run, we can be confident that repeated use of induction will lead the community of inquirers eventually to eliminate error and reach the truth (*see* PROBLEMS OF INDUCTION). The logic of abduction is the logic of discovery: it is concerned with which hypotheses are worth taking seriously; and Peirce made many contributions to our understanding of these issues.

Peirce’s famous pragmatist principle is a rule of logic to be employed in clarifying our concepts and ideas. Consider the claim that the liquid in a flask is an acid. If we believe this, we expect that if we were to place blue litmus paper in the flask, it would turn red: we expect an action of ours to have certain experiential results. The pragmatist principle holds that listing the conditional expectations

of this kind that we associate with applications of a concept provides a *complete* clarification of the concept. This is relevant to the logic of abduction: clarification using the pragmatist principle provides all the information about the content of an hypothesis that is relevant to deciding whether it is worth testing.

The most famous application of the pragmatist principle is Peirce's account of reality: when we take something to be really the case, we think it is "fated to be agreed upon by all who investigate" the matter; in other words, if I believe that it is really the case that *p*, then I expect that if anyone were to inquire well enough and for long enough into whether *p*, they would arrive at the belief that *p*. It is not part of the theory that the experiential consequences of our actions should be specified in a narrowly empiricist vocabulary – Peirce insisted that perception was theory laden. Nor is it his view that the conditionals listed when we clarify a concept are all analytic. Moreover, in later writings, he urged that the pragmatist principle could only be plausible to someone who accepted metaphysical realism (see REALISM): it requires that "would-be's" are objective and real. Indeed much of Peirce's later work is concerned with developing a "scientific metaphysics" which vindicates his metaphysical realism. His use of the term "pragmatism" to describe his own position was designed to distinguish it from the views of avowed "pragmatists" like William James (see JAMES) who, Peirce thought, had failed to see the necessity of defending realism.

See also JAMES; PRAGMATISM.

#### WRITINGS

*Collected Papers of Charles S. Peirce*, 8 vols (Cambridge, MA: Harvard University Press, 1931–58).

*Charles S. Peirce: Selected Writings*, ed. P. P. Wiener (New York: Dover, 1966).

*Reasoning and the Logic of Things*, ed. K. Ketner with introductions by Ketner and H. Putnam (Cambridge, MA: Harvard University Press, 1991).

*The Writings of Charles S. Peirce: A Chronological Edition* (Indianapolis, IN: Indiana University Press, 1982–).

#### BIBLIOGRAPHY

Hookway, C. J.: *Peirce* (London: Routledge and Kegan Paul, 1985).

Scheffler, I.: *Four Pragmatists* (London: Routledge and Kegan Paul, 1974), pt 1.

Skagestad, P.: *The Road of Inquiry* (New York: Columbia University Press, 1981).

CHRISTOPHER HOOKWAY

**perceptual knowledge** Perceptual knowledge is knowledge acquired by or through the senses. This includes most of what we know. Some would say it includes *everything* we know. We cross intersections when we see the light turn green, head for the kitchen when we smell the roast burning, squeeze the fruit to determine its ripeness, and climb out of bed when we hear the alarm ring. In each case we come to know something – that the light has turned green, that the roast is burning, that the melon is overripe, and that it is time to get up – by some sensory means. Seeing that the light has turned green is learning something – that the light has turned green – by use of the eyes. Feeling that the melon is overripe is coming to know a fact – that the melon is overripe – by one's sense of touch. In each case the resulting knowledge is somehow *based on*, *derived from* or *grounded in* the sort of experience that characterizes the sense modality in question.

Seeing a rotten kumquat is not at all like the experience of smelling, tasting or feeling a rotten kumquat. Yet all these experiences can result in the same knowledge – knowledge that the kumquat is rotten. Although the experiences are much different, they must, if they are to yield knowledge, embody information about the kumquat: the information that it is rotten. Seeing that the fruit is rotten differs from smelling that it is rotten, not in *what* is known, but *how* it is known. In each case, the information has the same source – the rotten kumquat – but it is, so to speak, delivered

via different channels and coded in different experiences.

It is important to avoid confusing perceptual knowledge of facts (e.g. that the kumquat is rotten) with the perception of objects (e.g. rotten kumquats). It is one thing to see (taste, smell, feel) a rotten kumquat, quite another to *know* (by seeing or tasting) that it is a rotten kumquat. Some people, after all, don't know what kumquats look like. They see a kumquat but do not realize (do not *see* that) it is a kumquat. Some people don't know what kumquats smell like. They smell a rotten kumquat and – thinking, perhaps, that this is the way this strange fruit is supposed to smell – do not realize from the smell (i.e. do not *smell that*) it is rotten. In such cases people see and smell rotten kumquats – and in this sense *perceive* rotten kumquats – and never know that they are kumquats – let alone rotten kumquats. They cannot, not at least by seeing and smelling, and not until they have learned something about (rotten) kumquats, come to *know* that what they are seeing or smelling is a (rotten) kumquat. Since the topic of this essay is perceptual *knowledge* – knowing, by sensory means, that something is F – we will be primarily concerned with the question of what *more, beyond* the perception of Fs, is needed to see that (and thereby know that) they are F. The question is, not how we *see* kumquats (for even the ignorant can do this), but, how we know (if, indeed, we do) that that is what we see.

Much of our perceptual knowledge is indirect, dependent or derived. By this I mean that the facts we describe ourselves as learning, as coming to know, by perceptual means are pieces of knowledge that depend on our coming to know something else, some other fact, in a more direct way. We see, *by the gauge*, that we need gas; see, *by the newspapers*, that our team has lost again; see, *by her expression*, that she is nervous. This derived or dependent sort of knowledge is particularly prevalent in the case of vision, but it occurs, to a lesser degree, in every sense modality. We install bells and other noise-makers so that we can, for example, *hear* (by the bell) that someone is at the door and (by the alarm) that its time to get up. When we obtain knowledge in

this way, it is clear that unless one sees – hence, comes to know – something about the gauge (that it reads “empty”), the newspaper (what it says) and the person's expression, one would not see (hence, know) what one is described as coming to know by perceptual means. If one can't hear that the bell is ringing, one cannot – not at least in *this* way – hear that one's visitors have arrived. In such cases one sees (hears, smells, etc.) that *a* is F, coming to know thereby that *a* is F, *by* seeing (hearing, etc.) that some *other* condition, *b*'s being G, obtains. When this occurs, the knowledge (that *a* is F) is derived from, or dependent on, the more basic perceptual knowledge that *b* is G.

Though perceptual knowledge about objects is often, in this way, dependent on knowledge of facts about different objects, the derived knowledge is sometimes about the *same* object. That is, we see that *a* is F by seeing, not that some other object is G, but that *a* itself is G. We see, *by her expression*, that *she* is nervous. She tells that the fabric is silk (not polyester) by the characteristic “greasy” feel of the fabric itself (not, as I do, by what is printed on the label). We tell whether it's an oak tree, a Porsche, a geranium, an igneous rock or a misprint by its shape, colour, texture, size, behaviour and distinctive markings. Perceptual knowledge of this sort is also *derived* – *derived from* the more basic facts (about *a*) we use to make the identification. In this case the perceptual knowledge is still indirect because, although the same object is involved, the facts we come to know about it are different than the facts that enable us to know it.

Derived knowledge is sometimes described as *inferential*, but this is misleading. At the conscious level there is no passage of the mind from premise to conclusion, no reasoning, no problem-solving. The observer, the one who sees that *a* is F by seeing that *b* (or *a* itself) is G, needn't be (and typically isn't) aware of any process of inference, any passage of the mind from one belief to another. The resulting knowledge, though logically derivative, is psychologically immediate. I could *see* that she was getting angry; so I moved my hand. I did not – at least not at any conscious level – *infer* (from her expression and



behaviour) that she was getting angry. I could (or so it seems to me) *see* that she was getting angry. It is this psychological immediacy that makes indirect perceptual knowledge a species of *perceptual* knowledge.

The psychological immediacy that characterizes so much of our perceptual knowledge – even (sometimes) the most indirect and derived forms of it – does not mean that *learning* is not required to know in this way. One isn't born with (may, in fact, never develop) the ability to recognize daffodils, muskrats and angry companions. It is only after a long experience that one is able visually to identify such things. Beginners may do something corresponding to inference: they recognize relevant features of trees, birds and flowers, features they already know how to perceptually identify, and *then* infer (conclude), on the basis of what they see, and under the guidance of more expert observers, that it's an oak, a finch or a geranium. But the experts (and we are all experts on many aspects of our familiar surroundings) do not typically go through such a process. The expert just *sees* that it's an oak, a finch or a geranium. The perceptual knowledge of the expert is still dependent, of course, since even an expert can't see what kind of flower it is if she can't first see its colour and shape, but it is to say that the expert has developed identificatory skills that no longer require the sort of conscious inferential processes that characterize a beginner's efforts.

Coming to know that *a* is F by seeing that *b* is G obviously requires some background assumption on the part of the observer, an assumption to the effect that *a* is F (or perhaps only probably F) when *b* is G. If one doesn't assume (take it for granted) that the gauge is *properly connected*, doesn't (thereby) assume that it would not register "Empty" unless the tank was nearly empty, then even if one could see that it registered "Empty", one wouldn't learn (hence wouldn't see) that one needed gas. At least one wouldn't see it *by* consulting the gauge. Likewise, in trying to identify birds, it's no use being able to see their markings if one doesn't know something about which birds have which marks – something of the form: a bird with these markings is (probably) a finch.

It would seem, moreover, that these background assumptions, if they are to yield *knowledge* that *a* is F, as they must if the observer is to *see* (by *b*'s being G) that *a* is F, must themselves qualify as knowledge. For if this background fact isn't known, if it isn't known whether *a* is F *when b* is G, then the knowledge of *b*'s being G is, taken by itself, powerless to generate the knowledge that *a* is F. If the conclusion is to be *known* to be true, *both* the premises used to reach that conclusion must be known to be true. Or so it would seem.

Externalists (see EXTERNALISM/INTERNALISM), however, argue that the indirect knowledge that *a* is F, though it may depend on the knowledge that *b* is G, does not require *knowledge* of the connecting fact, the fact that *a* is F *when b* is G. Simple belief (or, perhaps, justified belief; there are stronger and weaker versions of externalism) in the connecting fact is sufficient to confer a knowledge of the connected fact. Even if, strictly speaking, I don't *know* she is nervous whenever she fidgets like that, I can none the less *see* (hence, know) that she is nervous (by the way she fidgets) if I (correctly) assume that this behaviour is a reliable expression of nervousness. One needn't *know* the gauge is working well to make observations (acquire observational knowledge) with it. All that is required, besides the observer *believing* that the gauge is reliable, is that the gauge, in fact, be reliable (i.e. that the observer's background beliefs be *true*). Critics of externalism have been quick to point out that this theory has the unpalatable consequence that knowledge can be made possible by – and, in this sense, be made to rest on – lucky hunches (that turn out true) and unsupported (even irrational) beliefs. Surely, internalists argue, if one is going to know that *a* is F on the basis of *b*'s being G, one should have (as a bare minimum) some justification for thinking that *a* is F, or is probably F, when *b* is G.

Whatever view one takes about these matters (with the possible exception of extreme externalism), indirect perception obviously requires some understanding (knowledge? justification? belief?) of the general relationship

between the fact one comes to know (that *a* is *F*) and the facts (that *b* is *G*) that enable one to know it. And it is this requirement on background knowledge or understanding that leads to questions about the possibility of indirect perceptual knowledge. Is it really knowledge? Even if it is, is it really *perceptual* knowledge? The first question is inspired by sceptical doubts about whether we can ever know the connecting facts in question. How is it possible to learn, to acquire knowledge of, the connecting facts knowledge of which is necessary to see (by *b*'s being *G*) that *a* is *F*? These connecting facts do not appear to be *perceptually* knowable. Quite the contrary; they appear to be general truths knowable (if knowable at all) by inductive inference from past observations. And if one is sceptical about obtaining knowledge in this indirect, inductive, way, one is, perforce, sceptical about the existence of the kind of indirect knowledge, including indirect perceptual knowledge of the sort described above, that depends on it.

Even if one puts aside such sceptical questions, however, there remains a legitimate concern about the *perceptual* character of this kind of knowledge. If one sees that *a* is *F* by seeing that *b* is *G*, is one really *seeing* that *a* is *F*? Isn't perception merely a part – and, indeed, from an epistemological standpoint, the less significant part – of the process whereby one comes to know that *a* is *F*. One must, it is true, see that *b* is *G*, but this is only one of the premises needed to reach the conclusion (knowledge) that *a* is *F*. There is also the background knowledge that is essential to the process. If we think of a *theory* as any factual proposition, or set of factual propositions, that cannot itself be known in some direct observational way, we can express this worry by saying that indirect perception is always theory-loaded: seeing (indirectly) that *a* is *F* is only possible if the observer *already* has knowledge of (justification for, belief in) some theory, the theory “connecting” the fact one comes to know (that *a* is *F*) with the fact (that *b* is *G*) that enables one to know it.

This, of course, reverses the standard (foundationalist) picture (see FOUNDATIONALISM) of human knowledge. Instead of theoretical

knowledge depending on, and being derived from, perception, perception (of the indirect sort) presupposes a prior knowledge of theories.

Foundationalists are quick to point out that this apparent reversal in the structure of human knowledge is *only* apparent. Our indirect perception of facts depends on theory, yes, but this merely shows that indirect perceptual knowledge is not part of the foundation. To reach the kind of perceptual knowledge that lies at the foundation, we need to look at a form of perception that is purified of all theoretical elements. This, then, will be perceptual knowledge *pure* and *direct*. No background knowledge or assumptions about connecting regularities are needed in direct perception because the known facts are presented directly and immediately and not (as in indirect perception) on the basis of some *other* facts. In direct perception *all* the justification (needed for knowledge) is right there in the experience itself.

What, then, about the possibility of perceptual knowledge pure and direct, the possibility of coming to know, on the basis of sensory experience, that *a* is *F* where this does *not* require, and in no way presupposes, background assumptions or knowledge that has a source outside the experience itself? Where is this epistemological “pure gold” to be found?

There are, basically, two views about the nature of direct perceptual knowledge (coherentists would deny that any of our knowledge is basic in this sense) (see COHERENTISM). These views (following traditional nomenclature) can be called direct realism (see DIRECT REALISM) and representationalism (or representative realism) (see REPRESENTATIVE REALISM). A representationalist restricts direct perceptual knowledge to objects of a very special sort: ideas, impressions, or sensations (sometimes called sense-data) – entities in the mind of the observer. One *directly* perceives a fact (e.g. that *b* is *G*) only when *b* is a mental entity of some sort – a subjective appearance or sense-datum – and *G* is a property of this datum. Knowledge of these sensory states is supposed to be certain and infallible. These sensory facts are, so to speak, right up against the mind's eye. One cannot be mistaken

about these facts for these facts are, in reality, facts about the way things appear to be, and one can't be mistaken about the way things appear to be. Normal perception of external conditions, then, turns out to be (always!) a type of *indirect* perception. One "sees" that there is a tomato in front of one *by* seeing that the appearances (of the tomato) have a certain quality (reddish and bulgy) and inferring (this is typically said to be automatic and unconscious), on the basis of certain background assumptions (e.g. that there typically is a tomato in front of one when one has experiences of this sort) that there is a tomato in front of one. All knowledge of objective reality, then, even what commonsense regards as the most direct perceptual knowledge, is based on an even more direct knowledge of the appearances.

For the representationalist, then, perceptual knowledge of our physical surroundings is always theory-loaded and indirect. Such perception is 'loaded' with the theory that there is some regular, some uniform, correlation between the way things appear (known in a perceptually direct way) and the way things actually are (known, if known at all, in a perceptually indirect way).

The second view, *direct* realism, refuses to restrict direct perceptual knowledge to an inner world of subjective experience. Though the direct realist is willing to concede that much of our knowledge of the physical world is indirect (however direct and immediate it may sometimes *feel*), *some* perceptual knowledge of physical reality is direct. What makes it direct is that such knowledge is not based on, nor in any way dependent on, other knowledge and belief. The justification needed for the knowledge is right there in the experience itself.

To understand the way this is supposed to work, consider an ordinary example. *S* identifies a banana (learns that it is a banana) *by* noting its shape and colour – perhaps even tasting and smelling it (to make sure its not wax). In this case the perceptual knowledge that it is a banana is (the direct realist admits) indirect, dependent on *S*'s perceptual knowledge of its shape, colour, smell and taste. *S* learns that it is a banana by seeing that

it is yellow, banana-shaped, etc. None the less, *S*'s perception of the banana's colour and shape is not indirect. *S* does not see that the object is yellow, for example, by seeing (knowing, believing) anything more basic – either about the banana or anything else (e.g. his own sensations of the banana). *S* has *learned* to identify such features, of course, but what *S* learned to do is not make an inference, even a unconscious inference, from other things he believes. What *S* acquired was a cognitive skill, a disposition to believe of yellow objects he saw that they were yellow. The exercise of this skill does not require, and in no way depends on, the having of any *other* beliefs. *S*'s identificatory successes will depend on his operating in certain special conditions, of course. *S* will not, perhaps, be able to visually identify yellow objects in drastically reduced lighting, at funny viewing angles, or when afflicted with certain nervous disorders. But these facts about *when* *S* can see that something is yellow does not show that his perceptual knowledge (that *a* is yellow) in any way depends on a *belief* (let alone *knowledge*) that he is in such special conditions. It merely shows that direct perceptual knowledge is the result of exercising a skill, an identificatory skill, that like any skill, requires certain conditions for its successful exercise. An expert basketball player can't shoot accurately in a hurricane. He needs normal conditions to do what he has learned to do. So also with individuals who have developed perceptual (cognitive) skills. They need normal conditions to do what they have learned to do. They need normal conditions to see, for example, that something is yellow. But they don't, any more than the basketball player, have to know they are in these conditions to do what being in these conditions enables them to do.

This means, of course, that for the direct realist direct perceptual knowledge is fallible and corrigible. Whether *S* sees that *a* is *F* depends on his being caused to believe that *a* is *F* in conditions that are appropriate for an exercise of that cognitive skill. If conditions *are* right, then *S* sees (hence, knows) that *a* is *F*. If they aren't, he doesn't. Whether or not *S* knows depends, then, not on what else (if anything)

S believes, but on the circumstances in *which* S comes to believe. This being so, this type of direct realism is a form of externalism (see the discussion above and EXTERNALISM/INTERNALISM). Direct perception of objective facts, pure perceptual knowledge of external events, is made possible because what is needed (by way of justification) for such knowledge has been reduced. Background knowledge – and, in particular, the knowledge that the experience does, indeed, suffice for knowing – isn't needed.

This means that the foundations of knowledge are fallible. None the less, though fallible, they are in no way derived. That is what makes them foundations. Even if they are brittle, as foundations sometimes are, everything else rests upon them.

See also DIRECT REALISM; EXPERIENCE; REPRESENTATIVE REALISM; SENSATION/COGNITION.

## BIBLIOGRAPHY

- BonJour, L.: *The Structure of Empirical Knowledge* (Cambridge: Cambridge University Press, 1985).  
 Chisholm, R.: *Perceiving: A Philosophical Study* (Ithaca, NY: Cornell University Press, 1957).  
 Dretske, F.: *Seeing and Knowing* (Chicago, IL: University of Chicago Press, 1969).  
 Dretske, F.: *Knowledge and the Flow of Information* (Cambridge, MA: MIT Press, 1981).  
 Goldman, A.: "Discrimination and Perceptual Knowledge," *Journal of Philosophy* 73 (1976), 771–91.  
 Jackson, F.: *Perception* (Cambridge: Cambridge University Press, 1977).  
 Lehrer, K.: *Knowledge* (Oxford: Clarendon Press, 1974).  
 Pitcher, G. A.: *Theory of Perception* (Princeton, NJ: Princeton University Press, 1971).  
 Price, H. H.: *Perception* (Oxford: Methuen, 1932).  
 Swartz, R., ed.: *Perceiving, Sensing and Knowing* (Berkeley, CA: University of California Press, 1965).

FRED DRETSKE

**perspectivism** see NIETZSCHE.

**phenomenalism** Classical phenomenalism is the view that propositions asserting the

existence of physical objects are analytically equivalent to propositions asserting that subjects would have certain sequences of sensations were they to have certain others. Although it is primarily a metaphysical view, phenomenalism is a theory with important epistemological implications and, indeed, it is often accepted by philosophers for its alleged epistemological advantages.

The basic idea behind phenomenalism is compatible with a number of different analyses of the self or conscious subject. Thus a phenomenalist might understand the self as a mind of the sort that Berkeley (see BERKELEY) was committed to, or as a Humean construct of actual and possible experience (see HUME). Also, as I use the term, a phenomenalist might adopt any number of different analyses of the visual, tactile, auditory, olfactory, gustatory and kinesthetic sensations described in the antecedents and consequents of the subjunctive conditionals that, according to the phenomenalist, analyse physical object propositions. Probably the most common analysis of sensations adopted by the phenomenalist is a sense-datum theory (see SENSE-DATA) with the sense-data construed as mind-dependent entities (see ACT/OBJECT ANALYSIS). But there is nothing to prevent a phenomenalist from accepting an adverbial theory (see ADVERBIAL THEORY) instead.

## HISTORICAL ORIGINS

The historical origins of classical phenomenalism are difficult to trace, in part because, as one would expect, early statements of the view were usually not very careful. In his *Dialogues*, Berkeley hinted at classical phenomenalism when he had Philonous explain how he could reconcile an ontology containing only minds and ideas with the story of a creation that took place before the existence of humans:

Why, I imagine that if I had been present at the Creation, I should have seen things produced into being; that is, become perceptible, in the order described by the sacred historian.

(Berkeley, 1713, p. 251)

More often, however, Berkeley seemed to rely on actual ideas in the mind of God to secure the existence of a physical world that is independent of the existence of any finite being.

John Stuart Mill (1867) (see MILL) may have been the first philosopher to put forth a phenomenalistic analysis. In that work he says that matter is a “permanent possibility of sensation” and in explaining what permanent possibilities of sensation are Mill often seems to suggest that they could be understood in terms of the sensations one would have under certain conditions.

The attraction of classical phenomenalism grew with the rise of logical positivism (see LOGICAL POSITIVISM) and increasing acceptance of verificationism (see VERIFICATIONISM). Phenomenalism, the argument went, is the only view that can accommodate both the ordinary conception of physical objects as mind-independent, enduring entities and the commonsense view that it is possible to confirm through experience the existence of physical objects. To understand the argument in support of this claim it is necessary to understand the epistemological framework presupposed by most phenomenologists. Classical phenomenologists were invariably foundationalists (see FOUNDATIONALISM), who further endorsed the radical empiricist’s claim that the only contingent propositions we know directly are propositions describing the contents of our minds. If any belief about the physical world is to be justified at all it must be *inferentially* justified from what we know about the contents of our minds.

Classical phenomenologists were also usually implicitly *internalists* in epistemology (see EXTERNALISM/INTERNALISM) – at least they were implicitly committed to the view one might call *inferential* internalism. According to the inferential internalist one can be justified in believing one proposition *p* on the basis of another *e* only if one is justified in believing that *e* makes it likely that *p*. But how can one establish the occurrence of certain sensations as evidence for the existence of physical objects when all one knows directly is that certain sensations occur? If one’s paradigm for establishing one thing as the sign of something else is inductive reasoning, there is obviously

a fundamental and apparently insuperable problem (see PROBLEM OF THE EXTERNAL WORLD). To establish inductively that sensations are signs of physical objects one would have to *observe* a correlation between the occurrence of certain sensations and the existence of certain physical objects. But to observe such a correlation in order to establish a connection, one would need independent access to physical objects and, by hypothesis, this one cannot have. If one further adopts the verificationist’s stance that the ability to comprehend is parasitic on the ability to confirm, one can easily be driven to Hume’s conclusion (see HUME):

Let us chace our imagination to the heavens, or to the utmost limits of the universe; we never really advance a step beyond ourselves, nor can conceive any kind of existence, but those perceptions, which have appear’d in that narrow compass. This is the universe of the imagination, nor have we have any idea but what is there produc’d.

(Hume, 1739–40, pp. 67–8)

If one reaches such a conclusion but wants to maintain the intelligibility and verifiability of assertions about the physical world, one can go either the idealistic or the phenomenalistic route. The idealist (Berkeley, for example) tries to identify physical objects with bundles of sensations. The obvious difficulty (of which Berkeley was certainly aware) is how to preserve the mind-independent status of physical objects. The ordinary conception of a physical object just is the concept of something that could exist even in the absence of any minds (and their sensations). As an alternative to idealism, the phenomenologist proposes that we invoke the conceptual machinery of subjunctive conditionals. To say that a given physical object exists is not to say that there exists some entity of an ontologically different sort from that with which we are directly acquainted in sensation; nor, of course, is it to say that someone is actually having some sensation. Rather, we should view such propositions as equivalent in meaning to assertions about what sensations or sequences of sensation a subject would have were he to have certain others. The truth or falsity of such



propositions is mind-independent and so with this analysis we could secure the mind-independent status of the physical world. What is more, to the empiricist's great relief, it looks as if such propositions could be established inductively. Subjunctives that assert connections holding between sensations can presumably be justified without having to correlate anything but sensations.

#### OBJECTIONS TO PHENOMENALISM

Many philosophers today would reject the epistemological, ontological and metaphysical presuppositions with which phenomenologists approached the problem of perception. Foundationalism is hardly the received view in contemporary epistemology, and even those sympathetic to a kind of foundationalism are more likely to embrace an externalist version of the view, a version that circumvents the sceptical problems the phenomenologist was trying to solve with a phenomenological analysis. Other philosophers reject the conception of philosophical analysis as an attempt to specify necessary and sufficient conditions for the truth of propositions (see PARADOXES OF ANALYSIS). Some of these embrace an externalist or causal theory of meaning which makes the meaning of expressions in our language largely an empirical question inaccessible to an *a priori* method of analysis.

Still, the idea behind phenomenalism was, in the abstract, enormously attractive for many philosophers sympathetic to the presuppositions of radical empiricism and to suggest that it is rejected for the above reasons is probably to put the cart before the horse. In fact, it may have been the widespread conclusion that phenomenalism is false that led to rejection of some of the philosophical presuppositions on which it rested. Many of the objections which led to a general abandonment of phenomenalism surfaced with the attempt to spell out in detail how the theory is supposed to work.

One preliminary difficulty concerns the analysis of contingent subjunctive conditionals. The concepts of a law of nature, causation and

the relation expressed by contingent subjunctive conditionals are all closely intertwined, and the problem has been to find an analysis of one of these concepts that does not presuppose an understanding of the others. The relationship between laws of nature and contingent subjunctive conditionals is also the source of at least two objections to phenomenalism. Some philosophers would argue that the subjunctive conditionals employed in the phenomenologist's analysis presuppose the existence of lawful regularities between patterns of sensations. But it is problematic to suppose that there are any genuine laws whose antecedents and consequents refer only to the occurrence of sensations. Unless the antecedents of such generalizations are protected by qualifications referring to the *physical* conditions under which certain sensations have occurred, it is implausible, the argument goes, to suppose that there are any lawful consequences concerning subsequent sensations. But if the assertion of such conditionals carries with it implicit reference to physical conditions, the phenomenologist programme of fully reducing talk about the physical world to talk about sensation fails. Furthermore, if anything like a regularity theory of law were correct, the existence of lawful connections between patterns of sensations might presuppose once again the existence of minds and sensations and thus make both the truth of the phenomenologist's subjunctives and the propositions about the physical world they are supposed to analyse parasitic upon the existence of minds – just what the phenomenologist is trying to avoid with an alternative to idealism. The phenomenologist, you will recall, is trying to accommodate the commonsense conception of the physical world as something that could have existed even in the absence of any conscious beings.

The above are certainly concerns, but the argument against phenomenalism that was so widely viewed as decisive is the argument from perceptual relativity, most clearly and concisely presented by Roderick Chisholm (1948) (see CHISHOLM). Chisholm offers, in effect, a strategy for attacking any phenomenological analysis. The first move in the strategy

is to force the phenomenalist into giving at least one example of an alleged analytic consequence (expressed in purely phenomenal language) of a proposition asserting the existence of some physical object. When one gets the example, one simply describes a hypothetical situation in which, though the physical object proposition is true, its alleged analytic consequence would obviously be false. If the physical object proposition really did entail the experiential proposition, then there could be no hypothetical situation in which the one is true and the other false, and so we would have constructed a *reductio* of the proposed analysis. C. I. Lewis (1946, p. 240), for example (see LEWIS), claimed that the proposition that there is a doorknob in front of me and to the left (*p*) entails the proposition that if I should seem to see such a doorknob in front of me and to the left and should seem to be initiating a certain grasping motion, then in all probability the feeling of contacting a doorknob would follow (*r*). (Lewis used the “seems to perceive” terminology to report the occurrence of mind-dependent sensations.) Chisholm argues that *p* does not entail *r*, for there is another proposition *q* (the proposition that I am unable to move my limbs and my hands but am subject to delusions such that I think I am moving them; I often seem to myself to be initiating a certain grasping motion, but when I do I never have the feeling of contacting anything), which is obviously consistent with *p* and which when conjoined with *p* entails not-*r*.

#### RESPONSES

To escape the argument from perceptual relativity the phenomenalist might try to “protect” the conditionals he employs in his analysis with a “normal or standard conditions” clause added to the antecedents. It is crucial, however, that the phenomenalist not refer to the external or internal *physical* conditions of the subject for to do so would be to violate the conditions for a successful phenomenalist analysis. A primary goal of phenomenism was to reduce (completely) talk about the physical world to talk about sensa-

tions so as to make physical object propositions epistemically accessible. Can one protect the antecedents of the conditionals without resorting to physical description? One option would be to embed conditionals within conditionals but since we know that such conditionals are really designed to eliminate the existence of various physical conditions that might distort the normal sequence of sensations, we also know that any subjunctives designed to guarantee the absence of such distorting conditions will fail to do so because they will themselves be subject to distorting physical conditions.

There is one other avenue open to the phenomenalist, though this involves a subtle move away from classical versions of the view. One can attempt to introduce a normal or standard conditions clause in the antecedents of the phenomenalist’s subjunctive whose purpose is to *denote* those conditions *whatever they are* that normally (defined statistically) accompany certain sequences of sensations. The conditions denoted by such a clause might include other facts about what sensations would follow others, facts about Kantian things-in-themselves, or facts about the intentions of a Berkeleian God. The claim might be made that even though a normal conditions clause of this sort involves denoting things in ontological categories other than sensations, such denotation is epistemically harmless, for one is always justified in believing, *ceteris paribus*, that things are as they usually are.

Once one modifies phenomenism enough to allow into the phenomenalist analysis expressions that might denote things other than sensations, however, one can argue that we might as well embrace a version of a causal theory of objects that is much more closely related to classical phenomenism than the more familiar representative realism (see REPRESENTATIVE REALISM) replete with its primary/secondary quality distinction and a conception of objects that *resemble* in important respects the contents of our minds. On this “phenomenalist” causal theory, to assert the existence of a physical object is to assert the existence of a thing (whatever it is – its intrinsic character might be in principle

unknowable) that has the potential to produce certain sensations and that would produce certain sequences of sensations were it to produce certain others *under normal conditions*. This version of a causal theory contains no ontological commitments that extend beyond our modified phenomenism and it seems that it allows one a much more natural way of analyzing bare existential statements, e.g. there exists a table (somewhere, some time). Such statements are a nightmare for classical phenomenism for they provide no "setting" that makes even *prima facie* plausible the entailment of any conditional about what any particular subject would experience. Because the causal theorist's analysis of such statements begins with the bare existential claim about the existence of a potential cause of sensations, we eliminate that problem. Indeed, when Mill identified objects with the permanent possibilities of sensations he may well have been pointing not to classical phenomenism but to the causal theory that is closely related to it. Notice that this causal theory faces precisely the same problem of perceptual relativity as pure versions of phenomenism. One still needs a way of specifying subjunctively the "powers" that define the cause as a physical object of a given kind, and to avoid a regress one must define such powers without presupposing an understanding of physical object propositions. If so, two views long considered radically different may have a vested interest in finding common solutions to common problems.

See also FIRTH; LOGICAL POSITIVISM; PROBLEM OF THE EXTERNAL WORLD.

## BIBLIOGRAPHY

- Ayer, A. J.: *Language, Truth and Logic*, 2nd edn (New York: Dover, 1946).  
 Berkeley, G.: *Three Dialogues between Hylas and Philonous* (1713), in *Berkeley: Philosophical Works*, ed. M. R. Ayers (London: Dent, 1975).  
 Chisholm, R.: "The Problem of Empiricism," *Journal of Philosophy* 45 (1948), 512–17.  
 Hume, D.: *A Treatise of Human Nature* (1739–40); ed. L. A. Selby-Bigge, revised P. H. Nidditch (Oxford: Oxford University Press, 1978).

Lewis, C. I.: *An Analysis of Knowledge and Valuation* (LaSalle, IL: Open Court, 1946).

Mackie, J. L.: "What's Really Wrong with Phenomenism," *Proceedings of the British Academy* 55 (1969), 113–27.

Mill, J. S.: *An Examination of Sir William Hamilton's Philosophy* (London: Longmans Green, 1867).

RICHARD FUMERTON

**phenomenology** As summed up in Husserl's slogan, "To the things themselves!" (see HUSSERL), the aim of phenomenology is to bypass the presuppositions built into traditional theories (including psychology, physiology and epistemology) in order to describe what shows up in the flow of lived experience prior to reflection. The focus is solely on the essential structures of experience itself. The key discovery is that all forms of consciousness are characterized by "intentionality", a directedness toward things such that consciousness is always of or *about* something. Husserl distinguished the intentional act (*noesis*), which is occasional and transient, from the act's content (*noema*), the timeless, intersubjective object-as-meant. Later phenomenologists like Heidegger (see HEIDEGGER) and Merleau-Ponty (see MERLEAU-PONTY) rejected the distinctions of consciousness vs. object or act vs. content, and tried to describe the "natural conception of the world" prior to the *epoché* (see EPOCHE).

CHARLES GUIGNON

**philosophical knowledge** A traditional view of philosophical knowledge can be sketched by comparing and contrasting philosophical and scientific investigation, as follows. The two types of investigations differ both in their methods (the former is *a priori*, and the latter *a posteriori*) and in the metaphysical status of their results (the former yields facts that are metaphysically necessary and the latter yields facts that are metaphysically contingent). Yet the two types of investigations resemble each other in that both, if successful, uncover new facts, and these facts, although *expressed* in language, are generally not *about* language (except for investigations

in such specialized areas as philosophy of language and empirical linguistics).

This view of philosophical knowledge has considerable appeal. But it faces problems. First, the conclusions of some common philosophical arguments seem preposterous. Such positions as that it is no more reasonable to eat bread than arsenic (because it is only in the past that arsenic poisoned people), or that one can never know he is not dreaming, may seem to go so far against commonsense as to be for that reason unacceptable. Second, philosophical investigation does not lead to consensus among philosophers. Philosophy, unlike the sciences, lacks an established body of generally agreed upon truths. Moreover, philosophy lacks an unequivocally applicable method of settling disagreements. (The qualifier "unequivocally applicable" is to forestall the objection that philosophical disagreements are settled by the method of a priori argumentation; there is often unresolvable disagreement about which side has won a philosophical argument.)

In the face of these and other considerations, various philosophical movements have repudiated the above traditional view of philosophical knowledge. Thus, verificationism (see VERIFICATIONISM) responds to the unresolvability of traditional philosophical disagreements by putting forth a criterion of literal meaningfulness that renders such questions literally meaningless: "A statement is held to be literally meaningful if and only if it is either analytic or empirically verifiable" (Ayer, 1952, p. 9), where a statement is analytic iff it is just a matter of definition, and traditional controversial philosophical views, such as that it is metaphysically impossible to have knowledge of the world outside one's own mind, would count as neither analytic nor empirically verifiable (see LOGICAL POSITIVISM).

Various objections have been raised to this verification principle. The most important is that the principle is self-refuting, i.e. that when one attempts to apply the verification principle to itself, the result is that the principle comes out as literally meaningless (hence not true) because it is neither empirically verifiable nor analytic. This move may

seem like a trick. But it reveals a deep methodological problem with the verificationist approach. The verification principle is intended to delegitimize all controversy that is not resolvable either empirically or by recourse to definitions. The principle itself, however, cannot be established either empirically or by recourse to definitions. The principle is an attempt to rule out synthetic a priori controversy; yet the principle itself is both synthetic a priori and controversial. It is ironic that the self-refutingness of the verification principle is one of the very few points on which philosophers nowadays approach consensus.

Ordinary language philosophy, another twentieth-century attempt to delegitimize traditional philosophical problems, faces a parallel but unrecognized problem of self-refutingness. Just as verificationism can be characterized as reacting against unresolvable a priori controversy, ordinary language philosophy can be characterized as reacting against a priori counterintuitiveness. The ordinary language philosopher rejects counterintuitive philosophical positions (such as the view that time is unreal or that one can never know anything about other minds) by saying that these views "*go against ordinary language*" (Malcolm, in Rorty, 1970, p. 113); i.e. that these views go against the way the ordinary person uses such terms as "know" and "unreal", since the ordinary person would reject the above counterintuitive statements about knowledge and time. On the ordinary language view, it follows that the sceptic does not mean the same thing by "know" as does the non-philosopher, since they use the terms differently and meaning is use. Thus, on this view, sceptics and anti-sceptics no more have a non-linguistic disagreement about knowledge than someone who says "Banks are financial institutions" and someone who says "Banks are the shores of rivers" have a non-linguistic disagreement about banks.

An obvious objection here is that many factors besides meaning help to determine use. For example, two people who disagree about whether the world is round use the word "round" differently in that one applies it to the world while the other does not; yet

they do not thereby mean different things by “world” or “round”. Ordinary language philosophy allows that this aspect of use is not part of the meaning, since it rests on a disagreement about empirical facts. But in relegating all non-empirical disagreements to differences in linguistic meaning, the ordinary language philosopher denies the possibility of substantive, non-linguistic disagreement over a priori facts and thus, like the verificationist, disallows the synthetic a priori. Malcolm claims that “if a child who was learning the language were to say, in a situation where we were sitting in a room with chairs about, that it was ‘highly probable’ that there were chairs there, we should smile *and correct his language*” (1970, p. 116). Malcolm may be right about this particular case, since it is so unlikely that a child would have independently developed a sceptical philosophy. But a parallel response seems obviously inappropriate as a reply to a philosopher who says “One can never know that one is not dreaming”, or, for that matter, as a reply to an inept arithmetic student who says “ $33 = 12 + 19$ ”. If it were true that a philosopher uttering the first of these sentences were not using “know” in the usual sense, he could not convey his philosophical views to a French speaker by uttering that sentence’s French translation (“On ne peut jamais savoir qu’on ne rêve pas”), any more than one can convey his 8-year-old cousin Mary’s opinion that her teacher is vicious by saying “Mary’s teacher is viscous” if Mary wrongly thinks “viscous” means vicious and is using it that way. But it seems obvious that failure to translate “know” or its cognates into their French synonyms would prevent an English-speaking sceptic from accurately representing his views in French at all. The ordinary language view that all non-empirical disagreements are linguistic disagreements entails that if someone believes the sentence “a is F” when this sentence expresses the a priori proposition that a is F, then having the property he takes “F” to express is part of what he means by “a”. But this obviously goes against the Malcolmian “ordinary use” of the term “meaning”, i.e. what ordinary people, once they understand the term “meaning”, believe on a priori

grounds about the extension of the term “meaning”. For example, the ordinary man would deny that the inept student mentioned above cannot be using his words with our usual meaning when he says “ $33 = 12 + 19$ ”. Like the earlier objection of self-refutingness to verificationism, this objection reveals a deep methodological problem. Just as synthetic a priori controversy cannot be ruled out by a principle that is both synthetic a priori and controversial, a priori counterintuitiveness cannot be ruled out by a principle that is both a priori and counterintuitive.

Although verificationism and ordinary language philosophy thus are both self-refuting, the problems that helped motivate these positions need to be addressed. What are we to say about the fact that (a) many philosophical conclusions seem wildly counterintuitive and (b) philosophical investigations do not lead to philosophical consensus?

To put the first problem in perspective, it is important to see that even highly counterintuitive philosophical views generally have arguments behind them – arguments that “start with something so simple as not to seem worth stating”, and proceed by steps so obvious as not to seem worth taking, before “[ending] with something so paradoxical that no one will believe it” (Russell, 1956, p. 193). But since repeated applications of commonsense can thus lead to philosophical conclusions that conflict with commonsense, commonsense is a problematic criterion for assessing philosophical views. It is true that, once we have weighed the relevant arguments, we must ultimately rely on our judgment about whether, in the light of these arguments, it just seems reasonable to accept a given philosophical view. But this truism should not be confused with the problematic position that our considered philosophical judgment in the light of philosophical arguments must not conflict with our commonsense *pre-philosophical* views.

As for philosophers’ inability to reach consensus, it is important to see that this does not entail that there is no fact of the matter as to who is right. There are other possible explanations for this inability (*see* Rescher, 1978). Moreover, supposing that the existence of



unresolvable a priori disagreement over the truth of *p* shows that *p* lacks a truth-value would make the matter of whether *p* has a truth-value too dependent on which people happen to exist and what they can be persuaded to believe.

Both verificationism and ordinary language philosophy deny the synthetic a priori. Quine (see QUINE) goes further; he denies the analytic a priori as well; he denies both the analytic–synthetic distinction and the a priori–a posteriori distinction. In “Two Dogmas of Empiricism” Quine considers several reductive definitions of analyticity synonymy, argues that all are inadequate, and concludes that there is no analytic and synthetic distinction. But clearly there is a substantial gap in this argument. One would not conclude from the absence of adequate reductive definitions of “red” and “blue” that there is no red-blue distinction, or no such thing as redness. Instead, one would hold that such terms as “red” and “blue” are defined by example. But this also seems plausible for such terms as “synonymous” and “analytic” (Grice and Strawson, 1956).

On Quine’s view, the distinction between philosophical and scientific inquiry is a matter of degree. His later writings indicate that the sort of account he would require to make analyticity, necessity, or a priority acceptable is one that explicates these notions in terms of “people’s dispositions to overt behavior” in response to socially observable stimuli (Quine, 1969, p. 29). A discussion of this behaviourism (see BEHAVIOURISM), however, lies beyond the scope of this essay.

See also ANALYTICITY; NATURALIZED EPISTEMOLOGY; PSYCHOLOGY AND EPISTEMOLOGY; RYLE.

#### BIBLIOGRAPHY

- Ayer, A. J.: *Language, Truth and Logic* (New York: Dover, 1952).  
 Grice, H. P. and Strawson, P. F.: “In Defense of a Dogma,” *Philosophical Review* 65 (1956), 141–68.  
 Quine, W. V.: “Two Dogmas of Empiricism,” in his *From a Logical Point of View* (Cambridge, MA: Harvard University Press, 1953), ch. 2.

- Quine, W. V.: *The Ways of Paradox and Other Essays* (New York: Random House, 1966), chs 18 and 20.  
 Quine, W. V.: *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969).  
 Rescher, N.: “Philosophical Disagreement,” *Review of Metaphysics* 22 (1978), 217–51.  
 Rorty, R., ed.: *The Linguistic Turn* (Chicago, IL: University of Chicago Press, 1970).  
 Russell, B.: “The Philosophy of Logical Atomism,” in his *Logic and Knowledge*, ed. R. C. Marsh (London: Allen and Unwin, 1956), 177–281.  
 Wittgenstein, L.: *Philosophical Investigations* (Oxford: Blackwell, 1953).

FELICIA ACKERMAN

**Plato (c.429–347 BC)** Greek philosopher, founder of the Academy at Athens. Plato’s thinking about epistemology has two aspects. One concerns questions about what statements a person is justified in making and what he can justifiably claim to know. Here there is discussion of the grounds on which a statement might be established or refuted, and of the kinds of arguments that can be given for or against something. This part of Plato’s view is quite close to modern epistemology. The second aspect of his thinking is much closer to what is nowadays thought of as metaphysics. This part grows mainly out of the question, What must we and the world be like if it is to be the case that we know something? Here the issues have to do, not so much with ways of refuting and establishing statements, but rather with what appropriate conditions on knowledge are and with what the world and we ourselves must be like if we are to be able to fulfil them. It seems to Plato crucial that knowledge be thought of as concerning objective facts that hold non-relatively to a particular observer or circumstance of observation.

The first aspect of Plato’s epistemology was strongly influenced by Socrates, particularly that part of his activity that suggests a sceptical stance. Socrates is said to have claimed that he knew either nothing or very little (e.g. *Meno* 80, 86). In Plato’s works this claim is loosely associated with the portrayal of Socrates as practising a procedure of

*elenchus*, “scrutiny” (often also translated “refutation”). Arguably this procedure is capable only of refuting a statement, never of establishing one (except perhaps the negation of the statement refuted, though the state of logic in Plato does not allow this issue to be fruitfully raised). At any rate Socrates gained a reputation for skill at refuting things that others said, and he apparently practised the skill on himself, thereby becoming disinclined to make claims to knowledge.

This aspect of Plato’s thinking appears mostly in earlier works, such as the *Laches*, the *Euthyphro* and parts of the *Meno*, as well as *Republic* I, which is probably not early but mimics his earlier works. (The classification of Plato’s works as early, middle and late involves some controversy and conjecture, but is widely enough agreed upon to be presupposed here.)

The first aspect of Plato’s epistemology shares important preoccupations with epistemology since Descartes. First, as indicated, Socrates is portrayed as finding reason to reject most claims to knowledge. Next, Plato’s early investigations are launched mainly from a first-person standpoint: one must ask whether one’s own judgements are justified, and try to find justifications that will seem compelling to oneself. In this way they adopt a standpoint similar to that of what are now called “internalist” epistemological views (see EXTERNALISM/INTERNALISM). Third, it is often said to be important to make statements on the basis only of what oneself believes, and not to rely on hearsay or other’s beliefs. In addition, although Plato does not explicitly introduce the notion of “doubt”, he emphasizes the instability of most beliefs, i.e. the fact that they can often be easily dislodged by argument or other kinds of persuasion.

The second aspect of Plato’s epistemology is prominent in his middle period, when he is usually taken to have moved beyond Socrates’ influence and to have developed a positive doctrine of his own. This phase begins in the *Meno*, continues in the *Phaedo*, the *Symposium* and the *Republic*, and shows itself also in the *Timaeus* (usually regarded as one of Plato’s last works).

This part of Plato’s thinking is given over to developing a metaphysical theory about how we and the world must be if we are to have knowledge (*episteme*). It is either assumed or perhaps argued (see *Phdo.* 74, *Rep.* 475–80, *Tim.* 51) that we do have knowledge, and taken as obvious that knowledge is distinct from certain other cognitive states, in particular “belief” or “opinion”, *doxa*. Plato then tries to show what the world must contain, and what we ourselves must be like, if these two things are to hold. In this phase virtually no attempt is made at dogged socratic scrutiny of Plato’s own theory, at least until the *Theaetetus* and the *Parmenides*. In those enigmatic works Plato raises various questions about what state of affairs must obtain if knowledge is to be possible, including (it seems) questions about whether his own theory really accounts for that possibility. In the end it remains unclear what his answers to these questions are.

Before further examining this second phase of Plato’s thinking, which produces his most influential ideas in epistemology, it will be helpful to consider how it arose from the first phase.

Virtually without exception, each of Plato’s early works is dominated by an attempt to define some notion, usually an evaluative notion like moderation, courage, holiness, virtue, and so on. Most of the refutations appearing on those works are directed at the definitions that are offered, and these do always end up refuted. As Plato presents things, Socrates’ view was that before a person can know certain facts about virtue (for example, how it is acquired), one has to know a definition of it, or, as Plato also says, to know “what it is” (*Protagoras* 360–1, *Rep.* 354). In early dialogues a definition is often refuted by citing something that fits the definiendum but not the definiens, or vice versa. Sometimes, though, a definiens will be attacked for being not sufficiently clearer than the definiendum is (*Meno* 75–7), or for introducing some kind of circularity (e.g. *Meno* 79).

Plato’s middle works, however, are not organized wholly around searches for definitions. Rather, they expound a metaphysical

theory that is designed to show how knowledge is possible. It seems natural to suppose that the failure of earlier searches for definitions somehow led him to construct this theory. The way in which this happened, however, is not easy to reconstruct. One possibility is that the theory is supposed to provide a way of discovering definitions which does not rely on the use of *elenchus*. Another possibility is that Plato decided that finding definitions is not, after all, the first step of the epistemological enterprise, and hoped that his theory would show a different way to gain knowledge, and eventually also definitions, of the things that he had earlier taken to need defining at the first stage of any discussion involving them. Thus, for example, the *Republic* is able to say quite a lot about justice before it finally (at 443) arrives at an account (and even then not a formally stated one) of what justice is.

Plato's theory is built around the idea that there are objects, often called "Forms" (*eide*), that we do not sense but can have knowledge about. The idea of knowledge here is closely intertwined, in a way not easy to disentangle, with a notion of the understanding of terms. Some of this knowledge-*cum*-understanding is expressed in the form of propositions, for example the proposition that the Good is not pleasure (*Rep.* 505), and sometimes it is described as a kind of intellectual "viewing" (500, 511) – which shares something with (and led historically to) Frege's notion of "grasping" a sense and Russell's notion of "acquaintance" with a universal.

Much of Plato's metaphysics of Forms arises from his reasons for thinking that knowledge in the proper sense cannot consist simply in the use of the senses and cannot concern sensible objects. On the other hand, this view leads him into severe problems, many of which are investigated in his later works.

In particular it strikes him as important that a sensible object that has a certain predicate true of it can both present the simultaneous appearance of having the contrary predicate true of it (*Phdo.* 74, *Rep.* 479), and also can have had, or come to have, that contrary predicate true of it at some other time (*Tim.* 49–50). For this reason he tries to con-

struct his theory about Forms so that unlike a sensible, a Form will be depicted as incapable, at least under favourable circumstances, of presenting to the mind any appearance that is, so to speak, contrary to that of being the particular Form that it is. Thus, for example, the Form of equality is said to be incapable of appearing to be inequality (*Phdo.* 74), and Forms are in general said to be incapable of change whereas sensible objects are said to be constantly changing (*Phdo.* 78–9, *Rep.* 484–5, *Tim.* 51). However, at certain points in his later works Plato may suggest that nothing at all can satisfy this stringent condition (*Soph.* 248–9), and that therefore his view about Forms may need either more careful formulation, qualification, or rethinking. This same line of thought seemingly makes Plato suggest that it is possible to have knowledge only about Forms, and that knowledge about sensible objects is impossible (*Rep.* 477–80, *Tim.* 51). Moreover he also seems to hold sometimes that we cannot have about Forms that kind of cognition, belief or opinion, that we do have of sensibles (*ibid.*). On the other hand, he allows that it is possible to make mistakes about Forms, and also to be in a cognitive state vis-à-vis a Form that seems indistinguishable from what he seems obliged to call false belief or opinion. This idea, too, requires some kind of further explanation of the distinction between Forms and sensibles – a requirement that Plato seems to see some difficulty in satisfying (e.g. *Tht.* 187–200).

Although in this phase of his work Plato concentrates, as indicated, on constructing a metaphysics that will make room for the possibility of knowledge, he does at the same time pay some attention to the problems that are characteristic of the first phase of his epistemology. In the *Meno*, the *Phaedo* and the *Republic*, he develops what has been called the "method of hypothesis", which seems to be a method by which things can ultimately be demonstrated, unconditionally. In the *Meno* (87–9) and the *Phaedo* (99–101) he indicates that hypotheses are to be accepted only provisionally and not regarded as certain or unreviseable. In the *Republic*, however, he seems to maintain that one can somehow

reach an “unhypothesized” principle (511) which will somehow serve as the basis for demonstrating everything hitherto accepted merely hypothetically. He apparently implies that what is demonstrated thereby will have to do only with Forms (ibid.). He also makes a suggestion, not clearly explained, that this “principle” (*arche*) has something to do with the Form of the Good (533–4). There is no generally accepted interpretation of what Plato says here, but it seems to indicate that he accepted or was seriously considering some kind of *foundationalist* epistemological position, which would start from some unshakable principle and derive from it the rest of what there is to be known about Forms. (As often, however, Plato here seems to waver between thinking of the principle and what is to be derived as possessing propositional structure and treating them as non-propositionally structured objects.)

This method of hypothesis is earlier offered as something that is used by “dialectic”, the style of philosophizing that takes place through conversational questions and answers (see DIALECTIC). In later works, however, such as the *Phaedrus* and the *Sophist*, Plato says that dialectic makes use of the “method of collection and division”, which is a method for constructing taxonomies and definitions. Nothing in his description of this method, however, indicates that it could be used to demonstrate or justify definitions. (In *Posterior Analytics* II. 5 Aristotle attacks the contention, presumably due to someone other than Plato, that division could be used to demonstrate the correctness of a definition.) Interestingly, Plato here never tries to show clearly how such definitions might be subjected to, or could stand up against, the sort of scrutiny that is practised in the early dialogues.

Plato’s most concerted effort to explain what knowledge is appears in the *Theaetetus*. This work has not yet been given a satisfactory interpretation. It represents itself as a failed attempt, in the manner of earlier dialogues, to find a definition, this time, of knowledge. Among other points of obscurity, it is disputed whether the *Theaetetus* accepts, is neutral on or rejects the metaphysical

views that dominate the *Phaedo*, *Republic*, *Symposium* and *Timaeus*. In addition, the work is divided into sections whose relation to each other, as well as to Plato’s views as expressed elsewhere, is not fully clear. The first part (to 186) is a refutation of the thesis that knowledge is perception (*aisthesis*, a word that may also be translated “sensation”). The next part (187–200) is a failed attempt to defend the possibility of false belief and particularly false identity-beliefs against some objections. The last part (200–10) is a failed attempt to define knowledge as “true belief with an account (*logos*)”.

The first part of the work plainly has something to do with Plato’s claim in his middle period that knowledge has to do not with sensible objects but with Forms. But although this part of the dialogue definitely denies that to know is to perceive, it does not explicitly deny that there is knowledge concerning sensible objects, and some interpreters take the dialogue as a whole, on the contrary, to imply or even to assert that there is.

The latter two parts of the dialogue do not explicitly mention Forms, but many interpreters have thought that they tacitly deal with them. It seems more likely that they intended to be neutral on the question whether Forms are the only objects of knowledge, and to deal with problems that arise for any kind of objects that it might reasonably be taken to be possible to have knowledge about, whether these be Forms or something else. Indeed, the problems discussed here are more abstract than the ones that Plato deals with when he is expounding his metaphysics of knowledge, as in the *Republic*, and even when he is scrutinizing it and raising problems for it, as in the *Parmenides*. Indeed, some of the issues raised in the *Theaetetus*, and also in the *Sophist*, involve questions about the conditions under which a person can be said to have designated, or be thinking about, a particular object or concept. As such, they would nowadays be associated less with epistemology than with the philosophy of language. On the other hand, *Theaetetus* 187–200 seems to presuppose, against what seems to be suggested in the *Phaedo* and the

*Republic* (see *supra*), that an object of knowledge can in some sense present to the mind an appearance somehow conflicting with its being the thing that it is.

Although like early works the *Theaetetus* aims unsuccessfully at a definition, its main concern is with the metaphysics of thought and knowledge, and it is not centred on questions about justification of belief or knowledge. It thus shows Plato to have moved farther away than ever from concerns that were characteristic of the first phase of his epistemological thinking.

#### WRITINGS

- Platonis Opera*, ed. J. Burnet, 5 vols (Oxford: Clarendon Press, 1900–7).  
*The Collected Dialogues of Plato*, ed. E. Hamilton and H. Cairns (Princeton, NJ: Princeton University Press, 1961).

#### BIBLIOGRAPHY

- Bostock, D.: *Plato's Theaetetus* (Oxford: Clarendon Press, 1988).  
 Burnyeat, M.: *The Theaetetus of Plato* (Indianapolis, IN: Hackett, 1990).  
 Cooper, J. H.: "Plato on Sense Perception and Knowledge," *Phronesis* 15 (1970), 123–46.  
 Fine, G.: "Knowledge and Belief in *Republic* V," *Archiv für Geschichte der Philosophie* 58 (1978), 121–39.  
 Matthews, G. B. and Blackson, T. A.: "Causes in the *Phaedo*," *Synthesis* 79 (1989), 581–91.  
 McDowell, J.: *Plato, Theaetetus* (Oxford: Clarendon Press, 1973).  
 Penner, T.: *The Ascent from Nominalism* (Dordrecht: Reidel, 1987).  
 Robinson, R.: *Plato's Earlier Dialectic*, 2nd edn (Oxford: Clarendon Press, 1953).  
 Vlastos, G.: "Degrees of Reality in Plato," in *New Essays on Plato and Aristotle*, ed. R. Bambrough (New York: Routledge, 1965), 1–18.  
 White, N. P.: *Plato on Knowledge and Reality* (Indianapolis, IN: Hackett, 1976).  
 White, N. P.: "Plato's Epistemological Metaphysics," in *Cambridge Companion to Plato*, ed. R. Kraut (Cambridge: Cambridge University Press, 1991).

NICHOLAS WHITE

**Popper, Karl (1902–94)** Austrian-born philosopher who settled in Britain. Although widely regarded as primarily a philosopher of science and then a political philosopher, Popper might equally be regarded as first and foremost an epistemologist in that his conclusions in these other areas derive from his epistemological stance. Indeed, in his early *Logic of Scientific Discovery* he writes that "the main problem of philosophy is the critical analysis of the appeal to the authority of experience" (pp. 51–2). In the preface to the same work, Popper writes that the central problem of epistemology is the problem of the growth of knowledge, and that this is best-studied by studying the growth of scientific knowledge. His philosophy of science may then be seen as part of his epistemology, rather than vice versa. And both his attitude to politics and society and his later work in the philosophy of biology derive from his epistemology.

Popper's epistemology itself rests on a deep scepticism about the validity of inductive inference – a scepticism which, in the opinion of many critics, undermines his own attempt to establish a non-inductive account of knowledge (see PROBLEMS OF INDUCTION). On induction, Popper is basically Humean, but instead of following normal practice and attempting to rebut Hume directly, Popper develops an epistemology which supposedly does without induction altogether. The fundamental insight here is to exploit the asymmetry in empirical matters between proof and disproof. While no universal empirical theory can be proved, owing to our ignorance of the totality of phenomena, a universal theory can be disproved by only one counter-instance to it. Popper believed that not only the grand theories of mature science should be regarded as universal theories in the relevant sense. Even the statement "Here is a glass of water" involves universal terms ("glass", "water") and cannot be verified by observational experience, as it implies law-like behaviour on the part of the object in question – behaviour which is beyond our powers to verify fully. Nevertheless, just as one planet moving once in an elliptical orbit would be enough to refute the theory that all planets move in circles, so one unglass-like bit



of behaviour would convince us that what we had before us was no glass.

How, then, are we to proceed epistemologically? Popper's view is that we should consciously seek theories which are falsifiable by counter-evidence; that we should seek to falsify them, and those which survive testing should then be tentatively accepted and regarded as corroborated or closer to the truth than the ones which have been falsified. How this method – of conjecture and refutation, Popper calls it – differs from standard types of induction in which we also rely on severely tested theories is something which has escaped most critics of Popper. Others have questioned his right to base anything as substantial as the falsification of a theory on singular observations, given his belief that they too are theory-laden – a problem Popper wrestles with somewhat inconclusively in *The Logic of Scientific Discovery*.

If Popper's specific attempt to solve the problem of induction through a non-inductive epistemology has convinced few of his readers, his approach to science and knowledge generally has won many followers. The idea that we cannot know but only guess, and that we should avoid dogmatism (see DOGMATISM) in science and human affairs more generally, is attractive on a number of levels.

First, it sets human knowledge in a Darwinian context, something to which Popper increasingly turned his attention in later years. Our perceptual faculties and our theories are seen *modo biologico*, as natural or human attempts to anticipate the environment, attempts which will be weeded out if the environment does not match them sufficiently well or if the environment itself changes (see EVOLUTIONARY EPISTEMOLOGY; NATURALIZED EPISTEMOLOGY).

Second, Popper's fallibilism (see FALLIBILISM) led him to see science itself as a construction of the human imagination, rather than something simply read passively and mechanically off the facts of nature. This aspect of his thinking has been found liberating by many scientists. Connected to his fallibilism concerning science, though more controversial, are Popper's famous attempts to demarcate true science from pseudo-science in

terms of empirical falsifiability and also his antiessentialism (see NECESSARY/CONTINGENT). The former view would have true scientists not just proposing bold theories but, crucially, then seeking to reject them (in contrast, according to Popper, to proponents of such influential systems as Marxism and psychoanalysis). Unfortunately for the demarcation criterion, though, it seems that even the physical sciences need a higher degree of dogmatism than Popper's criterion can allow. His anti-essentialism amounts to little more than the view that, as any current theory of science is likely to turn out to be false, we should not regard it as giving us the essence of the world. But it does not follow from this that science and scientists should not be seeking natural kinds or essences.

Finally, Popper's attitude to politics, adumbrated passionately and memorably in *The Open Society and Its Enemies*, is that any social policy, like any other proposal, is bound to have unsuspected and unintended consequences. Society should, therefore, be organized so that criticisms can be heard from those affected by policies, and governments should be changed peacefully in response to the wishes of the governed. From our point of view, what is significant here is not so much the analysis Popper gives of the open society as the way his social philosophy follows directly from his epistemological fallibilism.

See also HISTORICISM; NATURAL SCIENCE.

## WRITINGS

*The Logic of Scientific Discovery* (London: Hutchinson, 1959); his translation of *Logik der Forschung* (Vienna, 1935).

*The Open Society and Its Enemies*, 2 vols (London: Routledge and Kegan Paul, 1945).

*Conjectures and Refutations* (London: Routledge and Kegan Paul, 1963).

*Objective Knowledge* (Oxford: Clarendon Press, 1972).

*Realism and the Aim of Science* (London: Hutchinson, 1983).

"Intellectual Autobiography" and "Replies to Critics," in *The Philosophy of Karl Popper*, ed. P. A. Schillp (LaSalle, IL: Open Court, 1974), 1–181 and 959–1197.

## BIBLIOGRAPHY

- Lakatos, I. and Musgrave, A., eds: *Criticism and the Growth of Knowledge* (Cambridge: Cambridge University Press, 1970).  
 O'Hear, A.: *Karl Popper* (London: Routledge and Kegan Paul, 1980).  
 Stove, D. C.: *Popper and After: Four Modern Irrationalists* (Oxford: Pergamon Press, 1982).  
 Watkins, J. W. N.: *Science and Scepticism* (London: Hutchinson, 1984).

ANTHONY O'HEAR

**positivism** see LOGICAL POSITIVISM.

**pragmatism** A (primarily American) school of philosophy initiated by Peirce (see PEIRCE) and James (see JAMES); characterized by the "pragmatic maxim", according to which the meaning of a concept is to be sought in the experiential or practical consequences of its application. The epistemology of pragmatism is typically anti-Cartesian, fallibilistic, naturalistic; in some versions it is also realistic, in others not.

"It has probably never happened", Peirce wrote in 1905, "that a philosopher has attempted to give a general name to his own doctrine without that name's soon acquiring in common philosophical usage, a signification much broader than was originally intended." His "pragmatism", he continued, had by then acquired a signification so much broader than, and indeed so much at odds with, his original intention, that it was "time to kiss his child good-bye" and "to announce the birth of the word 'pragmaticism', which is ugly enough to be safe from kidnappers" (*Collected Papers* 5.414).

As this suggests, the problem of giving an accurate brief characterization of the philosophical tendencies known as "pragmatism" is far from trivial. It is hard enough to specify what important philosophical ideas were shared by Peirce and James, the founders of pragmatism; harder yet to find a characterization that would also comfortably accommodate Dewey, Schiller and Mead; nearly impossible to extend it to include more recent neo-pragmatists and sympathizers, overt and covert, as diverse as Ramsey, Lewis, Quine,

Sellars, Putnam, Apel, Rorty, Rescher, etc. There is a large element of truth in Schiller's observation, that there are as many pragmatisms as pragmatists.

With respect specifically to the theory of knowledge, however, it is possible, not indeed to identify a set of shared doctrines, but instead to discern two broad styles of pragmatism, sharing the conviction that a Cartesian approach (see CARTESIANISM) is fundamentally flawed, but responding very differently to that failure. Repudiating the requirement of absolute certainty in knowledge, insisting on the connection of knowledge with action, pragmatism of a reformist stripe still acknowledges the legitimacy of traditional questions about the truth-conduciveness of our cognitive practices, and sustains a conception of truth objective enough to give those questions bite. Pragmatism of a revolutionary stripe, by contrast, relinquishing the objectivity of truth, acknowledges no legitimate epistemological questions over and above those internal to our current cognitive conventions.

This distinction between reformists and revolutionaries (adapted from one in Migotti, 1988) will not map perfectly on to my list of pragmatists, neo-pragmatists and sympathizers. Peirce would count as a reformist, as would Lewis, Ramsey, Sellars and Rescher; Schiller would count as a revolutionary, as would Rorty. But in James, to a degree, and more markedly in Dewey – arguably also in Quine and Putnam – elements of both tendencies are to be found. But this does not detract from the usefulness of the distinction as a diagnostic tool, either historically or philosophically; it will provide both a framework for understanding something of the shifts within pragmatism from Peirce through James and Dewey to Schiller, and the backdrop to an assessment of what is of most permanent value in the pragmatists' contributions to the theory of knowledge.

At the heart of Peirce's pragmatism (see PIERCE) lies the pragmatic maxim: "if one can define accurately all the conceivable experimental phenomena which the affirmation or denial of a concept could imply, one will have therein a complete definition of the concept,

and *there is absolutely nothing more in it*" (ibid., 5.412). Meaning is a matter of the conceivable experimental, experiential consequences of a concept's applying: *pragmatisch*, in the Kantian sense. Pragmatism is thus, Peirce acknowledges, a kind of "prope-Positivism"; one role of the pragmatic maxim is to reveal that "almost every proposition of ontological metaphysics . . . is gibberish" (5.423). But the maxim is not intended to rule out metaphysics altogether, but rather to discriminate the illegitimate, the pragmatically meaningless, from *scientific* metaphysics, which uses the method of the sciences, observation and reasoning. Philosophy, where it is legitimate, is an observational science, differing from other sciences not in its method, but in its reliance on the most familiar kinds of experience, the least sophisticated kinds of observation.

Peirce's thorough-going critique of Cartesian epistemology, and his own engagingly naturalistic theory of inquiry, are of a piece with this conception of a reformed, scientific philosophy.

With Descartes, philosophy "put off childish things, and began to be a conceited young man" (4.71). Descartes' method (see DESCARTES) is a sham, a matter of feigned doubt which inevitably leads to the eventual reinstatement of all the beliefs supposedly doubted. There is no such faculty as the intuition on which Descartes' criterion of clearness and distinctness relies, and no such intuitive self-consciousness as Descartes' reliance on the *cogito* (see COGITO) as the indubitable starting point for the reconstruction of knowledge requires. Descartes' aspiration to certainty is misplaced, his subjective stance viciously individualistic.

Unlike Descartes, and in strikingly Darwinian spirit, Peirce sees human belief as continuous with animal expectation, human inquiry as continuous with animals' exploration of their environment. He agrees that inquiry begins with doubt, but insists that it must be *real* doubt. Peirce conceives of belief as a habit of action, a disposition to behave, and of doubt as the result of some belief-habit's being interrupted by recalcitrance on the part of experience. Real doubt is thus

involuntary, and unpleasant; inquiry is a homeostatic process by which the organism strives to return to equilibrium, a new habit, a revised belief. Since what prompts inquiry is the urge to eliminate the irritation of doubt, the end of inquiry is stable, permanently doubt-proof belief.

Among possible methods for the "fixation of belief" the scientific method, Peirce holds, is distinguished by its appropriateness to the end of inquiry. Unlike, e.g., the *a priori* method favoured by traditional metaphysics, the scientific method, if it were sufficiently persisted in, would enable inquiry eventually to come to rest with beliefs which are stable because permanently safe from recalcitrance. Scientific method, according to Peirce, accommodates the perceptual judgements forced upon one by brute experience in an explanatory framework arrived at by reasoning of three types: abduction, the postulation of a hypothesis to explain some puzzling phenomenon; deduction of consequences from such abductive hypotheses; and inductive testing of those hypotheses (see ABDUCTION; INTUITION AND DEDUCTION; INDUCTION). Though mathematical truths are necessary, even mathematical inquiry, according to Peirce, is in a sense experiential – though in mathematics what is relevant is not outer but inner experience, the construction and observation of imagined diagrams (see MATHEMATICAL KNOWLEDGE). Peirce's disinclination to accept the Cartesian requirement that any knowledge worthy of the name must be certain or indubitable moves into sharper focus in the context of his insistence that the scientific inquirer is distinguished by his "contrite fallibilism", his readiness to "dump the whole cartload of his beliefs, the moment experience is against them" (1.14, 1.55). And his distaste for the individualism of Cartesian epistemology moves into sharper focus in the context of his conception of the individual scientific inquirer as just one contributor to a vast co-operative enterprise, extending both within and across generations.

Fallible and imperfect as scientific inquiry is, however, if this vast co-operative enterprise were to continue long enough – Peirce is aware there is no guarantee that it will

– eventually a final, stable opinion would be reached. The idea that evolutionary adaptation has given human beings an instinct for guessing right which enables them to make successful abductions, and the thesis that induction is in a sense self-corrective, play an obvious role here. Less obvious, but no less important, is the role of the thesis Peirce calls “scholastic realism”. This is the thesis that there are natural kinds and laws, “generals” as Peirce calls them, which are real, i.e. independent of how we think about or characterize the world. There is a pattern of generals, natural kinds and laws, underlying the particular facts and events we observe, which is “independent of what you, or I, or any number of men think”. And so the “arbitrary, accidental element” (8.13) in inquiry introduced by the peculiar circumstances and idiosyncrasies of individual inquirers can be expected gradually to be discarded as co-operative scientific inquiry proceeds, and the real pattern eventually to emerge.

Peirce contrasts his scholastic realism with “nominalism”, by which he means the idea that generals are “figments”, i.e. dependent on how we think about or describe things. But he is equally opposed to what one might call “noumenism”, the idea that the really real is in principle inaccessible to human cognition. This is a characteristically pragmatist attitude, for the pragmatic maxim disqualifies as bogus any question which would resist settlement however long scientific inquiry were to continue.

Peirce does not think it false to say that truth is correspondence with reality, but shallow – at best a nominal definition, giving no insight into the pragmatic meaning of the concept. His pragmatic definition identifies the truth with the hypothetical final upshot of scientific inquiry, and the real with the object of that opinion. “Truth is that concordance of [a] . . . statement with the ideal limit towards which endless investigation would tend to bring scientific beliefs”; “. . . any truth more perfect than this destined conclusion, any reality more absolute than what is thought in it, is a fiction of metaphysics” (5.564, 8.13). Peirce’s is a pragmatic realism, between nominalism and noumenism, neither idealist

(see IDEALISM) nor transcendentalist; his definition of truth aims at a delicate compromise between the twin desiderata of objectivity and accessibility.

“There can be no difference anywhere that doesn’t *make* a difference elsewhere” (James, 1907, p. 30): his version of the pragmatic maxim lies at the heart of James’s pragmatism too. Unlike Peirce, however, James thought philosophy would do well to go round Kant, rather than through him; and his interpretation of the maxim, stressing *praxis*, the practical, not simply the experiential, consequences of a concept’s applying, reflects this. Related, no doubt, is his willingness to include emotions among experiential consequences, and to include the practical consequences of *a subject’s believing a proposition*, as well as the practical consequences of *the belief’s being true* – which connect, in turn, with his doctrine of the “will to believe”, the thesis that religious beliefs which cannot by their nature be verified or falsified may nevertheless be legitimated by their salutary effects on the believer’s conduct of life. Another difference is less often stressed but no less significant. James is, in Peirce’s sense, a nominalist; when he writes of a belief’s practical effects he is concerned with its *particular* practical effects; and he takes it that all classifications are human constructs to be judged rather by their convenience and utility than by their coincidence with real kinds.

Like Peirce, James thinks it not false but inadequate to say that truth is correspondence with reality. Again much like Peirce, he characterizes “truth absolute” as “an ideal set of formulations towards which all opinions may in the long run of experience be expected to converge” (1909, p. 147). The difference between true beliefs and false ones is that true beliefs are “verifiable”, i.e. they would be confirmed by experience. The true, James sometimes says, is the satisfactory, the useful in the way of belief; true beliefs *work*. Critics like Moore and Russell were scandalized by what they took to be a crass identification of truth and utility: a criticism James describes as a “slander” (1909, p. 147), for, though indeed he had written that “*the true . . . is only the expedient in the way of belief*”, he

had gone on to explain, "... expedient in the long run and on the whole of course; for what meets expediently all the experience in sight won't necessarily meet all further experience equally satisfactorily. Experience ... has ways of *boiling over*, and making us correct our present formulas" (1907, p. 106).

But unlike Peirce James cannot appeal to the real constitution of the world to explain why, in the long run of experience, opinions can be expected to converge; and he is preoccupied more with truth in the concrete than with truth in the abstract, manifesting discomfort with the notion of *verifiability* and preferring to talk of particular truths actually verified. These nominalist predilections lead to apparent inconsistencies, as James seems sometimes to allow that beliefs which are verified are thereby shown to have been true all along, sometimes to suggest that beliefs become true when they are verified. Consistency may be restored by means of a distinction James sometimes makes between "abstract" or "absolute", and "concrete" or "relative" truth, identifying abstract truth with the verifiable and concrete truth with the verified. But what James calls concrete truth isn't really *truth* at all; his distinction is really between a belief's being true and its being shown to be true – or more accurately, since James allows that what is at one time "verified" may later turn out to be false, between a belief's being true and its being confirmed. James's stress on the particular, the concrete, amounts in effect to a tendency to downplay the concept of truth, as such, and to emphasize what is confirmed as inquiry proceeds.

In Dewey (*see* DEWEY), though he describes one of Peirce's best-known characterizations – "[t]he opinion which is fated to be agreed by all who investigate" – as "the best definition of truth" (Peirce, 5.414; Dewey, 1938, p. 345n.), this shift of emphasis is even more marked. Dewey, like James, is drawn to the idea of the mutability of truth; though he admits that what is verified is thereby shown to have been true all along, he interprets this as meaning only that it was going to be verified. Not surprisingly (given that concrete, mutable "truth" isn't really truth at all) he is also

drawn to the idea that it might be as well to stop talking of truth, and to work, instead, with the concept of warranted assertibility.

The attraction, for Dewey, of the doctrine of the mutability of truth no doubt relates to its consonance with his theory of inquiry, which shows some markedly revolutionary tendencies. Like Peirce and James, Dewey repudiates the "quest for certainty". Unlike them, he goes on to suggest a psychosociological diagnosis of the desire for certainty: it arose, he conjectures, from the sharp dichotomy of theory versus practice, and the distaste for the practical, the changeable, the uncertain, embodied in the slaveowning culture of ancient Greece. The most Hegelian of the pragmatists, Dewey is suspicious of traditional philosophical dualisms; and this is reflected in his epistemological writings, which are critical of the whole tradition from Plato through Descartes to his own contemporaries, because of its dependence on the dichotomies of subject/object, fact/value, theory/practice. "Special theories of knowledge differ enormously from one another. Their quarrels ... fill the air. The din thus created makes us deaf to the way in which they all say one thing in common. ... They all hold that the operation of inquiry excludes any element of practical activity that enters into the construction of the object known" (Dewey, 1929, p. 22). Dewey, by contrast, insists that knowing is not isolated from, but is itself a kind of, practice – to be judged, like other practices, by its purposive success rather than by some supposed standards of accuracy of its reflection of its objects; for the object of knowledge is not an immutable, independent reality, but is changed and even in part constituted by our cognitive interactions with it. Inquiry transforms an indeterminate situation into a determinate one.

Peirce's theory of inquiry is behaviouristic, naturalistic, fallibilistic; thoroughly anti-Cartesian, but essentially reformist. James's is predominantly reformist too, though his nominalistic preference for concrete "truths" introduces a potentially revolutionary element. In Dewey's theory revolutionary themes, especially the denial of an independent reality to which our theories may or may not



conform, are unmistakable. But reformist elements are also discernible. Laws are “intellectual instrumentalities”, he says, and there is no question of their literally conforming to what exists; but this doesn’t mean that they are merely mental, or that they need not “take account” of what exists (1929, pp. 205, 207). Important reformist themes in Dewey include: a distinction between the state and the content senses of belief; a conception of experience (*see* EXPERIENCE) much richer, thicker, than the old “sensationalist” idea; the aspiration to transcend the old dichotomy of rationalism (*see* RATIONALISM) versus empiricism (*see* EMPIRICISM) and allow a more realistic interplay of experience with reason.

Schiller is uncompromisingly revolutionary. While in both James and Dewey one finds a shift between the identification of truth with verifiability and the identification of truth with verification, and a more (James) or less (Dewey) equivocal commitment to the mutability of truth, in Schiller one finds a straightforward identification of truth with verification, and an unequivocal commitment to the mutability of truth. James acknowledges that his account of concrete truths could not stand alone: “to admit, as we pragmatists do, that we are liable to correction . . . involves the use on our part of an ideal standard” (1909, p. 142). But Schiller offers a theory of concrete truths as *a complete theory of truth*. He denies outright the idea that truth is correspondence with reality, an idea which he describes as not only worthless as a criterion but absurd in itself. Truth is practical working. “True”, he says, means “valued by us”; a proposition is true if it “forwards our ends”. Truth is mutable, since propositions become true only when successfully applied; a “truth which will not . . . submit to verification, is not yet a truth at all” (Schiller, 1907, p. 8). Reality is also mutable, growing as truth grows. Truth is dependent on us, relative to our purposes. And so is reality; facts are not simply discovered, but selected, even made, by us. In a spirit, no doubt, of deliberate provocation, Schiller likens his views to those of Protagoras; and, indeed, his revolutionary, relativistic humanism seems very far from Peirce’s realistic pragmatism. As

Peirce himself was well aware: “It seems to me a pity”, he wrote in 1908, “that [Mr Schiller and the pragmatists of today] should allow a philosophy so instinct with life to become infected with the seeds of death in such notions as . . . the mutability of truth” (Peirce, 6.485).

A poignant, and prophetic, observation: Rorty (*see* RORTY), most radical of contemporary self-styled pragmatists, and the closest to Schiller, uses “pragmatism” as the contrast to “realism”; takes the pragmatist view to be that truth is not the kind of thing there can be an interesting theory about; holds that there is nothing to be said about the criteria of rational belief, about what counts as good or flimsy evidence, over and above the conventions of our cognitive practices; conceives of pragmatism, in short, as a revolutionary repudiation of the questions, as well as the answers, of the epistemological tradition.

Other contemporaries, Bernstein and Margolis for example, have more or less self-consciously sought habitable middle ground between the revolutionary and the reformist wings of the pragmatist tradition; an enterprise which seems, however, rather to trade on than to resolve the relevant ambiguities.

It is rather, to my mind, within the tradition of reformist pragmatism that the most enduringly interesting epistemological contributions are to be found: for example, Mead’s theory of the social construction of the self, inspired by Peirce’s critique of the intuitive self-consciousness assumed by Descartes; Lewis’s – somewhat nominalistic – “*pragmatic a priori*” (*see* LEWIS), itself an inspiration for Quine’s call (*see* QUINE) for “a more thorough-going pragmatism” (Quine, 1953, p. 46); Ramsey’s behaviouristic approach to belief, and Quine’s, also, to meaning (Quine cites Dewey: “meaning is . . . primarily a property of behavior” (Quine, 1969, p. 27; Dewey, 1925, p. 179)); Quine’s association of natural kinds, induction and evolutionary epistemology (*see* EVOLUTIONARY EPISTEMOLOGY); Reichenbach’s pragmatic vindication of induction (*see* REICHENBACH); Hanson’s defence of the idea of an abductive logic of scientific discovery, Sellars’s appeal to the notion of explanatory coherence (*see*

SELLARS), Harman's to inference to the best explanation (*see* INFERENCE TO THE BEST EXPLANATION); Putnam's explorations of conceptions of truth intermediate between metaphysical realism and relativism (*see* PUTNAM), Apel's of consensual theories and their relation to the social dimensions of inquiry; Rescher's investigations of criteria of success and improvement of cognitive methods (*see* RESCHER), Jardine's of scientific progress; and many more. In my own work, such key ideas as explanatory integration, central to the articulated quasi-holism of my account of evidential support, and the distinction of belief-states and belief-contents, central to my account of the role of experience, were mined, as I can testify, from the same vein.

Unifying this rich but, it must be admitted, formidably diverse profusion of philosophical ideas is what one might call the ongoing project of reformist pragmatism: the aspiration to find a middle ground between dogmatism and scepticism (*see* SCEPTICISM); a conception of truth accessible enough to be realistically aspired to, yet objective enough to be worthy of the name; an articulation of the interplay between the world's contribution to knowledge, and ours. This is the essential spirit of reformist pragmatism, succinctly summed up by James: "... please observe ... that when ... we give up the doctrine of objective certitude, we do not thereby give up the quest or hope of truth itself" (1897, p. 17). So conceived, the tradition of reformist pragmatism still flourishes; and, though very far as yet from the "catholic consent" Peirce saw (8.13) as the end of inquiry, it is, indeed, "instinct with life".

*See also* EMPIRICISM; EVOLUTIONARY EPISTEMOLOGY; EXPERIENCE; GENETIC EPISTEMOLOGY; NATURALIZED EPISTEMOLOGY; NIHILISM; REALISM.

#### PRAGMATIST WRITINGS

- Dewey, J.: *Logic, the Theory of Inquiry* (New York: Henry Holt, 1938).  
 Dewey, J.: *Experience and Nature* (1925) (LaSalle, IL: Open Court, 1958).  
 Dewey, J.: *The Quest for Certainty* (1929) (New York: Capricorn Books, 1960).

- James, W.: *The Will to Believe and Other Essays in Popular Philosophy* (1897) (New York: Dover, 1956).  
 James, W.: *Pragmatism* (1907), ed. F. Burkhardt and F. Bowers (Cambridge, MA: Harvard University Press, 1975).  
 James, W.: *The Meaning of Truth* (1909), ed. F. Burkhardt and F. Bowers (Cambridge, MA: Harvard University Press, 1975).  
 Peirce, C. S.: *Collected Papers*, ed. C. Hartshorne, P. Weiss and A. Burks (Cambridge, MA: Harvard University Press, 1931–58). References to Peirce are by volume and paragraph number in this edition.  
 Schiller, F. C. S.: *Studies in Humanism* (London: Macmillan, 1907).  
 Thayer, H. S., ed.: *Pragmatism: The Classic Writings* (New York: New American Library, 1970).

#### BIBLIOGRAPHY

- Dewey, J.: "The Development of American Pragmatism," in his *Philosophy and Civilization* (New York: Minton, Balch, 1931).  
 Haack, S.: "'Extreme scholastic realism': Its Relevance to Philosophy of Science Today," *Transactions of the Charles S. Peirce Society* XXVIII: I (1992), 19–50.  
 Migotti, M.: "Recent Work in Pragmatism: Revolution or Reform in the Theory of Knowledge?," *Philosophical Books* 29, (1988), 65–73.  
 Quine, W. V.: "Two Dogmas of Empiricism," in his *From a Logical Point of View* (Cambridge, MA: Harper Torchbooks, 1953), 20–46.  
 Quine, W. V.: *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969).  
 Rorty, R.: *The Consequences of Pragmatism* (Hassocks: Harvester, 1982).

SUSAN HACK

**preface paradox** The paradox of the preface poses a problem that concerns the nature of rational belief. The paradox was introduced by D. C. Makinson (1965). The following example illustrates how the problem arises.

In his latest work an author asserts some of his rationally held beliefs: b<sub>1</sub>, b<sub>2</sub> ... b<sub>n</sub>. Since the author realizes that his previous works have all turned out to contain false assertions in spite of his best efforts, he

modestly and reasonably acknowledges in the preface of his new book that it too contains some false assertion somewhere. This acknowledgement serves to convey the author's belief in a proposition which has the logical form:  $\neg(b_1 \& b_2 \& \dots \& b_n)$ , a proposition that we can call "the prefatory belief".

It is logically impossible for all of the beliefs that have just been attributed to the author to be true. We should suppose that the author knows this. None the less, there seems to be no good reason to deny that the beliefs could all be rational under such circumstances. Is this situation really possible? If so, what is the relevant difference between such a case and cases of known logical conflict among beliefs that exemplify irrational belief? If the situation is not really possible, why not? Answering these questions constitutes the problem.

This is a serious problem. There is no easy way to show that the situation is either impossible or unproblematic. It is instructive to observe that a common approach to solving the lottery paradox (*see* LOTTERY PARADOX) cannot work here. The lottery paradox involves certain predictive propositions for each of which someone has a high level of purely statistical support. These propositions seem to be rationally believed by the person on the basis of that evidence, even though the person knows that there is a falsehood somewhere among them. A common approach to solving the lottery paradox rests on denying that purely statistical evidence is ever sufficient for the rational justification of a belief. However well that approach works for the lottery problem, it does not solve the general problem posed by the paradox of the preface. Statistical evidence is not essential to the latter. The author's beliefs that are expressed in the body of his work can be made rational by any sort of evidence that can be rationally believed to allow mistakes. This includes diverse familiar forms of evidence, e.g. perception, memory, testimony, formal reasoning and introspection, as well as statistical evidence. In order to follow closely Makinson's original exposition of the paradox, the case described above provides support for the prefatory belief from essentially statis-

tical considerations. But in equally plausible examples the prefatory belief is supported in other ways. For instance, the author may be given reliable testimonial evidence to the effect that his new manuscript contains some false assertion, or the author may seem to recall having previously discovered some error in the work without being able any longer to recall what it is, or the work may be a formal effort in which  $b_1, b_2 \dots b_n$  purport to be proven, while the prefatory belief is based on the author's last minute discovery of a proof that  $b_1, b_2, \dots, b_n$  together imply a contradiction. Thus, the inconsistent beliefs can be supported by virtually any sort of evidence. It is noteworthy that the above sorts of evidence for the prefatory belief give the author no indication of any particular flaw in his rational support for each of  $b_1, b_2, \dots, b_n$ . The general problem posed by the preface paradox cannot be solved on the basis of a denial that statistical evidence suffices for the rational justification of belief.

In light of the diversity of plausible cases of the phenomenon, it may seem most reasonable to accept that inconsistent rational beliefs are possible. Makinson reports in opposition to this view that we feel compelled to hold on general grounds that inconsistent rational beliefs are not possible (1965, p. 206). Makinson does not identify any of these general grounds, but John Pollock suggests one. Pollock contends that denying a need for consistency among rational beliefs would reduce to insignificance the role of deductive inference in rational belief acquisition (1986, p. 249). One problem with inconsistent rational beliefs that involves deduction is brought out by the following argument.

The argument begins with the assumption that every proposition is a deductive consequence of logically inconsistent premises. The second assumption is that valid deductive inference from rationally held beliefs is a sufficient basis for rational belief in the deduced conclusion. Given the hypothesis that there are logically inconsistent rational beliefs, these assumptions imply that any proposition whatever is rationally accepted as long as it is accepted on the basis of its deduction from inconsistent rational beliefs. Yet it

is clear that the author in our example does not gain rational belief in wild, arbitrary propositions simply by accepting them on the basis of deductions from  $b_1, b_2, \dots, b_n$  together with the prefatory belief. The argument concludes that the correct way to avoid this result is to deny that inconsistency among rational beliefs is possible.

This argument helps to draw attention to a problem involving deduction, but it does not show that inconsistent rational beliefs are impossible. The argument employs a premiss asserting that it is always rational to believe a proposition on the basis of a valid deduction from rational beliefs. This premiss can be refuted on entirely independent grounds. In certain cases, a *modus ponens* argument's premisses are not merely rationally believed but known, and yet it is not rational to accept the conclusion of the argument by deducing it from the premisses (Conee, 1987). The rationality of believing a proposition sometimes depends on what the believing is foreseen to bring about. In some cases not involving inconsistent beliefs, believing a given proposition is foreseen to bring with it the falsehood of that proposition. (One way this can happen is for the proposition in effect to say of itself that it is not believed.) When this is foreseen and nothing else is at stake, it is not rational to believe the proposition. This is true even if the proposition is known to be a logical implication of rationally believed or even known premisses. Hence, even propositions known to follow from rational beliefs are not always rational to believe, whether or not there can be inconsistent rational beliefs. The argument under consideration against the possibility of inconsistent rational beliefs thus fails.

A difficult task remains, however. Deduction from rational beliefs does very often yield new rational beliefs, and the new beliefs seem to be rational primarily because they are so inferred. Mathematics provides numerous illustrations of this. More poignant examples are provided by the cases that constitute the preface paradox. Virtually any valid deduction that the author could make from  $b_1, b_2, \dots, b_n$ , or from the prefatory belief, would afford an adequate basis for rational belief in the

inferred proposition. Any tenable theory of rational belief must allow deduction from rational beliefs to play some important role in rendering belief rational. Those who accept the possibility of inconsistent rational beliefs face the delicate task of reconciling this important role of deduction with the manifest failure of deduction to yield rational belief in many (though not all) of the deductive consequences of the inconsistent beliefs which they count as rational. Those who reject the possibility of inconsistent rational beliefs face the formidable challenge of explaining why the myriad apparent cases all turn out to be spurious or impossible.

*See also* PARADOX.

# BIBLIOGRAPHY

- Conee, E. B.: "Evident, but Rationally Unacceptable," *Australasian Journal of Philosophy* 65 (1987), 316–26.
- Hoffman, R.: "Mr Makinson's Paradox," *Mind* 77 (1968), 122–3.
- Hoffman, R.: "The Paradox of the Preface Again," *Mind* 82 (1973), 441.
- Klein, P.: "The Virtues of Inconsistency," *The Monist* 68 (1985), 105–35.
- Lacey, A. R.: "The Paradox of the Preface," *Mind* 79 (1970), 614–15.
- Makinson, D. C.: "The Paradox of the Preface," *Analysis* 25 (1965), 205–7.
- Pollock, J. L.: "Epistemology and probability," *Synthese* 55 (1983), 231–52.
- Pollock, J. L.: "The Paradox of the Preface," *Philosophy of Science* 53 (1986), 246–58.

EARL CONEE

**presence** *see* IMMEDIACY.

**presocratic epistemology** The beginning of philosophical interest in knowledge has been dated by some to the critical challenge to knowledge raised by fifth-century Sophists (Hamlyn, 1961), but the inability of human intelligence to comprehend the larger scheme of events was a commonplace of the earliest Greek poetry (cf. *Iliad* II, 484ff; *Odyssey* XVIII, 137ff.; Archilochus Fr. 70;

etc.). Snell (1953) traced the poet's pessimism to an identification of knowledge with eye-witness experience (cf. the Greek *oida*, "I know", lit. "I have seen"): what can be known is only what can be seen or otherwise perceived within the narrow limits of a human lifetime. Presocratic reflection on knowledge begins from this assumption of the joint inadequacy and indispensability of sense experience for knowledge.

Xenophanes of Colophon (c.565–c.470 bc) claimed that no man ever saw "the clear and certain truth" (*to saphes*), opinion (*dokos*) being "allotted to all" (Fr. 34.1, 4). The character of his scepticism and its supporting rationale have been the subject of debate since antiquity. Since the topic mentioned at 34.2 is "the gods and such as I say about all things" (i.e. divine attributes and the basic principles of the physical cosmos that Xenophanes has so far identified; cf. his comments on "all things" in Fr. 27 and 29), Fr. 34 falls well short of the universal scepticism characteristic of a later period. While Fränkel's thesis (1974) that Xenophanes' word for "know" just meant "know by means of sight" has not gained acceptance (see Barnes, 1979), it is clear that Xenophanes' nearcontemporary Herodotus held that knowledge, i.e. *clear and certain* awareness of truth, required confirmation on the basis of first-hand observation (cf. *History* II, 44). Xenophanes probably assumed, then, that our knowledge could extend no further than our direct experience, and drew the appropriate conclusion for the (inaccessible) realm of the divine and the (universal) principles of nature he had identified.

Whether Xenophanes countenanced *a priori* knowledge (e.g. about the divine, as proposed by Barnes and Hussey) is less clear. Fr. 34 would lead one to expect only "opinion" in this area; his explicit remarks on the "one greatest god" (Fr. 23–6) are all flat assertion; and the complex set of logical deductions credited to him in the pseudo-Aristotelian treatise *de Melisso Gorgia Xenophane* are almost certainly constructions of a later period. Xenophanes encouraged inquiry into the wonders of nature (Fr. 18, 27–33) and reflection about general principles

based on knowledge of new-found facts (cf. A 33 – his theory of alternating periods of flood and drought based on the discovery far inland of remains of ancient sea creatures), even if clear and certain knowledge of basic principles lay beyond mortal capacities.

Heraclitus of Ephesus (early fifth century) also spoke of inquiry as essential to knowledge (Fr. 35, 55), but listed Xenophanes among those who proved that "much-learning does not teach *noos*" (Fr. 40, roughly translated: "one can fail to see the forest for the trees"). It has been argued (Hamlyn, 1961; Kahn, 1979) that Heraclitus was less interested in knowledge than in the unity of the cosmos, but Fr. 1, 17, 34, 45, 72 appear designed to provoke reflection on the nature of knowledge by simultaneously identifying and contrasting sense experience with genuine understanding (Leshner, 1983). Other fragments highlight the roles played by the concepts one, many, same, different, motion, and rest (Mackenzie, 1988) as well as the importance of thinking, reflection, and interpretation (Fr. 19, 93, 107, 113, 116).

Pythagoras of Samos (late sixth century bc) was famous for claiming knowledge (Xenophanes Fr. 7; Heraclitus Fr. 129) rather than for explaining it, but two fifth-century(?) followers, Alcmaeon and Philolaus, discussed human cognitive capacities in relation to those of other animals (Fr. 1a and 13 respectively) and the divine (Fr. 1 and 6). The empiricist account of knowledge stated in Plato's *Phaedo* (96b) may go back to Alcmaeon (Barnes, 1979), but it sits uncomfortably with Alcmaeon's view of human awareness as mere "conjecture from signs" (Fr. 1). Philolaus' thesis that number was a condition for thought and knowledge (Fr. 3, 4, 6, 11) has been challenged as too explicit an epistemological opinion for so early a period (Kirk and Raven, 1957), but a connection between knowledge, measures and the limits of things appeared in Fr. 16 of Solon (c.640–c.561 bc) and ran through Heraclitus' Fr. 30, 31b, 45, 94 and 120. Recent accounts credit Philolaus with an epistemology along Pythagorean (mathematical) lines developed either in response to attacks on pluralism launched by followers of Parmenides (Huffman, 1988) or



by extension from Parmenides' own principles (Hussey, 1990).

Parmenides of Elea (late sixth–early fifth century) set out an account of “the existent” or “what is” (*to eon*) and of what can be thought and known about it, evidently in order to correct the confused way of thinking displayed in earlier cosmologies. After presenting a serial “critique” (*elenchos*) of the possible ways of thinking (“it is”, “it is not”, “it is and it is not”, and “it is or it is not” – arguably the world’s first Boolean analysis), he concluded only “it is” can be said and thought, a verdict based partly on the claim that all thought, knowledge, and meaningful speech about “the non-existent” are impossible (for interpretations of this thesis see Furth, 1968; Mourelatos, 1970; Owen, 1960).

The goddess of Parmenides’ poem promised ‘the unshaking heart of well rounded – or well persuasive – truth’ (Fr. 1.29; cf. “trustworthy account” at Fr. 8.50). Her preference for “tied down”, fixed, or invariably true ways of speaking (“that the existent exists”, “either it exists or it does not exist”, “what does not exist cannot possibly exist”, etc.) can be restated in more contemporary terms as a restriction of the class of knowable propositions to necessarily true ones. Parmenides’ serial critique of mortal opinion and repudiation of sense perception in pursuit of “well-persuasive” truth has invited comparison with Plato’s preference for a priori knowledge (at *Phaedo*, 100dff and *Rep.* 511bff) and with Descartes’ meditative review of his opinions in pursuit of those (clear and distinct) ideas that have the power to convince him completely (Owen, Kirk et al.).

The impact of Parmenides’ teaching is evident from Empedocles Fr. 11, Melissus Fr. 2, Zeno Frr. 1–4, and Anaxagoras Fr. 17 (for Leucippus and Democritus, cf. Aristotle, *de gen. et corr.*, 325a2ff). His successors did, however, offer materialistic explanations of thought of just the sort he appears to have repudiated in the *doxa* section of his poem (cf. Alcmaeon A 5; Empedocles Frr. 2, 109; Anaxagoras A 92; Democritus A 135; for this reading of the *doxa*, see Long, 1963; for a contrasting view of the Parmenidean character of the *doxa* section, see Coxon, 1986).

The essential but inadequate contribution to knowledge made by sense perception also led Anaxagoras and Democritus to flirt with scepticism (Frr. 21 and 7, 9, 11, 125 respectively). Democritus tied his scepticism to a distinction between the real properties of things and the qualities (e.g. sweetness) perceived “by custom” (*nomôi*, Fr. 9 – a distinction anticipated by Xenophanes’ comment (Fr. 38) that men would think figs were far sweeter if they had never tasted honey). Empedocles held that ordinary men lacked “knowledge of the whole” (Fr. 2) but (perhaps invoking the idea of multiple paths to understanding – Fr. 3) assigned a vast knowledge to himself (Frr. 17, 23).

Neither Plato nor Aristotle gave entirely accurate summaries of presocratic ideas about knowledge, but both spoke of the physical causes of sensation and the relation of sense experience to knowledge as questions raised by their predecessors (cf. Plato, *Phaedo*, 96b; *Theaetetus*, 152ff; Aristotle, *de anima* I, 2; III, 4; *Meta.* I, 5). Later sceptics appropriately traced their philosophical pedigree back to views expressed by the presocratics (cf. Cicero, *Academica* I.43ff: “Democritus, Anaxagoras, Empedocles, and almost all the ancients who said that nothing could be grasped or cognized or known, saying that the senses are restricted, the mind weak, the course of life short and that . . . truth has been submerged in an abyss . . .”; cf. also D. Laertius, 9.66; Timon Frr. 59, 60, 818–20, etc.).

See also ARISTOTLE; PLATO.

#### BIBLIOGRAPHY

- Barnes, J.: *The Presocratics* (London: Routledge and Kegan Paul, 1979).
- Coxon, A.: *The Fragments of Parmenides*, vol. III in the *Phronesis Supplemental Series* (Assen/Maastricht, 1986).
- Diels, K. and Kranz, W.: *Die Fragmente der Vorsokratiker*, 6th edn (Berlin: Weidmann, 1951); references to fragments and *testimonia* are to the B-fragment and A-*testimonia* sections.
- Fränkel, H.: “Xenophanes’ Empiricism and His Critique of Knowledge,” in *The Pre-Socratics*, ed.

- A. P. D. Mourelatos (Garden City, NY: Anchor Press/Doubleday, 1974), 118–31.
- Furth, M.: "Elements of Eleatic Ontology," *Journal of the History of Philosophy* 6: 3 (1968), 1–32.
- Hamlyn, D. W.: *Sensation and Perception* (London: Routledge and Kegan Paul, 1961).
- Huffman, C.: "The Role of Number in Philolaus' Philosophy," *Phronesis* 33 (1988), 1–29.
- Hussey, E.: "The Beginnings of Epistemology: From Homer to Philolaus," in *Epistemology*, ed. S. Everson (Cambridge: Cambridge University Press, 1990), 11–38.
- Kahn, C.: *The Art and Thought of Heraclitus* (Cambridge: Cambridge University Press, 1979).
- Kirk, G. S., Raven, J. E. and Schofield, M.: *The Presocratic Philosophers* (Cambridge: Cambridge University Press, 1957; 2nd edn 1983).
- Leshner, J. H.: "Heraclitus' Epistemological Vocabulary," *Hermes* III (1983), 155–70.
- Long, A. A.: "The Principles of Parmenides' Cosmology," *Phronesis* 8 (1963), 90–107.
- Mackenzie, M. M.: "Heraclitus and the Art of Paradox," *Oxford Studies in Ancient Philosophy* 6 (1988), 1–37.
- Mourelatos, A. P. D.: *The Route of Parmenides* (New Haven, CT, and London: Yale University Press, 1970).
- Nussbaum, M.: "Eleatic Conventionalism and Philolaus on the Conditions of Thought," *Harvard Studies in Classical Philology* 83 (1979), 63–108.
- Owen, G. E. L.: "Eleatic Questions," *Classical Quarterly* 10 (1960), 84–102.
- Snell, B.: *The Discovery of the Mind*, trans. T. G. Rosenmeyer (Oxford: Oxford University Press, 1953).

J. H. LESHNER

**prima facie reasons** There are two notions of the *prima facie* used in epistemology. The first is familiar from the law. For someone to be committed to stand trial, a *prima facie* case has to be made against them. A case of this sort is one which is strong enough to need an answer. It is called a *prima facie* case because of the Latin meaning of the terms: a *prima facie* case is one which *at first sight* looks impressive enough to need an answer. Such a case may collapse completely under further scrutiny.

The second notion of the *prima facie* is a technical use that derives from the moral philosophy of W. D. Ross. He introduced a notion of a *prima facie* duty in the following way: we

have a *prima facie* duty to keep our promises if every action of promise-keeping is to that extent right – if all actions of promise-keeping are the better for it. An action may be a *prima facie* duty (in virtue of some property it has) in this sense even though it is wrong overall, and so not a "duty proper", in Ross's terms.

Those who speak of *prima facie* reasons may do so in either of the above senses, but they should be clear which they intend since the two senses are so different. The main difference is that reasons of the first sort may collapse completely under scrutiny, so that something that seemed to be a reason (was a *prima facie* reason) ceases to be so on further enquiry; those of the second sort always remain as reasons, though they may be overridden by stronger *prima facie* reasons on the other side.

## BIBLIOGRAPHY

- Dancy, J.: "Intuitionism in Meta-epistemology," *Philosophical Studies* 42 (1982), 395–408.
- Ross, W. D.: *The Right and the Good* (Oxford: Clarendon Press, 1930), ch. 2.

JONATHAN DANCY

**primary and secondary qualities** A metaphysical distinction was drawn in antiquity between qualities which really belong to objects in the world and qualities, which only appear to belong to them, or which human beings only believe to belong to them, because of the effects those objects produce in human beings, typically through the sense organs. Thus Democritus: "by convention (*nomoi*) colour, by convention the sweet, by convention bitter; but in truth (*etei*) atoms and the void". Colour, sweetness, bitterness are here said to exist only "by convention": as something that does not hold everywhere by nature, but is produced in or contributed by human beings in their interaction with a world which really contains only atoms of certain kinds in a void. To think that some objects in the world are coloured, or sweet, or bitter is to attribute to objects qualities which on this view they do not actually possess.

Objects must possess some qualities or other in order to produce their effects, so the view is not that there are no qualities at all in the objects which cause perceivers to impute certain qualities to them. Rather, it is only that some of the qualities which are imputed to objects (e.g. colour, sweetness, bitterness) are not possessed by those objects. Knowledge of nature is knowledge of what qualities objects actually have, and of how they bring about their effects. For Democritus, atoms really possess those qualities (e.g. shape, size, motion) which are responsible for their having all the effects they have. To claim such knowledge is to impute certain qualities to objects; the richer one's knowledge, the more such qualities will be imputed. But when the imputation is true, or amounts to knowledge, the qualities are not merely imputed; they are also in fact present in the objects. The metaphysical view holds that those are the only qualities which objects really have. The rest of our conception of the world has a human source.

Galileo drew a similar distinction in explaining the wide gap between the way the world normally appears to perceiving beings on earth and the truth revealed about it by the "new science". If the sense organs of animals were taken away, he said, the figure, the number, and the motions of bodies would remain, but all colours, odours, sounds, etc., would be "abolished and annihilated". For him, all such qualities, "without the living animal", are "nothing but names". Although we have words for such things, we do not succeed in speaking of anything that really belongs to objects in the world. Objects possess only those qualities referred to in a perfected mathematical science which would explain why everything in the world happens as it does.

This is so far not a distinction between two kinds of qualities ("primary" and "secondary") which objects possess, or between qualities which are imputed to objects and qualities which are not, but rather between qualities which objects really have and qualities which are merely imputed to them but which they do not in fact possess. It is a claim about what is really so.

Descartes (*see* DESCARTES) found nothing but confusion in the attempt even to impute

to objects those very effects which they bring about through the senses. The "sensations" caused in people's minds by the qualities of bodies which affect them could not themselves be in the external objects. Nor does it make sense to suppose that bodies could in some way "resemble" those sensory effects. For Descartes the essence of body is extension, so no quality that is not a mode of extension could possibly belong to body at all. Colours, odours, sounds, etc., are on his view nothing but sensations. "When we say that we perceive colours in objects, this is really just the same as saying that we perceive something in the objects whose nature we do not know, but which produces in us a certain very clear and vivid sensation which we call the sensation of colour." If we try to think of colours as something real outside our minds "there is no way of understanding what sort of things they are".

This again is not a distinction between two kinds of qualities which belong to bodies; it is a distinction between qualities which belong to bodies (all of which are modes of extension such as shape, position, motion, etc.) and what we unreflectively and confusedly come to think are qualities of bodies.

The term "secondary quality" appears to have been coined by Robert Boyle (1627–91) whose "corpuscular philosophy" was shared by Locke. But it is not easy to say what either he or Locke meant by the term. They were not consistent in their use of it. Locke (*see* LOCKE), like Boyle, distinguished an object's qualities from the powers it has to produce effects. It has such powers only in virtue of possessing some "primary" or "real" qualities. The effects it is capable of producing occur either in other bodies or in minds. If in minds, the effects are "ideas" (e.g. of colour or sweetness or bitterness, or of roundness or squareness or motion). These ideas in turn are employed in thoughts to the effect that the object in question is, e.g., coloured or sweet or bitter, or round or square or moving. We have such thoughts, according to Locke, by thinking that the object in question "resembles" the idea we have in the mind.

Boyle and Locke sometimes call colour, sweetness, bitterness, etc., "secondary" qualities. In the view of Democritus, Galileo and

Descartes, colour, sweetness, bitterness, etc. are only mistakenly or confusedly believed to belong to objects. That would imply that objects do not really have such "secondary" qualities. But Locke also identifies "secondary qualities" as "such qualities which in truth are nothing in the objects themselves but powers to produce various sensations in us by their primary qualities" (*Essay*, 2.8.9). This can be taken in at least two ways. It could mean that, in addition to its "primary" qualities, all there really is in an object we call coloured, sweet, or bitter, etc. is its power to produce ideas of colour, sweetness, or bitterness, etc. in us by virtue of the operation of those "primary" or "real" qualities. That is compatible with the earlier view that colour, sweetness, bitterness, etc. are not really in objects. Or it could (and does seem to) mean that "secondary qualities" such as colour, sweetness, bitterness, etc. are themselves nothing more than certain powers which objects have to affect us in certain ways. But such powers, on Locke's view, really do belong to objects endowed with the appropriate "primary" or "real" properties. To identify "secondary qualities" with such powers in this way would imply that such "secondary qualities" as colour, sweetness, bitterness, etc., since they are nothing but powers, really do belong to or exist in objects after all. Imputations of colour, sweetness, etc. to objects would then be true, not false or confused, as on those earlier views.

A distinction drawn in this way between "primary" and "secondary" qualities would not be a distinction between qualities which objects really possess and qualities which we only mistakenly or confusedly think they possess. Nor would it even be a distinction between two kinds of *qualities*, strictly speaking. Rather it would be a distinction between qualities and (certain kinds of) powers, both of which really belong to objects. But Locke confusingly sometimes calls both of them "qualities".

He also held that our ideas of "primary" qualities such as bulk, figure, motion, etc. "resemble" qualities in bodies, but our ideas of "secondary qualities" such as colour, sweetness, bitterness, etc. do not. In the latter case,

but not the former, "there is nothing like our ideas, existing in the bodies themselves" (*Essay*, 2.8.15). This is Locke's way of saying what really belongs to the objects around us: only what the "corpuscular philosophy" says about them is so. We only mistakenly impute "secondary" qualities to objects; but in the case of the "primary" qualities the imputations are true. But that is inconsistent with the idea that the "secondary" qualities which we impute are nothing but powers, since the imputations would then be imputations of certain powers, and so would be true of all objects with the appropriate "primary" or "real" qualities.

Berkeley (*see* BERKELEY) objected to Locke that it is nonsense to speak of a "resemblance" between an idea and an object, just as Descartes had ridiculed the idea that a sensation could resemble the object that causes it. "An idea can be like nothing but an idea," Berkeley says (*Principles* §8). This is a general rejection of Locke's account of how we are able to think of things existing independently of the mind. It is correct, it works as much against what Locke says of our ideas of "primary" qualities as it does against what he says of our ideas of "secondary" qualities.

Boyle speaks of the "texture" of a body whose minute corpuscles are arranged in a certain way. It is in virtue of possessing that "texture" that the body is "disposed" or has the power to produce ideas of certain kinds in perceivers, even if no one is perceiving it at the moment. For Locke, objects have the powers they have only because their minute parts are arranged in the ways they are (and the laws of nature are what they are). In each case there is acknowledged to be a categorical "base" of the power; the object *can* do such-and-such only because it is so-and-so, even if the relevant way it *is* happens to remain unknown to us. This has tempted some philosophers in recent years to identify "secondary" qualities, not with the powers which objects have to affect us in certain ways, but with the qualitative "bases" of those causal powers. The colour or sweetness or bitterness, etc. of an object would then be some real (but possibly unknown) quality of the object which is responsible for the specific effects it has on us.

This again would imply that “secondary” qualities, so understood, are really in objects. And it would have the consequence that “secondary” qualities are true qualities, not just powers. But it would seem to leave no room for a distinction between “secondary” and “primary” or “real” qualities of bodies. The “bases” of all the causal powers of objects are to be understood in terms of their “primary” or “real” qualities.

Defence of a distinction between those qualities which really belong to objects and qualities which are only mistakenly thought to belong to them faces the epistemic problem of how we can know, of any particular kind of quality, whether it belongs to the first group or the second. Scientific knowledge of nature purports to tell us what objects there are and what sorts of qualities they have. Democritus, Galileo, Descartes, Locke and most other philosophers who have invoked the distinction thought they possessed some such knowledge. They relied on it to identify the first group of qualities. But atomistic or corpuscular or nuclear or any other specific physical science at best says only what is so; it does not also say what qualities objects do *not* have. So the metaphysical theory must establish in addition the further claim that the kinds of qualities mentioned in the preferred science are the *only* kinds of qualities which objects have.

Some have apparently thought that the wide variability among humans’ perceptions of the colours, sweetness or bitterness, etc. of objects, as contrasted with the uniformity in their perceptions of their shape, size, or position, etc., is enough to show that the former qualities do not belong to objects, while the latter do. But such appeals to “the relativity of perception” alone are at best inconclusive. It is not clear that there is in fact greater variability among our perceptions of the one kind of qualities than there is among those of the other. But even if there were, the most it would show is that we cannot tell by a single perception alone that an object is coloured, or sweet, or bitter, etc., not that it has no such qualities at all. Berkeley argued in this way against “modern philosophers” who tried to prove on “relativity” grounds that certain

“sensible qualities” do not exist outside the mind.

Two main strategies remain for accounting for such “secondary” qualities as colour, sweetness, bitterness, etc. in a world containing only objects with nothing but the “primary” or “real” qualities mentioned in a comprehensive physical science. One, in the spirit of Democritus, Galileo and Descartes, is to grant that we do have perceptions of and beliefs about such qualities, and to argue that all of them can none the less be explained without having to assume that any object anywhere actually has any colour, sweetness or bitterness, etc. The explanations would proceed solely in terms of the “primary” or “real” qualities mentioned in the preferred comprehensive physical science. They would thereby expose the perceptions as illusory and the beliefs as false or confused. This raises large issues about the relation between the mental and physical, and about the possibility of explaining psychological phenomena in exclusively physical terms.

Another strategy is to show that the qualities said to be perceived or thought about in such cases are really qualities that do belong to objects after all. This can take the form of arguing that, e.g., the word “coloured” just means the same as “has the power to produce perceptions of colour in human beings”, or means the same as “has that quality which produces perceptions of colour in human beings”, or means the same as that physical term, whatever it is, which denotes that quality which in fact produces perceptions of colour in human beings. These are all theses about the *meanings* of terms for allegedly “secondary” qualities. Or it might be held only that a so-called ‘secondary’ quality term in fact denotes the very same quality as is denoted by some purely physical term. This would simply identify the very quality in question with some physical quality or power (not two different qualities, but only one), without holding that the terms that denote it must have the same meaning. In either case it would have the consequence that when we see colour, or believe that an object is coloured, what we see, or what we believe to



belong to the object, is that very physical quality or power which colour is said to be. This again would leave no distinction between qualities which really belong to objects and qualities which are only mistakenly or confusedly imputed to them.

See also MOLYNEUX'S PROBLEM; NATURALISM; NOUMENAL/PHENOMENAL; OBJECTIVE/SUBJECTIVE; OBJECTIVITY; SUBJECTIVITY.

#### BIBLIOGRAPHY

- Bennett, J.: *Locke, Berkeley, Hume: Central Themes* (Oxford: Oxford University Press, 1971).  
 Berkeley, G.: *Principles of Human Knowledge* (1710), in *Berkeley: Philosophical Writings*, ed. M. R. Ayers (London: Dent, 1975).  
 Boyle, R.: "Experiments and Observations upon Colours" and "The Origins and Forms of Qualities," in his *Works*, vols 1, 3 (London: Birch, 1772).  
 Descartes, R.: "Optics" and "Principles of Philosophy," in *The Philosophical Writings of Descartes*, vol. 1, ed. J. Cottingham, R. Stoothoff and D. Murdoch (Cambridge: Cambridge University Press, 1985).  
 Galileo, G.: "The Assayer," in *Discoveries and Opinions of Galileo*, ed. S. Drake (New York: Doubleday, 1957).  
 Locke, J.: *An Essay Concerning Human Understanding* (1690), ed. P. H. Nidditch (Oxford: Clarendon Press, 1975).  
 Mackie, J.: *Problems from Locke* (Oxford: Oxford University Press, 1976).  
 Williams, B.: *Descartes: The Project of Pure Enquiry* (Harmondsworth: Penguin, 1978).

BARRY STROUD

**principle of charity** Davidson thinks – and this is one of his most characteristic doctrines – that *ceteris paribus* a sentence of the form "L-speakers hold S true in circumstance C" licenses the corresponding T-sentence "S is true (in L) iff C", but only if a "constitutive principle" of intentional ascription is presupposed; namely, that truth-conditions must be assigned to formulas of L under the constraint that most of the sentences held true by a speaker of L are true (by the interpreter's own lights). This is the principle of charity. It is

supposed to be holistic (see HOLISM) on the intended interpretation, which is that "most of the sentences" means a lot of them. It is primarily on the grounds that the principle of charity is constitutive of intentional ascription, but not of the ascription of physicalistic properties, that Davidson denies the possibility of psycho-physical laws. And it also has striking consequences for epistemology; if it is right, then the refutation of scepticism requires only the weak premise of belief coherence.

See also DAVIDSON; SOCIAL SCIENCES.

#### BIBLIOGRAPHY

- Davidson, D.: "Radical Interpretation," in his *Inquiries into Truth and Interpretation* (Oxford: Oxford University Press, 1984).  
 Davidson, D.: "Belief and the Basis of Meaning," in *ibid.*

ERNEST LEPORE

**principle of contradiction** This is a law of truth. Roughly speaking, a contradictory of a proposition  $p$  is one that can be expressed in the form *not-p*, or, if  $p$  can be expressed in the form *not-q*, then a contradictory is one that can be expressed in the form  $q$ . Thus, e.g., if  $p$  is  $2 + 1 = 4$ , then  $2 + 1 = 4$  is the contradictory of  $p$ , for  $2 + 1 \neq 4$  can be expressed in the form *not-(2 + 1 = 4)*. If  $p$  is  $2 + 1 \neq 4$ , then  $2 + 1 = 4$  is a contradictory of  $p$ , since  $2 + 1 \neq 4$  can be expressed in the form *not-(2 + 1 = 4)*. Thus, mutually contradictory propositions can be expressed in the form,  $r$ , *not-r*. The Principle of Contradiction says that mutually contradictory propositions cannot both be true and cannot both be false. Thus, by this principle, since if  $p$  is true, *not-p* is false, no proposition  $p$  can be at once true and false (otherwise both  $p$  and its contradictory would be false). In particular, for any predicate  $p$  and object  $x$ , it cannot be that  $p$  is at once true of  $x$  and also false of  $x$ . This is the classical formulation of the Principle of Contradiction. There are some senses in which the Principle of Contradiction is not above controversy (see Priest, 1985).

## BIBLIOGRAPHY

- Kneale, W. and Kneale, M.: *The Development of Logic* (Oxford: Clarendon Press, 1962).
- Priest, G., Routley, R. and Norman, J., eds: *Paraconsistent Logics* (Munich: Philosophia Verlag, 1985).

ROBERT S. TRAGESSEER

**principle of credulity** This is a term given by Thomas Reid, in *An Inquiry Into the Human Mind*, Pt VI, sec. 24 (1846, pp. 196–7) (see REID), to an “original principle” of human nature that involves a “disposition to confide in the veracity of others, and to believe what they tell us” (p. 196). This is an “original principle” in that it operates independently of learning or reasoning. We are so constituted that we naturally, unreflectively, tend to give credence to the testimony of our fellows. This “tendency is unlimited in children, until they meet with instances of deceit and falsehood; and it retains a very considerable degree of strength throughout life” (ibid.). As Reid points out, if “no proposition that is uttered in discourse would be believed, until it was examined and tried by reason . . . most men would be unable to find reasons for believing the thousandth part of what is told them. Such distrust and incredulity would deprive us of the greatest benefits of society . . .” (p. 197). Without such an original tendency children would be incapable of learning by instruction. It is only later that the person “sets bounds to that authority to which she was at first entirely subject” (ibid.) (see TESTIMONY). Reid pairs this principle with the “principle of veracity”, which is “a propensity to tell the truth, and to use the signs of language, so as to convey our real sentiments” (p. 196). And he analogizes the principle of credulity with the “inductive principle”, which is an original tendency to suppose that two things that have been constantly conjoined in the past will continue to be so in the future (p. 197). Here, too, Reid argues that if there were no original tendency to such suppositions we could never acquire any by reasoning (see PROBLEMS OF INDUCTION).

Richard Swinburne (1979) uses the term for a quite different principle according to which perceptual beliefs – beliefs based on perceptual experience about what a person seems to be perceiving – are likely to be correct, and hence are rationally acceptable, in the absence of specific reasons against this. “I suggest that it is a principle of rationality that (in the absence of special considerations) if it seems (epistemically) to a subject that  $x$  is present, then probably  $x$  is present; what one seems to perceive is probably so. How things seem to be is good grounds for a belief about how things are” (p. 254). Swinburne enunciates this principle in the course of defending an argument for the existence of God from religious experience (from putative experience of God), but it has much wider application, as he points out. (Reid enunciates a similar principle in his *Essays on the Intellectual Powers of Man*, Essay VI. ch. 5 (1846, p. 445).) Many philosophers are willing to apply it to sense perception, though many of those same philosophers balk at its application to alleged experience (perception) of God. Swinburne, along with other advocates, points out that unless perceptual beliefs are accorded a *prima facie* credibility (see PRIMA FACIE REASONS), just by virtue of being based on perception, it is difficult, if not impossible, to avoid a thoroughgoing scepticism about perception. (See e.g. Reid, (ibid.), Essay II; Moore, 1953, p. 125; Price, 1932, ch. VII.) Many attempts have been made in recent centuries to give a noncircular justification of the reliability of sense perception, but none of them have been generally accepted, and all are subject to severe criticism. (See Alston, 1991, ch. 3 for a critical survey of such attempts.)

We have seen that Swinburne claims only a defeasible justification for all perceptual beliefs (see DEFEASIBILITY), and it is evident that no stronger claim would be warranted. Since perceptual beliefs are not infrequently mutually contradictory, they cannot all be true. And we are familiar with many cases in which one can be shown to be mistaken in one’s suppositions about what one saw. Thus such a principle is only as good as the

account of defeasibility that accompanies it. Swinburne lists four kinds of defeaters:

1. "...the apparent perception was made under conditions or by a subject found in the past to be unreliable" (p. 260)
2. "...the perceptual claim was to have perceived an object of a certain kind in circumstances where similar perceptual claims have proved false" (p. 261)
3. "...one can challenge a perceptual claim to have perceived  $x$  either by showing that probably  $x$  was not present . . .
4. "...or by showing that even if  $x$  was present, it probably did not cause the experience of its seeming that  $x$  was present" (p. 261)

Though these are relevant considerations, I believe we can do better than this by way of a systematic list. For one thing, Swinburne adulterates the classification by mixing grounds for the defeater (inductive or not) with types of defeaters. Thus both (2) and (3) have to do with reasons for thinking the perceptual belief to be false (though in (3) this is somewhat masked by the restriction to what is perceived, in contrast to what is believed about it), and they differ primarily in the kinds of grounds adduced for this. For another, this orientation results in worthy candidates being omitted. Thus we can have reasons for thinking that the experience in question is not a sufficient indication of the truth of the belief. Some cases of this will fall under (1), (2) or (4), but others will not. I would suggest a much simpler classification into *rebutters* (reasons for thinking the belief to be false) and *underminers* (reasons for thinking that in this case the experience is not a sufficient ground for the belief in question). We can then proceed to distinguish various kinds of rebutters and underminers.

Swinburne argues that many religious experiences are not eliminated by the kinds of defeaters he lists. However many thinkers have argued that claims to have experienced God are of doubtful credibility because of more general considerations, such as the spotty distribution of such experiences and

the lack of the kinds of intersubjective checks we have with sense perception. For a response to such arguments, see Alston (1991), chs. 5 and 6.

#### BIBLIOGRAPHY

- Alston, W. P.: *Perceiving God* (Ithaca, NY: Cornell University Press, 1991).  
 Moore, G. E.: *Some Main Problems of Philosophy* (London: Allen and Unwin, 1953).  
 Price, H. H.: *Perception* (London: Methuen, 1932).  
 Reid, T.: *Works*, ed. Sir W. Hamilton, 2 vols (London: Longmans, Green, 1846).  
 Swinburne, R.: *The Existence of God* (Oxford: Clarendon Press, 1979).

WILLIAM P. ALSTON

**principle of identity** Since the same thing may look different at different times, or different things might look the same, or the same thing may be presented in different ways (as  $2 + 3$  and  $4 + 1$  present the same thing, 5, in different ways), we might well raise the question of which of many things (however presented) are the very same things. We would hope that we could answer this question as a matter of principle: given a thing  $x$ , and things of the same kind  $y, z, \dots$ , we might ask for a principle  $P_x$  that determines which of those things  $y, z, \dots$ , if any,  $= x$ . For any considered  $x$  there might be no such principle, or there might be a principle for  $x$  and other things of that kind, or there might be one principle for all things.

It does seem that, outside of mathematics, we do not have any successful, logically impeccable, principles of identity. In mathematics, or some major parts of mathematics, there is Leibniz's Principle of the Identity of Indiscernibles (see LEIBNIZ) – that  $x$  is the very same thing as  $y$  if, and only if,  $y$  has any property that  $x$  has, and conversely. One of Leibniz's main uses of the principle was to deny that there is any more to things than their properties (to deny, against, e.g., Locke, that there are property-bearing substances). But it is clear that this principle is much too strong for, say, trees and persons, at least with

respect to properties we might ordinarily attribute to them, for such of their properties are constantly changing. And yet, in some strong sense, despite persons and trees changing, they remain the same. But in exactly what sense? Or is there an exact sense? Are there logically exact principles legislating or articulating that sameness we sense in changing things?

Both the quest for principles of identity (e.g. principles of personal identity), and broodings over what might count as adequate principles of identity, form flourishing fields of philosophical research.

## BIBLIOGRAPHY

- Chisholm, R. M.: *Person and Object: A Metaphysical Study* (LaSalle, IL: Open Court, 1976).  
 Leibniz, G. W.: *New Essays on Human Understanding*, ed. and trans. P. Remnant and J. Bennett (Cambridge: Cambridge University Press, 1981).  
 Locke, J.: *An Essay Concerning Human Understanding*, ed. P. H. Nidditch (Oxford: Oxford University Press, 1975).  
 Salmon, N.: *Frege's Puzzle* (Cambridge, MA: MIT Press, 1986).  
 Wiggins, D.: *Identity and Spatio-temporal Continuity* (Oxford: Blackwell, 1971).

ROBERT A. TRAGESER

**private language argument** The expression “the private language argument” is sometimes used broadly to refer to a battery of arguments in Wittgenstein’s *Philosophical Investigations* §§243–315 which are concerned with the concepts of, and relations between, the mental and its behavioural manifestations (the inner and the outer), self-knowledge and knowledge of others’ mental states, avowals (see AVOWALS) of experiences and descriptions of experiences. It is sometimes used narrowly to refer to a single chain of argument in which Wittgenstein demonstrates the incoherence of the idea that sensation-names and names of experiences are given meaning by association with a mental “object” (e.g. the word “pain” by association with the sensation of pain) or by mental (private) ostensive

definition (see OSTENSIVE DEFINITION) in which a mental “entity” supposedly functions as a sample (e.g. a mental image, stored in memory, is conceived as providing a paradigm for the application of the name).

A “private language” is not a private code, which could be cracked by another person, nor a language spoken by only one person, which could be taught to others, but rather a putative language, the individual words of which refer to what can (apparently) be known only by the speaker, i.e. to his immediate private sensations or, to use empiricist jargon, to the “ideas” (see IDEA) in his mind. It has been a presupposition of the mainstream of modern philosophy, empiricist (see EMPIRICISM), rationalist (see RATIONALISM) and Kantian alike, of representational idealism (see IDEALISM) no less than of pure idealism, and of contemporary cognitive representationalism (see REPRESENTATION) that the languages we all speak are such private languages, that the foundations of language no less than the foundations of knowledge lie in private experience. To undermine this picture with all its complex ramifications is the purpose of Wittgenstein’s private language arguments (see also FOUNDATIONALISM; GIVEN).

The idea that the language each of us speaks is essentially private, that learning a language is a matter of associating words with, or ostensively defining words by reference to, subjective experiences (the “given”), and that communication is a matter of stimulating a pattern of associations in the mind of the hearer qualitatively identical with that in the mind of the speaker is linked with multiple mutually supporting misconceptions about language, experiences and their identity, the mental and its relation to behaviour, self-knowledge and knowledge of the states of mind of others.

1. The idea that there can be such a thing as a private language is one manifestation of a tacit commitment to what Wittgenstein called “Augustine’s picture of language” – a pre-theoretical picture according to which the essential function of words is to name items in reality, that the link between word and world is effected by ostensive definition (see OSTENSIVE DEFINITION),

and that the essential function of sentences is to describe a state of affairs. Applied to the mental, this preconception yields the following picture: one knows what a psychological predicate such as “pain” means if one knows, is acquainted with, what it stands for – a sensation one has. The word “pain” is linked to the sensation it names by way of private ostensive definition, which is effected by *concentrating* (the subjective analogue of pointing) on the sensation and undertaking to use the word of *that* sensation. First-person present tense psychological utterances, such as “I have a pain” are conceived to be descriptions which the speaker, as it were, reads off the facts which are privately accessible to him.

2. Experiences are conceived to be privately owned and inalienable – no one else can have my pain, but at best only a pain that is qualitatively, but not numerically, identical with mine. They are also thought to be epistemically private – only I *really* know that what I have is a pain, others can at best only believe or surmise that I am in pain.
3. Avowals of experience are expressions of self-knowledge. When I have an experience (e.g. a pain) I am conscious or aware of what I have by introspection (*see* INTROSPECTION) (conceived as a faculty of inner sense). Consequently, I have direct or immediate knowledge of my subjective experience. Since no one else can have what I have, or peer into my mind, my access is privileged. I *know*, and am *certain*, that I have a certain experience whenever I have it, for I cannot doubt that *this*, which I now have, is a pain.
4. One cannot gain introspective access to the experiences of others, so one can obtain only indirect knowledge or belief about them. They are hidden behind the observable behaviour, inaccessible to direct observation, and inferred either analogically (*see* ARGUMENT FROM ANALOGY) or as cause from effect (*see* INFERENCE TO THE BEST EXPLANATION).
5. The observable behaviour from which we thus infer consists of bare bodily move-

ments caused by inner mental events. The outer (behaviour) is not logically connected with the inner (the mental). Hence the mental is essentially private, known *strictu sensu* only to its owner, and the private and subjective is better known than the public.

The resultant picture leads first to scepticism then, ineluctably, to solipsism (*see* SOLIPSISM). Since pretence and deceit are always logically possible, one can never be sure whether another person is really having the experience he behaviourally appears to be having. But worse, if a given psychological predicate means *THIS* (which I have, and no one else *could* logically have – since experience is inalienable), then it is unintelligible that there should be any other subjects of experience. Similar scepticism about communication is unavoidable: if the defining samples of the primitive terms of a language are private, then I cannot be sure that what you mean by “red” or “pain” is not qualitatively identical with what I mean by “green” or “pleasure”. And nothing can stop us from concluding that all languages are private and strictly mutually unintelligible.

Philosophers had always been aware of the problematic nature of knowledge of other minds (*see* OTHER MINDS) and of mutual intelligibility of speech on their favoured picture. It is a manifestation of Wittgenstein’s genius to have launched his attack at the point which seemed incontestable – namely, not whether I can know of the experiences of others, but whether I can know of my own, not whether I can understand the “private language” of another in attempted communication, but whether I can understand my own allegedly private language.

#### THE UNINTELLIGIBILITY OF PRIVATE OSTENSIVE DEFINITION

Whether private ostensive definition is an intelligible notion depends on whether there are private (mental) analogues of the constitutive elements of public ostensive definition (*see* OSTENSIVE DEFINITION): (1) “stage-setting”,



which determines the grammatical category of the definiendum – the ostensive definition being only one rule among others, (2) an ostensive gesture, (3) a sample, (4) a method of projection.

- (1) Public ostensive definition gives a rule for the use of a word. Pointing at a tomato and saying “This is red” *by itself* no more determines the use of “red” than pointing at the moon and howling determines a use for a howl. “Red” is a colour word, and the grammatical category of *colour* fixes a whole host of rules. The grammar of the definiendum does not flow from the object pointed at. Concentrating on one’s toothache and saying “This is pain” does not determine what *this* is. It would have to presuppose the grammar of “sensation”, but that is a word in our public language and is not defined by private ostension.
- (2) The private analogue of pointing is supposedly concentrating one’s attention on one’s pain (which one can do). But concentrating one’s attention on a sensation is not a kind of pointing for oneself alone (although one can point, publicly, at one’s sensation). It does not determine a criterion of identity for subsequent uses of “pain”, and emphatically saying “This” does not either, for what is the “this” one concentrates on? (One cannot reply “a pain”, since that presupposes the very concept one is trying to determine. Nor can one reply “a certain sensation”, or even “a something”, for these are words in our public language with a determinate grammar of their own.)
- (3) A sample functions as a standard of correctness for indefinitely many applications. It must be preservable or reproducible, and the identity of the sample must be fixed (there must be a distinction between a correct and an incorrect selection of a sample in experience the use of a word). But one cannot preserve a sensation, sense-impression or experience for future use as a defining sample. Remembering that pain is *THIS* (conjuring up a mental image) is no substitute,

for on such presuppositions there can be no criterion for reproducing the *right* sample or mental surrogate. Here there would be no distinction between remembering correctly the connection between “S” and the paradigm that defines it and merely thinking one remembers, and there is no independent court of appeal to fix that distinction. But “S”, a sensation-name, does not mean: “whatever occurs to one when one says ‘S’”.

- (4) It must be possible to lay a sample alongside reality for match and mismatch. A method of projection is associated with each category of samples (the methods of projection for samples of colours, lengths, sounds or weights are altogether different). A mental image or representation cannot satisfy this requirement. One cannot *perceive* it (one *has* it). Nor can one lay it alongside reality for match or mismatch. One can say that the curtains are the colour one imagined, but not by *comparing* one’s mental image (which one cannot *see*) with the visible curtains. One may have the same pain in one’s left foot as in one’s right foot, but one cannot justify asserting that one has a pain in one’s left foot by saying “Pain is *THIS*” (and concentrating on one’s pain in the right foot) and adding “and what is in my left foot is *THIS*”. For nothing determines what *THIS* is, and there is here no technique of application for the definiendum, no method of laying the sample alongside reality and nothing that determines identity or difference between sample and the described item.

In short, a “private ostensive definition” cannot determine a rule for the use of a word, can provide neither a private explanation of meaning nor a norm of correctness for the application of a word. It cannot function in a practice with a determinate technique of application. What then replaces the private language account? Truly, “pain” is the name of a sensation, as “red” is the name of a colour. But the moot question is: What is it for a word “S” to name a sensation? It is, *inter alia*, for the utterance “I have an S” to be the

expression or manifestation of a sensation, and to constitute a *criterion* for others to assert "He has an S".

#### CONFUSIONS ABOUT THE PRIVACY OF EXPERIENCE

It is easy to construe "You can't have my pain" to mean that two people cannot have the same pain (i.e. the numerically identical pain) but only similar (qualitatively identical) pains. From this it seems to follow that no one else can really know whether I am in pain or what I really mean by "pain". This is mistaken. The distinction between numerical and qualitative identity which applies to substances has no application to sensations or experiences. One is inclined to think that since, e.g., your headache is in your head and mine is in my head, difference in location, by Leibniz's law, implies numerical difference. This is confused, since for two people to have the same pain in the same place just is for them to have a pain of such-and-such phenomenal characteristics in corresponding parts of their bodies. But one might waive this, and point out that Siamese twins might each suffer pain at the point of juncture; now it might be argued that for all that A's pain is his pain, and B's pain is distinct – for it belongs to him. This is muddled, for the subject of a pain is not a distinguishing mark of the pain, any more than an object is a distinguishing characteristic of its colour. The criteria of identity of a pain include phenomenal characteristics, intensity and location. If these tally between two people, then they do have the same pain.

#### CONFUSIONS ABOUT EPISTEMIC PRIVACY

The doctrines of epistemic privacy, privileged access and immediacy ("direct" knowledge of one's own states of mind) are distortions of various grammatical propositions, viz., that there is no such thing as my *not* knowing, my doubting or wondering, my not being certain whether I am, e.g., in pain, and no such

thing as my having behavioural grounds or evidence for being in pain, as there is no such thing as my misrecognizing and misidentifying my pains. But the grammatical exclusion of doubt does not make room for certainty – rather, it excludes it likewise, as the exclusion of ignorance precludes the intelligibility of knowledge. The grammatical exclusion of behavioural grounds for "self-ascription of experience" does not imply that there are directly observable (introspectible) inner grounds, which are akin to perception. It implies that avowals of experience are not self-ascriptions parallel to other-ascriptions, but groundless *expressions* of the inner – as a groan is a groundless expression of pain. The exclusion of error, misrecognition and misidentification does not ensure infallible knowledge, recognition and identification – rather it precludes any such thing. Hence, Wittgenstein insisted, "I know I am in pain" is either just an emphatic assertion that I am in pain (or a joke) or it is philosophers' nonsense. It is erroneous to think that we know how things are with us inwardly by the faculty of "introspection". Rather we *can* say what we feel, as we can say how things strike us perceptually, or what we intend, imagine or think. The avowal "I have a pain" is typically an *expression* of pain – a learnt extension and partial replacement of a groan (see AVOWALS). It is a criterion for others to ascribe pain to the speaker in a description "He has a pain", but it is not itself a description (though it may be a report). Description typically goes with observing, scrutinizing, examining and investigating; it characteristically involves perceptual competence exercised in various observational conditions, recognition and identification, skill and accuracy of representation (and ways of improving these by closer scrutiny, improved observational conditions, consulting others), the possibility of error (and ways of correcting it), and grounds of judgement. But in the case of expressive uses of first-person present tense psychological utterances (manifestations or avowals of the inner) no perception or perceptual skill is involved, there are no *observational* conditions, there is neither recognition nor misrecognition, identification or

misidentification, no checking by closer scrutiny, no consulting others or discovering from evidence how things are with oneself. One does not “read off” from the “inner facts” how things are with one and render a *description* of them in words for the benefit of others. And much the same goes for one’s sense-impressions, desires, thoughts and intentions – although there are also great differences here. The articulate expression of the inner is not as such a manifestation of self-knowledge, but it is true that a rich inner life is the prerogative of language-users. A dog may expect its master now, but it cannot now expect its master to return next week, for its behavioural repertoire is too impoverished. Nothing it can now do will count as a criterion for now expecting or wanting something to happen next week, or for feeling remorse over what it did last week. Such feelings and desires presuppose the mastery of linguistic skills and their manifestation in articulate expressive behaviour.

#### ASCRPTION OF EXPERIENCE TO OTHERS

The classical picture of our knowledge of “other minds” similarly rests on a wide range of misconceptions. It presupposes that psychological predicates are given meaning by private ostensive definition, and hence that other-ascription of experience involves attributing to others *THIS* (which one now has), on the basis of analogy or “inference to the best explanation”. But private ostensive definition is a *contradictio in adjectio*, and to say that since I know what it is for me to be in pain, therefore I must know what it is for another to be in pain is akin to thinking that since I know what it is to be 5 o’clock here I must know what it is to be 5 o’clock on the sun. This is incoherent. It must first be determined what *counts* as being in pain, i.e. what justifies the employment of this expression. The first-person present tense use is typically a manifestation of the inner, parallel to and in the simplest case a partial substitute for prelinguistic expressive behaviour. The utterance and the non-linguistic behaviour alike

constitute *logical criteria* for third-person ascriptions. More generally, third-person psychological propositions are justifiably asserted on the basis of appropriate behavioural criteria, e.g. typical forms of pain behaviour (including avowals), in appropriate circumstances. These are not inductive evidence discovered by non-inductive identification of the relata and observation of regular correlation, but are logical (grammatical) grounds: *this* is what is called “a cry of pain”, “a scream of agony”, etc. An avowal of experience and an avowal of the identity of a current experience with an antecedent one rest on no criteria, but such avowals together with other forms of expressive behaviour in appropriate circumstances constitute criteria, and criteria of sameness and difference, for the experiences of other people. But it is important to correct misconceptions of human behaviour, for what we call “behaviour”, what we observe when we observe our fellow men, is not “bare bodily movements”, but – laughing with joy, wincing in pain, smiling in amusement, etc. The joy, pain, or amusement are not *accompaniments* of “bare bodily movements” – as it were *hidden* behind the behaviour (i.e. in the mind). They are not hidden, but manifest; they do not accompany the behaviour (as thunder accompanies lightening) but infuse it; they are not behind the behaviour (as the movement of a clock is behind the dial) but visible in it.

To be sure, contrary to behaviourism (see BEHAVIOURISM), pain is not the same as pain-behaviour and joy is distinct from joyful behaviour. For one can be in pain and not show it, and feel joyful without manifesting it. But to feel pain or joy and not show it is not to *hide* anything. One hides one’s feelings when one deliberately *suppresses* them (as one hides one’s thoughts by keeping one’s diary under lock and key, not merely by thinking and not expressing one’s thoughts). When one avows a headache, expresses one’s pleasure or says one thinks it cannot be said that the utterances are mere words and that the inner is still hidden. Talk of the inner is a *metaphor*, and one must beware of looking for an inside *behind* that which in this metaphor is the inside.

We do often know when others are, e.g., in pain, and can be as certain of it as of " $2 + 2 = 4$ ". One cannot say of someone screaming in agony after an accident "Maybe he is not really in pain". One sees the pain on his face, sees that he is suffering. Such knowledge is not *indirect*, for there is no more direct way of knowing that a person is in pain. *He* does not know "directly", since he cannot be said to *know* that he is in pain. Rather, he is in pain and says so! (See PERCEPTUAL KNOWLEDGE.)

It can be misleading to say that one *infers* that someone is in pain from his behaviour, although one might infer from the fact that someone has arthritis that he has pains in his joints. Of course, I may justify saying that I knew he was in pain on the grounds that I saw him writhing in agony (and here I describe the outer in terms of the inner), but it would be misleading to represent this as inferring that he was in pain from his mere behaviour and absurd to say "I saw only his behaviour and inferred that he was in pain".

It is true that pretence is possible and that our judgements here are fallible and defeasible. But it is not *always* possible. It is unintelligible to suppose that a neonate pretends, for pretence has to be *learnt*. Nor is it possible in all circumstances, e.g. when someone falls into flames. Rather, there are circumstance-dependent criteria for pretence no less than for that which is pretended. Hence the possibility of pretence is no more a ground for scepticism about other minds than the possibility of illusion is a ground for scepticism about the existence of objects (see ARGUMENT FROM ILLUSION).

Wittgenstein's private language arguments are not a form of behaviourism. On the contrary, he insists upon the distinction between pain and pain-behaviour and does not claim that the inner is a fiction, but rather that a certain philosophical picture of the inner is a *grammatical* fiction. His arguments do not rely upon a form of verificationism to rebut the supposition of the intelligibility of private ostensive definition, they merely remind us that if we are to talk of a rule for the use of a word, then there must be an operative distinction between the correct and the incorrect application of the rule. He did not argue that the

only refuge against scepticism with regard to language lies in community (public) consensus (and hence that language is essentially social), but rather that the idea of a rule (e.g. of grammar) that can in principle be understood by only one person is unintelligible. The private language arguments overturn the whole tradition of philosophical thought about the nature of the mind and about the relation between the inner and the outer. Their implications ramify widely through philosophy of language (undercutting the idea that the foundations of language lie in private experience), epistemology (undermining the thought that knowledge has private foundations) and philosophical psychology (demolishing the presuppositions of mentalism and behaviourism alike).

See also AVOWALS; FOUNDATIONALISM; INTROSPECTION; OSTENSIVE DEFINITION; OTHER MINDS; SELF-KNOWLEDGE AND SELF-IDENTITY; WITTGENSTEIN.

#### BIBLIOGRAPHY

- Cooke, J. W.: "Human Beings," in *Studies in the Philosophy of Wittgenstein*, ed. P. Winch (London: Routledge and Kegan Paul, 1969), 117–51.
- Hacker, P. M. S.: *Wittgenstein: Meaning and Mind*, vol. 3 of *An Analytical Commentary on the Philosophical Investigations* (Oxford: Blackwell, 1990), 1–286.
- Hacker, P. M. S.: *Insight and Illusion: Themes in the Philosophy of Wittgenstein*, rev. edn (Oxford: Oxford University Press, 1986), 245–336.
- Hanfling, O.: *Wittgenstein's Later Philosophy* (London: Macmillan, 1989), 88–126.
- Kenny, A. J. P.: *Wittgenstein* (London: Allen Lane, 1973), 178–202.
- Malcolm, N.: "Consciousness and Causality," in D. M. Armstrong and N. Malcolm, *Consciousness and Causality: A Debate on the Nature of Mind* (Oxford: Blackwell, 1984), 3–102.
- Malcolm, N.: *Thought and Knowledge* (Ithaca, NY: Cornell University Press, 1977), 85–169.
- Wittgenstein, L.: *Philosophical Investigations* (Oxford: Blackwell, 1953), §§243–315.
- Wittgenstein, L.: "Notes for Lectures on 'Private Experience' and 'Sense Data,'" ed. R. Rhees, *Philosophical Review* 77 (1968), 275–320.

P. M. S. HACKER

**probability, theories of** Philosophical theories of the meaning of “probability” have dealt with a variety of concepts which may be loosely grouped into three families: *degree of belief*, *relative frequency* and *chance*. The mathematics of probability has been less controversial than its interpretation, largely because all theories treat probability as some kind of proportion. Frequentists treat probability as the proportion of actual cases displaying an attribute. Degree of belief and chance theories treat probability as a proportion of possible cases, with the measure on which the proportions are based reflecting respectively intensities of belief that one has or ought to have or intensities of causal tendency. Proportions add. If I have one half of the pie and you have one quarter of it (and we aren’t sharing any) then together we have three quarters of the pie. Proportions are normalized. All of the pie is 100 per cent. Proportions can’t be negative.

#### PROBABILITY CALCULUS

Formally, we can think of a probability as a function mapping elements of a Boolean algebra  $B$  (possible pie slices including the null slice and the whole pie) to real numbers in the interval  $[0, 1]$  which satisfies:

1. If  $b, c$  are disjoint then  $\text{pr}(b \cup c) = \text{pr}(b) + \text{pr}(c)$
2.  $\text{pr}(\text{universal element}) = 1$
3. For no  $b$ , is  $\text{pr}(b) < 0$ .

(It is often assumed that there is an underlying space (the set of pie atoms) such that the Boolean algebra in the foregoing is an algebra of subsets of that space, but this is not essential.) The probability of  $b$  conditional on  $a$  is then defined for elements,  $a$ , with nonzero probability as  $\text{pr}(a \cap b)/\text{pr}(a)$ . Taking probability conditional on  $a$  is just taking the proportion of  $a$  rather than the proportion of the universal element.

It is often mathematically convenient to strengthen the foregoing framework by requiring the operative Boolean algebra to

be closed under countably infinite Boolean operations and to strengthen the principle of additivity to that of countable additivity: If the elements of a countable set,  $\{b_i\}$ , are pairwise disjoint, then:

$$\text{pr} \cup_i (b_i) = \sum_i \text{pr}(b_i)$$

For example, we idealize to say that a countable infinity of disjoint pie slices of  $1/4, 1/8, 1/16 \dots$  of the pie taken together account for half of the pie. This is essentially the standard mathematical formulation of probability as set forth by Kolmogorov in 1933. It allows probability theory to be taken as a branch of measure theory, but on some interpretations of probability the additional requirements are problematic.

(Variations of the foregoing framework are possible. The requirement that the probability function be real-valued can be relaxed to allow infinitesimal probabilities. Intervalvalued rather than point-valued probabilities can be considered. Conditional probabilities can be taken as primitive. Domains for the probability function which do not have the full Boolean structure can be considered.)

#### DEGREE OF BELIEF

It is evident that belief need not be an all or nothing affair; belief comes in degrees. How actual degrees of belief work is a question for psychology, and our best evidence indicates that they systematically fail to satisfy the rules of the probability calculus. It is argued, however, by Frank Ramsey (1931) and by Bruno de Finetti (1937) that consistent or coherent degrees of belief should satisfy the rules of the probability calculus. From this point of view probability theory should be thought of as logic rather than psychology – that is, as the logic of partial belief.

There are different levels on which the argument can be carried out. At the level where complications involving the concept of utility are assumed away, we have the Dutch book arguments. At a deeper level there are representation theorems which show that a



rich, consistent preference ordering must be representable as generated by expected utility computed from underlying probability and utility assignments.

Suppose that a very rich and reliable bookie posts odds on a horse race. He values a contract which pays a pound of gold if Stewball wins at one ounce of gold. He values a contract which pays a pound of gold if Molly wins at one ounce of gold. He will buy or sell these contracts in any quantity at what he considers a fair price or better. Suppose that he also deals in contracts which pay a pound of gold if either Stewball or Molly win, and that he values these at four ounces of gold. You can buy separate contracts for Stewball and for Molly to win at a net outlay of two ounces of gold and sell back a disjunctive contract for either Stewball or Molly to win for four ounces of gold – for a net profit of two ounces of gold. You are able to make a *dutch book* against the bookie – a finite number of bets whose net gain is positive no matter what the outcome of the race – because his evaluation of the disjunctive contract does not cohere with his evaluation of the separate contracts on the disjuncts.

The bookie is said to be *coherent* if no dutch book can be made against him. Coherence requires that if the bookie values the separate contracts at one ounce of gold each, he values the disjunctive contract which pays off exactly as holding both separate contracts together would, at two ounces of gold. Coherence also requires that a contract that pays off a pound of gold no matter what, is worth a pound of gold and that a contract that pays off a pound of gold in some circumstances and requires no payment otherwise has non-negative value (i.e. the bookie will not pay someone to take it off his hands.) Now let us take the value to a person (in pounds of gold) of a contract which pays a pound of gold if  $p$  and nothing otherwise as that person's personal probability of  $p$ . Then it follows immediately from the foregoing that coherence requires that one's personal probabilities satisfy principles 1–3 of the probability calculus. Coherence, as defined, does not require that personal probabilities be countably additive. Strengthening the definition of coherence by

allowing a countable number of bets gets countable additivity, but the epistemological status of such a strengthened notion of coherence is controversial.

Along similar lines, dynamic coherence arguments can be made for updating by Bayes' rule of conditioning – to take one's new probabilities equal to one's old conditional on the evidence – and for more general rules in updating situations in which Bayes' rule is not directly applicable. If we allow the bookie a "spread" between buying and selling prices, we have a natural way of getting interval-valued degrees of belief. C. A. B. Smith (1961) shows that in such a setting coherence requires behaving as if one had a convex set of point probability measures and accepted a transaction just in case it had positive expected value according to all members of that set.

The Dutch book arguments idealize away complications arising from the value of goods, such as the declining marginal utility of gold. Ramsey gave a deeper analysis. Consistency requirements are put directly on a decision-maker's preference order, and a probability-utility representation is gotten from the preferences with the probabilities obeying principles 1–3, and preferences agreeing with expected utility. This general sort of representation theorem can be implemented in a variety of ways. The most well-known probability-utility representation – that of L. J. Savage – does not get countable additivity, but imposition of an additional continuity requirement on preferences will guarantee a countably additivity probability.

Some degree-of-belief theorists, for example de Finetti, hold that rational degrees of belief have to satisfy only the constraints imposed by the probability calculus. Others hold that prior to having any evidence everyone should have the same probability assignment – one determined by symmetry of ignorance or "insufficient reason" or, more recently, computational complexity. The origin of this sort of approach goes back to Bayes and Laplace. In modern times, Jeffreys, Carnap, Jaynes and Solomonoff are – in different ways – exponents of logical determination of the appropriate prior probabilities.

Suppose a six-sided die is roughly chiselled from stone, rolled one hundred times, and then smashed. We take as our events the boolean algebra generated by the possible outcomes on a trial. If we define the probability of an event as the number of trials which exhibit it divided by one hundred, we have a definition which clearly satisfies 1–3. However, such a finite frequency interpretation cannot capture what usually mean when we talk of the probability of an event, for we usually allow that an unlikely sequence of outcomes can occur and we do not want to canonize the frequencies in such a sequence as the true probabilities.

For this reason, frequentists have moved to limiting relative frequency interpretations. Limiting relative frequencies are defined relative to an infinite *sequence* of trials. Here, unlike in the finite case, the order of the sequence can make a difference – so that talk of a reference *class* is highly misleading. There are possible infinite sequences in which the relevant limits do not exist, so the definition must be restricted to those sequences in which they do. Even in this case it should be noted that limiting relative frequencies need not be countably additive. Consider the sequence of the integers, and the properties of being  $= 1$ ,  $= 2$ ,  $= 3$ , . . . Each of these properties has limiting relative frequency of zero, but their countable disjunction has limiting relative frequency of one.

Some frequentists – for example, Reichenbach – are content to deal with reference sequences where the relevant limiting relative frequencies exist, but others – notably von Mises – wished to restrict the theory to sequences which are objectively *random*. This notion of objective randomness is somewhat problematic in von Mises, but has subsequently been substantially clarified.

A version of the strong law of large numbers can be invoked for whatever it is worth. *An infinite sequence of independent and identically distributed trials will with probability 1 produce a random sequence with limiting relative frequencies of outcomes such that the relative frequency of an outcome is equal to its probability.*

But this justification must already refer to a different notion of probability according to which the probability of a single event – of an outcome on a given trial – makes sense. The probability calculus then guarantees agreement with the limiting relative frequencies with probability 1, where that probability 1 is in the antecedent non-frequentist sense of probability.

There are further difficulties with the limiting relative frequency interpretation. Why should the special case of independent identically distributed trials be used to model all probability? In a world in which causation operates many processes do not give independent trials. And if our stone die is to be rolled an infinite number of times, it will presumably become worn so that the trials may not preserve the probabilities. Finally, we may not have infinite sequences at all. Is there no probability if the universe is finite?

To deal with these difficulties a kind of counterfactual limiting relative frequency view has been proposed. The probability of a outcome is taken as the relative frequency that it would have if the experiment were repeated independently an infinite number of times without altering the relevant causal factors in the experimental set-up. This quite sophisticated version of the limiting relative frequency view is already to be found in John Venn's *The Logic of Chance* (1866). It, of course, raises its own set of philosophical questions: specifying the relevant causal factors which individuate the chance set-up, giving the relevant sense of independence in a way that is not circular and interpreting the key counterfactual conditionals.

#### CHANCE

Instead of trying to tell a story in which fictional relative frequencies agree with chances, one might take the course of directly postulating chances as theoretical entities. The chances for outcomes on a single throw of the die are conceived of as elements of physical reality which represent *probabilistic tendencies* or *propensities* to produce those outcomes. The connection between chances and frequencies

is then specified by the strong law of large numbers, the ergodic theorem and other limit theorems of the probability calculus itself.

Why should the chances have the mathematical structure of the probability calculus? This is simply postulated as part of the statistical model under consideration. So the use of countably additive probabilities in classical theories does not – on this view – pose any foundational problem. On the other hand, a propensity theorist might find it plausible to consider modifications of the classical probability calculus in physical theories where such a modified structure would be a useful theoretical tool. This flexibility is just a reflection of the lack of content of the propensity theory as stated. Some advocates of this general approach attempt to say more: for instance, that the chances are determined by the chance set-up, or even that propensities correspond to hypothetical limiting relative frequencies. Evidently, there is no sharp line between relative frequency and propensity conceptions of probability and hybrid philosophical theories are possible.

If chances are to be taken as theoretical entities, then it is important to ask how statements of chance are confirmed or disconfirmed. This is to bring the discussion full circle since it raises the question of rational degrees of belief about the chances. For example, consider our rough stone die. When it is created the true chances are unknown, although we may have some degrees of belief about where the true chances may be. After the die is tossed one hundred times we have some frequency evidence which should enable us to modify our degrees of belief about the true chances.

If, for each chance hypothesis, we have the degree-of-belief probability of the observed outcome sequence conditional on that chance hypothesis, then we can use Bayes' theorem to update our degrees of belief about chance. The natural thing to do – and what everyone does do – is to take the (degree of belief) probability of an outcome sequence conditional on a chance hypothesis to be equal to what the chance hypothesis says it is. That is, we use the following principle, or some close relation to it:

$$\text{pr}[\text{outcome}/\text{Chance}(\text{outcome})] = a \\ = a \quad (\text{M})$$

*Example: A coin with unknown bias is to be flipped. My degree of belief that the coin will come up heads conditional on the hypothesis that it is really biased 2 to 1 in favour of heads, is 2/3.*

The foregoing gives a brief indication of how the three conceptions of chance, frequency and degree-of-belief can interact within a parametric Bayesian framework. It raises another philosophical question: that of the status of (M).

Subjective Bayesians want to do without the metaphysics of chance, and deal with only degrees of belief (which, of course, include degrees of belief about relative frequencies). However, subjective Bayesians do not have to forgo the language of chance and the mathematics of parametric Bayesian statistics. Rather, they regard such talk as a manner of speaking that can be mathematically justified in the presence of symmetries in your degrees of belief. A case in point is de Finetti's famous representation theorem. If one's degrees of belief about a sequence of trials is exchangeable (invariant under finite permutations of trials) then one's degrees of belief are a mixture of probabilities which make the trials independent and identically distributed. Thus one's degrees of belief are *just as if* one had a parametric model with unknown chances. Generalizations of de Finetti's representation theorem handle other important cases. In this subjective setting one can reconstruct the random variable, *chance*, as probability conditional on an appropriate partition or sigma-algebra. One consequence of such a construction is that (M) becomes a theorem.

See also BAYESIAN EPISTEMOLOGY in Part I.

## BIBLIOGRAPHY

- Bayes, T.: "An Essay towards Solving a Problem in the Doctrine of Chances," *Philosophical Transactions of the Royal Society* 53 (1763), 370–418; 54: 296–325; reprinted in *Biometrika* 45 (1958), 293–315.

- Carnap, R.: *Logical Foundations of Probability* (Chicago, IL: University of Chicago Press, 1950).
- Diaconis, P. and Freedman, D.: "Partial Exchangeability and Sufficiency," in *Statistics: Applications and New Directions* (Calcutta: Indian Statistical Institute, 1981), 205–36.
- Diaconis, P. and Zabell, S.: "Updating Subjective Probability," *Journal of the American Statistical Association* 77 (1982), 822–30.
- de Finetti, B.: "La Prévision: ses lois logiques, ses sources subjectives," in *Annales de l'Institut Henri Poincaré* 7 (1937), 1–68; trans. as "Foresight: Its Logical Laws, Its Subjective Sources," in *Studies in Subjective Probability* ed. H. E. Kyburg, Jr and H. Smokler (New York: Kreiger, 1980).
- Fine, T.: *Theories of Probability* (New York: Academic Press, 1973).
- Jeffrey, R.: *The Logic of Decision*, 2nd edn (Chicago, IL: University of Chicago Press, 1983).
- Jeffreys, H.: *The Theory of Probability* (Oxford: Oxford University Press, 1939).
- Kolmogorov, A. N.: *Foundations of the Theory of Probability* (New York: Chelsea, 1959).
- Kyburg, H.: "Propensities and Probabilities," *British Journal for the Philosophy of Science* 25 (1974), 358–75.
- Lane, D. A. and Sudderth, W. D.: "Coherent Predictive Inference," *Sankhya*, ser A, 46 (1984), 166–85.
- Mises, R. von and Geiringer, H.: *The Mathematical Theory of Probability and Statistics* (New York: Academic Press, 1964).
- Ramsey, F. P.: "Truth and probability," in his *The Foundations of Mathematics and Other Essays*, ed. R. B. Braithwaite (New York: Harcourt Brace, 1931).
- Reichenbach, H.: *The Theory of Probability* (Berkeley, CA: University of California Press, 1949).
- Savage, L. J.: *The Foundations of Statistics* (New York: Wiley, 1954).
- Skyrms, B.: *Pragmatics and Empiricism* (New Haven, CT: Yale University Press, 1984).
- Skyrms, B.: "Stability and Chance," with an appendix by P. Vanderschraaf, in *Existence and Explanation*, ed. W. Spohn et al. (Dordrecht: Kluwer, 1991).
- Venn, J.: *The Logic of Chance* (1886); 4th edn (New York: Chelsea, 1962).

BRIAN SKYRMS

**problem of the criterion** This is the problem of how both to formulate the *criteria*, and to determine the *extent*, of knowledge and

justified belief. The following account will focus on justification. The problem arises from the seeming plausibility of the following two propositions:

- (1) I can identify instances (and thus determine the extent) of justified belief only if I already know the criteria of it.
- (2) I can know the criteria of justified belief only if I can already identify the instances of it.

If both (1) and (2) were true, I would be caught in a circle: I could know neither the criteria nor the extent of justified belief. In order to show that both can be known after all, a way out of the circle must be found. The nature of this task is best illustrated by considering the four positions that may be taken concerning the truth values of (1) and (2):

- (a) Scepticism as to the possibility of constructing a theory of justification: Both (1) and (2) are true; consequently, I can know neither the criteria nor the extent of justified belief. (This kind of scepticism is restricted in its scope to epistemic propositions. While it allows for the possibility of justified beliefs, it denies that we can *know* which beliefs are justified and which are not.)
- (b) (2) is true but (1) is false: I can identify instances of justification without applying a criterion.
- (c) (1) is true but (2) is false: I can identify the criteria of justified belief without prior knowledge of its instances.
- (d) Both (1) and (2) are false; I can know the extent of justified belief without applying criteria, and vice versa.

The problem of the criterion may be seen as the problem of providing a rationale for a non-sceptical response, that is, for either (b), (c), or (d).

Roderick Chisholm (see CHISHOLM), who has devoted particular attention to this problem, calls the second response *particularism* and the third *methodism*. Hume (see HUME), who draws a sceptical conclusion as to the extent of empirical knowledge using "deducibility from sense-experience" as the criterion

of justification, was a methodist. Thomas Reid (*see* REID) and G. E. Moore (*see* MOORE) were particularists; they rejected Hume's criterion on the ground that it turns obvious cases of knowledge into cases of ignorance. Chisholm advocates particularism as the correct response. His view, which has also become known as critical cognitivism (*see* COMMONSENSISM AND CRITICAL COGNITIVISM), may be summarised as follows. Criteria for the application of epistemic concepts are expressed by epistemic principles. The antecedent of such a principle states the non-normative ground on which the epistemic status ascribed by the consequent supervenes (*see* EPISTEMIC SUPERVENIENCE). (Cf. Chisholm, 1957, pp. 30–9, 1982, p. 12.) An example is the following:

If S is appeared to F-ly, then S is justified in believing that there is an F in front of S.

According to this principle, a criterion for justifiably believing that there is something red in front of me is "being appeared to redly". In constructing his theory of knowledge, Chisholm considers various principles of this kind, accepting or rejecting them depending on whether or not they fit what he identifies, without using any criterion, as the instances of justified belief. As the result of using this method, he rejects the principle above as too broad, and Hume's empiricist criterion (which, unlike the criteria Chisholm tries to formulate, states a necessary condition),

If S is justified in believing that there is an F in front of S, then S's belief is deducible from S's sense-experience

as too narrow. (Cf. Chisholm, 1982, ch. 5; 1977, chs 4 and 7.)

Regarding the viability of particularism, this approach raises the question of how it is possible to identify instances of justified belief without applying any criteria. Chisholm's answer rests on the premise that, in order to know, no criterion of knowledge or justification is needed (1982, p. 53). He claims that this holds also for knowledge of epistemic facts. Supposing I am justified in believing that *p*, what justifies me in believing that I am

justified in believing that *p* is the same body of evidence that justifies me in believing that *p*. Put differently, both *JJp* and *Jp* supervene on the same non-epistemic ground. (Cf. Chisholm 1982, ch. 4. For a dissenting view, *see* Alston, 1989, p. 24; cf. Van Cleve, 1979, p. 86.) Thus in order to become justified in believing myself to be justified in believing that *p*, I need not apply any criterion of justified belief; I need only *consider* the evidence supporting *p*. The key assumption of particularism, then, is that in order to acquire knowledge of an epistemic fact, one need not *apply*, but only *satisfy* the antecedent condition of, the epistemic principle that governs the fact in question. Hence it is possible to have knowledge of epistemic facts such as "I am justified in believing that there is an F in front of me" without applying epistemic principles, and to use this knowledge in order to reject those principles that are either too broad or too narrow.

According to methodism, the correct solution to the problem proceeds the opposite way: epistemic principles are to be formulated without using knowledge of epistemic facts. But how could methodists distinguish between correct and incorrect principles, given that an appeal to instances of epistemic knowledge is illegitimate? Against what could they check the correctness of a putative principle? Unless the correct criteria are immediately obvious, which is doubtful, it remains unclear how methodists could rationally prefer one principle over another. Thus Chisholm rejects Hume's criterion not only because of its sceptical implications but also on the ground of its arbitrariness: Hume "leaves us completely in the dark so far as concerns what *reason* he may have for adopting this particular criterion rather than some other" (1982, p. 67). Particularists, then, accept proposition (2), and thus reject responses (c) and (d), both of which affirm that (2) is false.

One problem for particularism is that it appears to beg the question against scepticism, (cf. Bonjour, 1985, p. 13). In order to evaluate this criticism, it must be kept in mind that particularists reject criteria with sceptical consequences on the basis of instances, whereas sceptics reject instances of justification on the basis of criteria. This



difference in methodology is illustrated by the following two arguments:

*An Anti-sceptical Argument*

(A) If the "deducibility from sense-experience" criterion is correct, then I am not justified in believing that these are my hands.

(B) I am justified in believing that these are my hands.

Therefore:

(-C) The "deducibility from sense-experience" criterion is not correct.

*A Sceptical Argument*

(A) If the "deducibility from sense-experience" criterion is correct, then I am not justified in believing that these are my hands.

(C) The "deducibility from sense-experience" criterion is correct.

Therefore:

(-B) I am not justified in believing that these are my hands.

The problematic premisses are (B) and (C). Particularists reject (C) on the basis of (B), and sceptics (B) on the basis of (C). Regarding question-begging, then, the situation is symmetrical: both beg the question against each other. Who, though, has the better argument? Particularists would say that accepting (B) is more reasonable than accepting (C) because the risk of making an error in accepting a general criterion is greater than in taking a specific belief to be justified.

The problem of the criterion is not restricted to epistemic justification and knowledge but is posed by any attempt to formulate general principles of philosophy or logic. In response to the problem of induction, Nelson Goodman (see GOODMAN) has proposed bringing the principles of inductive inference into agreement with the instances of inductive inference we accept. John Rawls has attempted to formulate principles of justice with the objective of bringing the principles and what we count as examples of justice into a state of reflective equilibrium (cf. Goodman, 1965, pp. 66–7; Rawls, 1971, pp. 19–21, 48–51). Goodman and Rawls believe that in order to identify the principles they seek their instances must be known to

begin with. But they believe also that in the process of bringing principles and instances into agreement, principles may be saved by sacrificing instances. They may, therefore, be considered advocates of a view analogous to response (iv), a hybrid of particularism and methodism. (For a critical discussion of Goodman's approach, see Stich, 1988; cf. Sosa, 1989.)

See also CHISHOLM; COMMONSENSISM AND CRITICAL COGNITIVISM; CRITERIA AND KNOWLEDGE; SEXTUS EMPIRICUS.

BIBLIOGRAPHY

- Alston, W. P.: *Epistemic Justification* (Ithaca, NY: Cornell University Press, 1989).
- BonJour, L.: *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1985).
- Chisholm, R.: *Perceiving* (Ithaca, NY: Cornell University Press, 1957).
- Chisholm, R.: *The Foundations of Knowing* (Minneapolis, MN: University of Minnesota Press, 1982).
- Chisholm, R.: *Theory of Knowledge*, 2nd edn (Englewood Cliffs, NJ: Prentice-Hall, 1977).
- Goodman, N.: *Fact, Fiction, and Forecast* (Indianapolis, IN: Bobbs-Merrill, 1965).
- Rawls, J.: *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971).
- Sosa, E.: "Equilibrium in coherence," in *The Current State of the Coherence Theory*, ed. J. W. Bender (Dordrecht: Kluwer, 1989), 242–50.
- Sosa, E.: "The Foundations of Foundationalism," *Nous* 14 (1980), 547–64.
- Stich, S.: "Reflective Equilibrium, Analytic Epistemology and the Problem of Cognitive Diversity," *Synthese* 74 (1988), 391–413.
- Van Cleve, J.: "Foundationalism, Epistemic Principles, and the Cartesian Circle," *Philosophical Review* 88 (1979), 55–91.

MATTHIAS STEUP

**problem of the external world** An external world, as philosophers have used the term, is not some distant planet external to earth. Nor is the external world, strictly speaking, a *world*. Rather, the external world consists of all those objects and events which

exist external to perceivers. So the table across the room is part of the external world, and so is its brown colour and roughly rectangular shape. Similarly, if the table falls apart when a heavy object is placed on it, the event of its disintegration is a part of the external world.

One object external to and distinct from any given perceiver is any other perceiver. So, relative to one perceiver, every other perceiver is a part of the external world. However, another way of understanding the external world results if we think of the objects and events external to and distinct from every perceiver. So conceived the set of all perceivers makes up a vast community, with all of the objects and events external to that community making up the external world. In this essay we will understand the notion of an external world in the former way. We will thus suppose that perceivers are entities which occupy physical space, if only because they are partly composed of items which take up physical space.

What, then, is the problem of the external world (hereafter the PEW)? Certainly it is not whether there is an external world; this much is taken for granted. Instead, the problem is an epistemological one which, in rough approximation, can be formulated by asking whether and if so how a person gains knowledge of the external world. So understood, the problem seems to admit of an easy solution. There is knowledge of the external world which persons acquire primarily by perceiving objects and events which make up the external world.

However, many philosophers have found this easy solution problematic. Indeed, the very statement of the PEW itself will be altered once we consider the main arguments against the easy solution.

#### AN EPISTEMIC ARGUMENT

One way in which the easy solution, mentioned above, has been further articulated is in terms of epistemological direct realism (*see* DIRECT REALISM) (hereafter, EDR). This theory is realist in so far as it claims that objects and

events in the external world, along with many of their various features, exist independently of and are generally unaffected by perceivers and acts of perception in which they engage. And this theory is epistemologically direct since it also claims that in perception people often, indeed typically, acquire immediate non-inferential knowledge of objects and events in the external world. It is on this latter point that it is thought to face serious problems.

The main reason for this is that knowledge of objects in the external world seems to be dependent on some other knowledge, and so would not qualify as immediate and non-inferential. It is claimed that I do not gain immediate non-inferential perceptual knowledge that there is a brown and rectangular table before me, because I would not know such a proposition unless I knew that something then appeared brown and rectangular. Hence, knowledge of the table is dependent upon knowledge of how it appears. Alternately expressed, if there is knowledge of the table at all, it is indirect knowledge, secured only if the proposition about the table may be inferred from propositions about appearances. If so, EDR is false (*see* Chisholm, 1957, pp. 55ff).

This argument suggests a new way to formulate the PEW:

PEW 1: Can one have knowledge of propositions about objects and events in the external world based upon propositions which describe how the external world appears, i.e. upon appearances?

Unlike our original formulation of the PEW, this formulation does not admit of an easy solution. Indeed, it has seemed to many philosophers that it admits of no solution at all, so that scepticism regarding the external world is the only remaining alternative.

#### PERCEPTUAL ARGUMENTS

If we think back to the easy solution to the first version of PEW, we note that it says that a

person gains knowledge of objects and events in the external world primarily by perceiving them. If we concentrate on perception, a slightly different version of the easy solution emerges, one which incorporates perceptual direct realism (PDR). This theory is realist in just the way described earlier, but it adds, secondly, that objects and events in the external world are typically directly perceived, as are many of their features such as their colours, shapes and textures.

Often, PDR is developed further by simply adding EDR to it. Such an addition is supported by claiming that direct perception of objects in the external world provides us with immediate non-inferential knowledge of such objects. Seen in this way, PDR is supposed to support EDR, though strictly speaking they are independent doctrines. One might consistently, perhaps even plausibly, hold one without also accepting the other. (I think Berkeley (*see* BERKELEY) did so; *see* Pappas, 1991.)

Direct perception is that perception which is not dependent on some other perception. The main opposition to the claim that we directly perceive external objects comes from indirect or representative realism (*see* REPRESENTATIVE REALISM). That theory holds that whenever an object in the external world is perceived, some other object is also perceived, namely a *sensum* – a phenomenal entity of some sort. Further, one would not perceive the external object if one were to fail to perceive the *sensum*. In this sense the *sensum* is a perceived intermediary, and the perception of the external object is dependent on the perception of the *sensum*. For such a theory, perception of the *sensum* is direct, since it is not dependent on some other perception, while perception of the external object is indirect. More generally, for the indirect realist, all directly perceived entities are *sensa*. On the other hand, those who accept PDR claim that perception of objects in the external world is typically direct, since that perception is not dependent on some perceived intermediaries such as *sensa*.

It has often been supposed, however, that the argument from illusion (*see* ARGUMENT FROM ILLUSION) suffices to refute all forms of PDR. The argument from illusion is actually

a family of different arguments rather than one argument (*see* Pitcher 1970; Cornman, 1971). Perhaps the most familiar argument in this family begins by noting that objects appear differently to different observers, and even to the same observer on different occasions or in different circumstances. For example, a round dish may appear round to a person viewing it from directly above and elliptical to another viewing it from one side. As one changes position the dish will appear to have still different shapes, more and more elliptical in some cases, closer and closer to round in others. In each such case, it is argued, the observer directly sees an entity with that apparent shape. Thus, when the dish appears elliptical, the observer is said to see directly something which is elliptical. Certainly this elliptical entity is not the dish nor the top surface of the dish, since that is round. This elliptical entity, a *sensum*, is thought to be wholly distinct from the dish.

In seeing the dish from straight above it appears round, and it might be thought that then one directly sees the dish rather than a *sensum*. But here too relativity sets in; the dish will appear different in size as one is placed at different distances from the dish. So even if in all of these cases the dish appears round, it will also appear to have many different diameters. Hence, in these cases as well, the observer is said to directly see some *sensum*, and not the dish.

This argument concerning the dish can be generalized in two ways. First, more or less the same argument can be mounted for all other cases of seeing and across the full range of sensible qualities – textures and colours in addition to shapes and sizes. Second, one can utilize related relativity arguments for other sense modalities. With the argument thus completed, one will have reached the conclusion that in all cases of non-hallucinatory perception, the observer directly perceives a *sensum*, and not an external physical object. Presumably in cases of hallucination a related result holds, so that one reaches the fully general result that in all cases of perceptual experience, what is directly perceived is a *sensum* or group of *sensa*, and not an external physical object. PDR, therefore, is deemed false.

Yet even if PDR is refuted, this by itself does not generate a problem of the external world. We need to add that if no person ever directly perceives an external physical object, then no person ever gains immediate non-inferential knowledge of such objects. Armed with this additional premise, we could conclude that if there is knowledge of external objects, it is indirect and based upon immediate knowledge of *sensa*. We can then formulate the problem of the external world in another way:

PEW2: Can one have knowledge of propositions about objects and events in the external world based upon propositions about directly perceived *sensa*?

It is worth noting the differences between PEW1 and PEW2. The arguments which lead to these problems are quite different, with those directing us to PEW2 concerned with perception in a way not mentioned in arguments leading to PEW1. Also, attempts to solve PEW2 require accounts of perception alternative to PDR, while this is not so for PEW1.

#### PROPOSED SOLUTIONS TO PEW 1

If the argument leading to PEW1 is satisfactory, then we have knowledge of the external world only if propositions about objects and events in the external world (hereafter E-propositions) are inferable from propositions about appearances (A-propositions). They will be so inferable only if they are either deducible from E-propositions, or are inferable by some cogent inductive inference. It is clear, however, that E-propositions are not deducible from any finite group of A-propositions. We can see this by noting that the A-propositions might all be true in a case of a hallucination, when the E-proposition in question is plainly false. Thus, A-propositions such as those expressed by "I seem to see something round, and red and spherical" and "I seem to be tasting something sweet and slightly tart" do not entail that one is seeing or tasting an apple.

Some philosophers have thought that if analytical phenomenalism (*see* PHENOMENALISM) were true, the situation would be different. Analytical phenomenalism is the doctrine that every E-proposition is fully analysable into, and thus is equivalent in meaning to, a group of A-propositions (*see* Ayer, 1940; Lewis, 1946). The numbers of A-propositions making up the analysis of any single E-proposition would likely be enormous, perhaps indefinitely many. Nevertheless, analytical phenomenalism might be of help in solving PEW1 because the required deductions of E-propositions from A-propositions could be readily made. An E-proposition equivalent in meaning to some A-propositions can certainly be deduced from those A-propositions.

However, matters are not as simple as these remarks make it seem. Even if analytical phenomenalism is true, no reasonably sized set of A-propositions will *entail* an E-proposition. For, given analytical phenomenalism there are indefinitely many A-propositions in the analysis of each E-proposition. Hence, the inference from any set of A-propositions to an E-proposition is apt to be inductive, even granting the truth of analytical phenomenalism. Moreover, most of the A-propositions into which we might hope to analyse an E-proposition would be complex subjunctive conditionals such as that expressed by "If I were to seem to see something red, round and spherical, and if I were to seem to try to taste what I seem to see, then most likely I would seem to taste something sweet and slightly tart." But A-propositions of this complex sort will not typically be immediately known, and thus knowledge of E-propositions will not generally be based upon immediate knowledge of such A-propositions.

Moreover, there is good reason to think that analytical phenomenalism is false. For each proposed translation of an E-proposition into A-propositions has been shown to be defective, and there is no reason to suppose that any new attempt at a translation of this sort will succeed where all others have failed (*see* Chisholm, in Swartz, 1965).

But neither are E-propositions *inductively* derivable from A-propositions. Plainly

enumerative induction is of no help in this regard, for that is an inference from premisses about observed objects in a certain class having some properties F and G, to unobserved objects in the same class having properties F and G. A-propositions, however, concern appearances while E-propositions concern external objects and events. So, the most likely inductive inference to consider is a causal one: we infer from certain effects, described by A-propositions, to their likely causes, described by P-propositions. But here, too, the inference is apt to prove problematic.

Consider the A-propositions expressed by "I seem to see something red, round and spherical" and "I seem to taste something sweet and slightly tart". To infer cogently from these propositions to that expressed by "There is an apple before me" we need additional information, such as that expressed by "Apples generally cause visual appearances of redness, roundness and spherical shape, and gustatory appearances of sweetness and tartness". With this additional information, the inference is a good one, and it is likely to be true that there is an apple there relative to those premisses. The cogency of the inference, however, depends squarely on this additional premise; relative only to the stated A-propositions, it is not highly probable that there is an apple there.

Defenders of indirect realism have sometimes appealed to an inference to the best explanation (see INFERENCE TO THE BEST EXPLANATION) to help justify E-propositions. We might say that the best explanation of the appearances is that they are caused by external objects. However, even if this is true, as no doubt it is, it is unclear how establishing this *general* hypothesis helps to justify *specific* E-propositions such as that these particular appearances are caused by that red apple.

The point here is a general one: cogent inductive inferences from A-propositions to E-propositions are available only with some added premiss expressing the requisite causal relation, or perhaps some other premiss describing some other sort of correlation between appearances and external objects. So there is no reason to think that PEW1 can be solved by exhibiting inductive inferences

from A-propositions to E-propositions (see Chisholm, 1957, pp. 73–4). And since deductive and inductive inferences from A-propositions to E-propositions seem to exhaust the options, no solution to PEW1 is at hand. So, unless there is some solution to PEW2, it would appear that scepticism concerning knowledge of the external world would be the most reasonable position to take.

#### PROPOSED SOLUTIONS TO PEW 2

PEW2 is generated by the supposed refutation of PDR, conjoined with the principle that one has immediate perceptual knowledge of something only if one directly perceives that object. Broadly speaking, there are two alternatives to PDR: (perceptual) indirect realism and (perceptual) phenomenalism. In contrast to indirect realism, described above, perceptual phenomenalism rejects realism outright, and holds instead that: (1) physical objects are collections of sensa; (2) in all cases of perception, at least one sensum is directly perceived; and (3) to perceive a physical object one directly perceives some of the sensa which are constituents of the collection making up that object.

Proponents of each of these positions try to solve PEW2 in different ways. But are they in fact any better able to solve PEW2 than related doctrines we discussed earlier were able to solve PEW1? The answer has seemed to most philosophers to be NO, for in indirect realists and phenomenologists have used strategies we have already considered and rejected.

To see this, let us use the term "S-proposition" for propositions which describe presently directly perceived sensa. Indirect realists typically claim that the inference from S-propositions to E-propositions would be an inductive one, specifically a causal inference from effects to causes. Inferences of such a sort will be perfectly cogent provided we can use a premiss which specifies that physical objects of a certain type are causally correlated with sensa of the sort currently directly perceived. Such a premiss will itself be justified, if at all, solely on the basis of S-propositions. Certainly for the indirect realist one never



directly perceives the causes of *sensa*. So, if one knows that, say, apples typically cause such-and-such visual *sensa*, one knows this only indirectly on the basis of knowledge of *sensa*. But no group of S-propositions by itself supports any inferences to causal correlations of this sort. Consequently, indirect realists are in no position to solve PEW2 by showing that E-propositions are inductively derivable from S-propositions.

Phenomenalists have often supported their position, in part, by noting the difficulties facing indirect realism. But phenomenalism is no better off with respect to PEW2. Phenomenalists construe physical objects as collections of *sensa*. So, to infer an E-proposition from some S-propositions is to infer a proposition about a collection from propositions about constituent members of the collection. This inference, too, will be an inductive one, albeit not a causal one. None the less, a related problem faces the phenomenalist, namely the inferences in question will require a premiss that such-and-such directly perceived *sensa* are constituents of some collection C, where C is some physical object such as an apple. The problem comes with trying to justify such a premiss. To do this, one will need some plausible account of what is meant by claiming that physical objects are collections of *sensa*. To explicate this idea, however, phenomenalists have typically turned to analytical phenomenalism: physical objects are collections of *sensa* in the sense that propositions about physical objects are analysable into propositions about *sensa*. And analytical phenomenalism, we have seen, has been discredited.

If neither PEW1 nor PEW2 can be easily solved, then scepticism about the external world is a doctrine we would be forced to adopt. One might even say that it is here that we locate the real problem of the external world: how can we avoid being forced into accepting scepticism.

#### AVOIDING SCEPTICISM

The best answer, I think, is to question the arguments which lead to PEW1 and PEW2.

Beginning with the latter, the crucial question is whether any part of the argument from illusion really forces us to abandon PDR. To help see that the answer is NO we may note that a key premise in the relativity argument considered earlier links how something appears with direct perception: the fact that the dish appears elliptical is supposed to entail that one directly perceives something which is elliptical (*see* Moore, 1965). But is there an entailment here? Certainly we do not think that the proposition expressed by "The book appears worn and dusty and more than two hundred years old" entails that the observer directly perceives something which is worn and dusty and more than two hundred years old (Chisholm, 1964, p. 95). And there are countless other examples like this one, where we will resist the inference from a property F appearing to someone to the claim that F is instantiated in some entity.

Proponents of the argument from illusion might complain that the inference they favour works only for certain adjectives, specifically for adjectives referring to nonrelational sensible qualities such as colour, taste, shape, and the like. Such a move, however, requires an argument which shows why the inference works in these restricted cases and fails in all others. No such argument has ever been provided, and it is difficult to see what it might be.

If the argument from illusion is defused, the major threat facing PDR will have been removed. Hence, there will no longer be any real motivation for PEW2. Of course, even if PDR is reinstated, this does not solve PEW1. That problem might arise even for one who accepts PDR. But here, as well, there is reason to be suspicious. In keeping with the argument used to generate PEW1, let us grant that one would not know that one is seeing something blue if one failed to know that something looked blue. In this sense there is a dependence of the former on the latter, as noted in the argument. What is not clear is whether the dependence is *epistemic* or *semantic*. It is the latter if, in order to understand what it is to see something blue, one must also understand what it is for something to look blue. This may be true, even when the belief that one is

seeing something blue is not epistemically dependent on (based upon) the belief that something looks blue. Merely claiming, as in the first argument, that there is a dependence relation does not discriminate between epistemic and semantic dependence. Moreover, there is reason to think it is not an epistemic dependence. For in general, observers rarely have beliefs about how objects appear, but this fact does not impugn their knowledge that they are seeing, e.g. blue objects (see Pollock, 1986, p. 61).

This criticism of the argument used for PEW1 is narrow, in the sense that it focuses only on individual elements within the argument but does not question the background assumptions on which the argument seems to be based. Those assumptions, broadly speaking, are foundationalist in character: knowledge and justified belief are divided into the basic, immediate and non-inferential cases, and the non-basic, inferential knowledge and justified belief which is supported by the basic. However, though foundationalism (see FOUNDATIONALISM) was widely assumed when the problem of the external world was given currency in Descartes (see DESCARTES) and the classical empiricists (see EMPIRICISM), it has been repeatedly challenged and there are in place well-worked alternative accounts of knowledge and justified belief, some of which seem to be as plausible as the most tenable version of foundationalism. So we have some good reason to suspect, quite as one might have initially thought, that the problem of the external world just does not arise, at least not in the forms in which it has usually been presented.

See also ARGUMENT FROM ILLUSION; DIRECT REALISM; EXPERIENCE; INFERENCE TO THE BEST EXPLANATION; PHENOMENALISM; REPRESENTATIVE REALISM; SCEPTICISM, CONTEMPORARY; TRANSCENDENTAL ARGUMENTS.

## BIBLIOGRAPHY

- Ayer, A. J.: *Foundations of Empirical Knowledge* (London: Macmillan, 1940).  
Chisholm, R.: "The Problem of Empiricism," in Swartz, 1965.

- Chisholm, R.: *Perceiving* (Ithaca, NY: Cornell University Press, 1957).  
Chisholm, R.: *Theory of Knowledge*, 1st edn (Englewood Cliffs, NJ: Prentice-Hall, 1964).  
Cornman, J.: *Materialism and Sensations* (New Haven, NJ: Yale University Press, 1971).  
Lewis, C. I.: *An Analysis of Knowledge and Valuation* (LaSalle, IL: Open Court, 1946).  
Moore, G. E.: "Visual Sense-Data," in Swartz, 1965.  
Pappas, G.: "Berkeley and Common Sense Realism," *History of Philosophy Quarterly* 8 (1991), 27–42.  
Pitcher, G.: *A Theory of Perception* (Princeton, NJ: Princeton University Press, 1970).  
Pollock, J.: *Contemporary Theories of Knowledge* (Totowa, NJ: Rowman and Littlefield, 1986).  
Swartz, R., ed.: *Perceiving, Sensing and Knowing* (New York: Doubleday, 1965).

GEORGE PAPPAS

**problem of induction** see PROBLEMS OF INDUCTION.

**problem of other minds** see OTHER MINDS.

**problem of rule-following** Rule-following is an intentional activity of the sort that may be involved in using words, moving chess-pieces, adopting local custom and thinking straight. It is the activity of intentionally conforming or trying to conform to the rules relevant in such areas. The problem of rule-following is that of explaining how such activity is possible. Rule-following requires the agent to identify something – a rule – that prescribes what to do in an indefinitely large and varied range of situations and then to try to remain faithful to the prescriptions of that rule. It is difficult to see what sort of thing, among the items that a human mind can access, could serve this indefinitely prescriptive function. The problem of rule-following is to resolve that difficulty.

The problem derives from the later work of Ludwig Wittgenstein (see WITTGENSTEIN). (See especially Wittgenstein, 1953, 1956.) Although it had attracted considerable attention in the first phase of Wittgenstein's influence, it tended to be eclipsed by issues associated with the private language argument (see PRIVATE LANGUAGE ARGUMENT). It was

only in the 1970s and 1980s that it came to the fore as a problem in its own right. This was due in particular to the work of Robert Fogelin (1987), Saul Kripke (1982) and Crispin Wright (1980). (*See also* Holtzman and Leich, 1981; Wright, 1984.)

There has been a variety of approaches canvassed to the solution of the rule-following problem. The possible solutions would include ones that take rules as platonic entities and that ascribe to us an ability to get in tune with those entities: to main-line them, as it were. But most approaches attempt to solve the problem within a naturalistic framework (*see* NATURALISM) that precludes the positing of such non-natural entities and the ascription of such non-routine abilities. They try to show that naturalists are not forced to the iconoclastic position described in Kripke (1982) – the so-called sceptical solution – according to which rule-following is an illusion; there is simply nothing of the kind going on. But the concern here is not with the different possible solutions to the problem. (For a survey of solutions, *see* Boghossian, 1989; *see also* Pettit, 1990a and b.) The concern is rather with the characterization of the problem of rule-following. I will pursue that concern by dealing in turn with three distinct questions. What are rules? What is it to follow rules? And what is the problem with the notion of following rules?

#### WHAT ARE RULES?

As I invoke the notion, and as it is commonly invoked, rules are normative constraints in types of decision or, more generally, judgement; decisions can be taken to be judgements as to what is best. That something is a normative constraint in a type of judgement means that it identifies one or more options among the alternatives to be adjudicated as more appropriate in some way than others. The option may be the most polite, as with a rule of etiquette; the most becoming, as with a rule of fashion; the just option, as with a rule of fairness; the proper verdict, as with a rule of evidence; the right thing to say, as with a rule of truth-telling; or whatever.

The rules with which the tradition has been concerned are relevant in an indefinitely large variety of situations and it will be useful if we build the assumption of such open-ended relevance into our conception of rules. Thus we take rules as normative constraints that are relevant in an indefinitely large number of judgement-types. Most of the rules with which we are familiar involve normative constraints that are relevant in an indefinite variety of situations. There is usually no mechanical way of specifying even the different situations where a rule of etiquette applies; such rules, as it is often put, are open-textured.

What sort of an entity is going to represent a normative constraint over an indefinite variety of situations? One type of entity that could serve in that role would be an indefinitely large set of pairs, one for every judgement-type to which the rule is relevant: the first member in the pair would be the type of situation involved, the other the appropriate type of option. We might refer to such an indefinitely extended object as the rule-in-extension. Another sort of entity that could serve in the role required would be the abstract object which is conceived as having the property of identifying the appropriate option for every relevant situation to which it is applied; the entity would be an abstract function which, given a situation-type as input, delivers the appropriate type of option as output. We might call this the rule-in-intension.

#### WHAT IS IT TO FOLLOW A RULE?

Here the important distinction to draw is between conforming to a rule and following it. To conform to a rule is to form judgements or make decisions which satisfy the normative constraint in question: to behave in the manner that is appropriate according to that rule. Conforming to a rule in this sense does not require a knowledge of what the rule is, or a knowledge that conformity has been achieved. All the more obviously it does not require a desire on the agent's part to conform, or an attempt to satisfy that desire. That a subject conforms to a certain rule is a relational

fact about its behaviour that tells us nothing, in itself, about the agent's state of mind.

To follow a rule however, as distinct from just conforming to it, is to exhibit a certain state of mind; it is not just to satisfy a behavioural specification. To follow a rule is to conform to the rule intentionally, to conform to the rule through trying to conform to it, to conform to the rule because of acting on the basis of a desire to achieve conformity. An agent follows a rule in the judgements it forms just in case its making judgements in accordance with that rule is intentional: just in case a rationalizing set of beliefs and desires suitably explains its making such judgements.

(There may seem to be a problem with the notion of an agent intentionally making judgements that conform to a rule. We know that it cannot be intentional on an agent's part that it judges that *p* or that *q*. So how can it be intentional that it judges in accordance with a certain rule? No problem. An agent may not judge intentionally that *p*, though it does judge intentionally in accordance with a certain rule, where so judging is just judging that *p*. That it judges that *p* is not due, even in part, to any desire that it has; that it judges in conformity to a rule is due, at least in part, to such a desire. In the context of its other beliefs the desire to conform to the rule explains the subject's judging in conformity to the rule but the desire does not in the same way explain the subject's judging that *p*; the desire would tend to promote a judgement that *q*, did the rule require that *q*.)

#### THE PROBLEM OF RULE-FOLLOWING

There are two conditions that must be fulfilled by any rule, if the rule is to be capable of being followed. It must meet the objective condition of being or fixing a normative constraint that applies in an indefinite variety of cases. And it must meet the subjective condition of engaging appropriately with our intentional projects: of being something to which a creature like one of us can try to conform. The problem of rule-following is how anything can meet both sorts of conditions at once.

The subjective condition breaks down into at least three distinct sub-conditions and it will be useful to distinguish these, if we are to get the measure of the problem on hand. The first sub-condition is that the rule must be determinable or identifiable by a finite subject, in particular that it must be determinable or identifiable independently of any particular application. If the subject is to try to conform, then there must be something presented to it to which it can address its efforts. And if the subject is to be in a position to try to conform in any instance, then the rule to which it is to try and conform must be presented independently of how the rule applies in that instance. Allow that the rule is partly identified as requiring such and such an option in this situation, and it makes no sense to think of the subject trying to be faithful to it in that situation.

The second sub-condition that a rule must satisfy if it is to engage with the intentional projects of a creature like one of us is that it should not only be identifiable as a target of conformity for a finite subject, it should also be capable of instructing the subject, so to speak, on what it requires in the different instances where the subject tries to conform. This means that the rule must be directly readable, in the sense that the finite subject can tell straight off what it seems to require – this is the case with basic rules – or can tell what it requires by the application of rules whose apparent requirements it can ultimately tell straight off: this is the case with non-basic rules. Unless a finite subject can read off the requirements of a rule in this way, then it is not in a position to try to conform.

The third sub-condition complements the second. Where the second says that a rule must be readable by a finite subject, the third says that it can only be fallibly readable. No matter how directly the rule speaks to the subject, no matter how quickly the subject can tell what the rule seems to require, that fact alone cannot provide an epistemic guarantee that it has got the requirement of the rule right. The subject must not be an infallible authority, in at least one sense of that phrase. It may be in a position to know what a rule requires in a given situation. It may even be

in a position to know that it will get the rule right in that situation. Whether these claims are allowed will depend on how precisely the limits of knowledge are drawn. But no matter how knowledge is understood, the subject cannot be in a position to rule out altogether the possibility that it might get a rule wrong; the subject cannot know it for a fact that error is impossible in its reading of a rule. Otherwise it would make no sense for us to think of the subject as trying to get the rule right.

To return then to the problem of rule-following, the challenge is to identify something that can simultaneously satisfy the objective condition of being a normative constraint that is relevant in an indefinite variety of situations and the subjective conditions of being independently identifiable, directly readable and fallibly readable. There are two ways in which we might think of meeting this challenge: by taking something which we know to satisfy the objective condition and then showing how it can also satisfy the subjective constraints; or by taking something which certainly satisfies the subjective constraints and then showing how it can also satisfy the objective condition. But both paths look to be blocked and that is the essence of the rule-following problem. In setting out the problem I follow Saul Kripke (1982), adopting roughly the same presentation as in Pettit (1990a).

Take the sorts of entities which we know to satisfy the objective condition: the rule-in-extension and the rule-in-intension. The rule-in-extension is not capable of satisfying the subjective conditions, because it is liable to be an infinitely large set. There is no way that I could get in touch appropriately with such an infinite object. There is no way that I could get in touch with the infinite extension of a property across actual and possible worlds – say, the extension associated with boxes or triangles or games – as I try to be faithful to the appropriate rule in descriptive classification. And, to take the sort of rule discussed by Kripke, there is no way I could get in touch with the rule-in-extension associated with the plus-function: the rule determining what number is the referent of “ $x + y$ ”, for any two numbers  $x$  and  $y$ . “The infinitely many cases of the table are not in my mind

for my future self to consult” (Kripke, 1982, p. 22).

What of the rule-in-intension? What, for example, of the addition function, as Frege would conceive of it, which determines the correct option in any judgement about the sum of two numbers? What is there against the idea that this abstract object might be able to satisfy the subjective conditions, engaging a finite mind appropriately? Here the problem is to explain how a creature like one of us is able to get in contact with such an abstract object. It does not affect our senses like a physical object and so we are not causally connected with it in the ordinary way. So how then does it become present to such a creature?

Moving from the entities which can clearly satisfy the objective condition on a rule to entities that look more likely to be able to satisfy the subjective conditions, the question here is whether such entities can be objectively satisfactory: whether they can serve as normative constraints over an indefinite variety of cases. Kripke mentions two main candidates for entities of this kind: first, actual or possible examples of the application of the rule in question, such as examples of a property or examples of addition; and secondly, introspectible states of consciousness, as for instance a suitable idea or feeling. But there is an objection that applies to all such candidates, so Kripke argues, and indeed to any finite object that is proposed for the role in question. The objection, and this is clearly derived from Wittgensteinian materials, is that no finite object can unambiguously identify a constraint that is normative over an indefinite variety of cases. Consider a series of examples of addition:  $1 + 1 = 2$ ,  $1 + 2 = 3$ ,  $2 + 2 = 4$ , and the like. Or consider any set of examples of boxes or triangles or games. For all that any such finite object can determine, the right way to go with a novel case remains open. “Plus”, as we understand it, forces us to say that  $68 + 57 = 125$  but the examples given do nothing to identify the plus-rule as distinct from, say, the *quus*-rule, where this says that the answer in the case of 68 and 57 is 5. “Is a triangle”, as we understand it, forces us to say that a square page, diagonally folded, is a triangle but the examples given, if they do



not include this case, will be consistent with the folded page's not being a triangle; perhaps the rule illustrated outlaws paper triangles or perhaps it outlaws triangles made by folding. The fact is that any finite set of examples, mathematical or otherwise, can be extrapolated in an infinite number of ways; equivalently, any finite set of examples instantiates an infinite number of rules.

The upshot of these considerations is that rules do not appear to be the sorts of things that our finite minds can identify as items to follow; or, looking at the matter from the other way around, that among the items that our finite minds can suitably identify there appears to be nothing that could put us in touch with rules. Rule-following is a mysterious activity. It is central to human life and thought but its very possibility is philosophically problematic.

The rule-following problem is an important challenge for philosophers, in particular for philosophers of a naturalistic bent. What in the world – what in the natural world – does rule-following involve? Perhaps the only widely agreed point is that it certainly involves the development of an extrapolative disposition, a disposition generated by some examples to apply the rule after a certain pattern in new cases. But such a disposition is not enough on its own to constitute rule-following. While it provides a mechanism for prompting responses, it does not provide something which might tell us how to go on in new instances, something from which we might intentionally take our guidance (Kripke, 1982, p. 24).

Perhaps the best hope of a naturalistic solution is not to try to reduce rule-following to the operation of such a disposition but to give an account, using the disposition, of how a subject can identify a rule to follow. Under the account favoured by the present author, for example, the extrapolative disposition serves a second role over and beyond that of prompting responses in new cases: it enables certain applications to exemplify the rule and to present it as something that the subject can try to follow; although the applications given as examples will instantiate an indefinite number of rules, as we noted above, the

extrapolative disposition may ensure that they exemplify only one. (Pettit, 1990a and b, 1992). Future discussions will probably centre on such attempts to make naturalistic sense of the rule-following phenomenon.

*See also* OSTENSIVE DEFINITION; PRIVATE LANGUAGE ARGUMENT; WITTGENSTEIN.

#### BIBLIOGRAPHY

- Boghossian, P.: "The Rule-following Considerations," *Mind* 98 (1989), 504–50.  
 Fogelin, R.: *Wittgenstein* (London: Routledge, 1987).  
 Holtzman, S. H. and Leich, C. M., eds: *Wittgenstein: To Follow a Rule* (London: Routledge and Kegan Paul, 1981).  
 Kripke, S. A.: *Wittgenstein on Rules and Private Language* (Oxford: Blackwell, 1982).  
 Pettit, P.: "The Reality of Rule-following," *Mind* 99 (1990(a)), 1–21.  
 Pettit, P.: "Affirming the Reality of Rule-following," *Mind* 99 (1990(b)), 433–9.  
 Pettit, P.: *The Common Mind* (New York: Oxford University Press, 1992).  
 Wittgenstein, L.: *Philosophical Investigations*, trans. G. E. M. Anscombe (Oxford: Blackwell, 1953).  
 Wittgenstein, L.: *Remarks on the Foundations of Mathematics* (Oxford: Blackwell, 1956).  
 Wright, C.: *Wittgenstein on the Foundations of Mathematics* (Cambridge, MA: Harvard University Press, 1980).  
 Wright, C., ed.: *Synthese: Special Issue on Rule-following* 58 (1984).

PHILIP PETTIT

#### problems of induction

##### THE HUMEAN PROBLEM OF INDUCTION

Suppose that there is some property A pertaining to an observational or experimental situation, and that out of a large number of observed instances of A, some fraction  $m/n$  (possibly equal to 1) have also been instances of some logically independent property B. Suppose further that the background circumstances not specified in these descriptions have been varied to a substantial degree and also that there is no collateral information

available concerning the frequency of B's among A's or concerning causal or nomological connections between instances of A and instances of B.

In this situation, an *enumerative* or *instantial* inductive inference would move from the premise that  $m/n$  of observed A's are B's to the conclusion that approximately  $m/n$  of all A's are B's. (The usual probability qualification will be assumed to apply to the inference, rather than being part of the conclusion.) Here the class of A's should be taken to include not only unobserved A's and future A's, but also possible or hypothetical A's. (An alternative conclusion would concern the probability or likelihood of the very next observed A being a B.)

The traditional or Humean problem of induction, often referred to simply as *the* problem of induction, is the problem of whether and why inferences that fit this schema should be considered rationally acceptable or justified from an epistemic or cognitive standpoint, i.e. whether and why reasoning in this way is likely to lead to true claims about the world. Is there any sort of argument or rationale that can be offered for thinking that conclusions reached in this way are likely to be true if the corresponding premiss is true – or even that their chances of truth are significantly enhanced?

Hume's discussion of this issue (see HUME) deals explicitly only with cases where *all* observed A's are B's and where A is claimed to be the cause of B, but his argument applies just as well to the more general case. His conclusion is entirely negative and sceptical: inductive inferences are not rationally justified, but are instead the result of an essentially a-rational process, custom or habit. Hume challenges the proponent of induction to supply a cogent line of reasoning that leads from an inductive premise to the corresponding conclusion and offers an extremely influential argument in the form of a dilemma (sometimes referred to as "Hume's fork") to show that there can be no such reasoning. Such reasoning would, he argues, have to be either a priori demonstrative reasoning concerning relations of ideas or "experimental" (i.e. empirical) reasoning

concerning matters of fact or existence. It cannot be the former, because all demonstrative reasoning relies on the avoidance of contradiction, and it is not a contradiction to suppose that "the course of nature may change", that an order that was observed in the past will not continue in the future; but it also cannot be the latter, since any empirical argument would appeal to the success of such reasoning in previous experience, and the justifiability of generalizing from previous experience is precisely what is at issue – so that any such appeal would be question-begging. Hence, Hume concludes, there can be no such reasoning (1748, pp. 35–6).

An alternative version of the problem may be obtained by formulating it with reference to the so-called *Principle of Induction*, which says roughly that the future will resemble the past or, somewhat better, that unobserved cases will resemble observed cases. An inductive argument may be viewed as enthymematic, with this principle serving as a suppressed premiss, in which case the issue is obviously how such a premiss can be justified. Hume's argument is then that no such justification is possible: the principle cannot be justified a priori because it is not contradictory to deny it; and it cannot be justified by appeal to its having been true in previous experience without obviously begging the question.

The predominant recent responses to the problem of induction, at least in the analytic tradition, in effect accept the main conclusion of Hume's argument, viz. that inductive inferences cannot be justified in the sense of showing that the conclusion of such an inference is likely to be true if the premise is true, and thus attempt to find some other sort of justification for induction. Such responses fall into two main categories: (1) pragmatic justifications or "vindications" of induction, mainly developed by Reichenbach (see REICHENBACH); and (2) ordinary language justifications of induction, whose most important proponent is Strawson (see STRAWSON). In contrast, some philosophers still attempt to reject Hume's dilemma by arguing either (3) that, contrary to appearances, induction can be inductively justified without vicious circularity, or (4) that an a

priori justification of induction is possible after all. We will look briefly at all four of these responses.

(1) Reichenbach's view (see REICHENBACH) is that induction is best regarded, not as a form of inference, but rather as a *method* for arriving at *posits* regarding, e.g., the proportion of A's that are also B's. Such a posit is not a claim asserted to be true, but is instead an intellectual wager analogous to a bet made by a gambler. Understood in this way, the inductive method says that one should posit that the observed proportion is, within some measure of approximation, the true proportion and then continually correct that initial posit as new information comes in.

The gambler's bet is normally an "appraised posit", i.e. he knows the chances or odds that the outcome on which he bets will actually occur. In contrast, the inductive bet is a "blind posit": we do not know the chances that it will succeed or even that success is possible. What we are gambling on when we make such a bet is the value of a certain proportion in the independent world, which Reichenbach construes as the limit of the observed proportion as the number of cases increases to infinity. But we have no way of knowing that there even is such a limit, no way of knowing that the proportion of A's that are also B's converges in the long run on some stable value rather than varying at random. And if we cannot know that this limit exists, then we obviously cannot know that we have any definite chance of finding it.

What we can know, according to Reichenbach, is that *if* there is a truth of this sort to be found, the inductive method will eventually find it. That this is so is an analytic consequence of Reichenbach's account of what it is for such a limit to exist. The only way that the inductive method of making an initial posit and then refining it in light of new observations can fail to eventually arrive at the true proportion is if the series of observed proportions never converges on any stable value, which means that there is no truth to be found concerning the proportion of A's that are B's. Thus induction is justified, not by showing that it will succeed or indeed that it has any definite likelihood of success,

but only by showing that it will succeed if success is possible. Reichenbach's claim is that no more than this can be established for any method, and hence that induction gives us our best chance for success, our best gamble in a situation where there is no alternative to gambling.

This pragmatic response to the problem of induction faces several serious problems. First, there are indefinitely many other "methods" for arriving at posits for which the same sort of defence can be given – methods which yield the same result as the inductive method in the long run but differ arbitrarily in the short run. Despite the efforts of Salmon and others, it is unclear that there is any satisfactory way to exclude such alternatives, in order to avoid the result that any arbitrarily chosen short-term posit is just as reasonable as the inductive posit. Second, even if there is a truth of the requisite sort to be found, the inductive method is only guaranteed to find it or even to come within any specifiable distance of it in the indefinitely long run. But any actual application of inductive results always takes place in the short run, making the relevance of the pragmatic justification to actual practice uncertain. Third, and most importantly, it needs to be emphasized that Reichenbach's response to the problem simply accepts the claim of the Humean sceptic that an inductive premiss never provides the slightest reason for thinking that the corresponding inductive conclusion is true. Reichenbach himself is quite candid on this point, but this does not alleviate the intuitive implausibility of saying that we have no more reason for thinking that our scientific and commonsense conclusions that result from induction are true than, to use Reichenbach's own analogy (1949, p. 482), a blind man wandering in the mountains who feels an apparent trail with his stick has for thinking that following it will lead him to safety.

An approach to induction that resembles Reichenbach's in claiming that particular inductive conclusions are posits or conjectures, rather than the conclusions of cogent inferences, is offered by Popper (see POPPER). But Popper's view is even more overtly sceptical: it amounts to saying that all that can ever

be said in favour of the truth of an inductive claim is that the claim has been tested and has not yet been shown to be false.

(2) The ordinary language response to the problem of induction has been advocated by many philosophers, but the discussion here will be restricted to Strawson's paradigmatic version (see STRAWSON). Strawson claims that the question whether induction is justified or reasonable makes sense only if it tacitly involves the demand that inductive reasoning meet the standards appropriate to deductive reasoning, i.e. that the inductive conclusion be shown to follow deductively from the inductive premiss. Such a demand cannot of course be met, but only because it is illegitimate: inductive and deductive reasoning are simply fundamentally different kinds of reasoning, each possessing its own autonomous standards, and there is no reason to demand or expect that one of these kinds meet the standards of the other. Whereas if induction is assessed by inductive standards, the only ones that are appropriate, then it is obviously justified.

The problem here is to understand what this allegedly obvious justification of induction amounts to. In his main discussion of the point (1952, pp. 256–7), Strawson claims that it is an analytic truth that it is reasonable to believe a conclusion for which there is strong evidence and also an analytic truth that inductive evidence of the sort captured by the schema presented earlier constitutes strong evidence for the corresponding inductive conclusion, thus apparently yielding the *analytic* conclusion that it is reasonable to believe a conclusion for which there is inductive evidence. But he also admits, indeed insists, that the claim that inductive conclusions will be true in the future is contingent, empirical, and may turn out to be false (1952, p. 261). Thus the notion of reasonable belief and the correlative notion of strong evidence must apparently be understood in ways that have nothing to do with likelihood of truth, presumably by appeal to the standards of reasonableness and strength of evidence that are accepted by the community and are embodied in ordinary usage.

Understood in this way, Strawson's response to the problem of induction does not

speak to the central issue raised by Humean scepticism: the issue of whether the conclusions of inductive arguments are likely to be true when the corresponding premises are true. It amounts to saying merely that if we reason in this way, we can correctly call ourselves "reasonable" and our evidence "strong", according to our accepted community standards. But to the underlying issue of whether following these standards is a good way to find the truth, the ordinary language response appears to have nothing to say.

(3) The main attempts to show that induction can be justified inductively have concentrated on showing how such a defense can avoid circularity. Skyrms (1975, pp. 30–6) formulates perhaps the clearest version of this general strategy. The basic idea is to distinguish different levels of inductive argument: a first level in which induction is applied to things other than arguments; a second level in which it is applied to arguments at the first level, arguing that they have been observed to succeed so far and hence are likely to succeed in general; a third level in which it is applied in the same way to arguments at the second level; and so on. Circularity is allegedly avoided by treating each of these levels as autonomous and justifying the arguments at each level by appeal to an argument at the next higher level.

One problem with this sort of move is that even if circularity is avoided, it seems clear that the movement to higher and higher levels will eventually fail simply for lack of evidence: a level will be reached at which there have not been enough successful inductive arguments to provide a basis for an inductive justification at the next higher level. And if this is so, then the whole series of justifications collapses. A more fundamental difficulty is that the epistemological significance of the distinction between levels is obscure. If the issue is whether reasoning in accord with the original schema offered above ever provides a good reason for thinking that the conclusion is likely to be true, then it still seems question-begging, even if not flatly circular, to answer this question by appeal to another argument of the same form.

(4) The idea that induction can be justified on a purely a priori basis is in one way the most natural response of all: it alone treats an inductive argument as an independently cogent piece of reasoning whose conclusion can be seen rationally to follow, albeit perhaps only with probability, from its premise. Such an approach has, however, only rarely been advocated (*see* Russell, 1912; BonJour, 1986), and is widely thought to be clearly and demonstrably hopeless.

Many of the reasons for this pessimistic view depend on general epistemological theses about the possibility or nature of a priori knowledge (*see* A PRIORI KNOWLEDGE in Part I). Thus if, as Quine (*see* QUINE) alleges, there is no a priori justification of any kind, then obviously an a priori justification for induction is ruled out. Or if, as more moderate empiricists have claimed, a priori knowledge must be analytic, then again an a priori justification for induction seems to be precluded, since the claim that if an inductive premise is true, then the conclusion is likely to be true does not fit the standard conceptions of analyticity (*see* ANALYTICITY). A consideration of these matters is beyond the scope of the present article.

There are, however, two more specific and quite influential reasons for thinking that an a priori approach is impossible that can be briefly considered here. First, there is the assumption, originating in Hume but since adopted by very many others, that an a priori defence of induction would have to involve "turning induction into deduction", i.e. showing, *per impossibile*, that the inductive conclusion follows deductively from the premise, so that it is a formal contradiction to accept the latter and deny the former. But it is unclear why an a priori approach need be committed to anything this strong. It would be enough if it could be argued that it is a priori unlikely that such a premiss is true and the corresponding conclusion false.

Second, Reichenbach defends his view that the pragmatic justification is the best that is possible by pointing out that a completely chaotic world in which there simply is no true conclusion to be found as to the proportion of A's that are B's is neither impossible nor unlikely from a purely a priori standpoint,

the suggestion being that therefore there can be no a priori reason for thinking that such a conclusion is true. But there is a subtle confusion lurking here: that a chaotic world is a priori neither impossible nor unlikely in the absence of any further evidence does not show that such a world is not a priori unlikely and a world containing a certain regularity a priori likely *in relation to* the occurrence of a long-run pattern of evidence in which a certain stable proportion of observed A's are B's – an occurrence, it might be claimed, that would be highly unlikely in a chaotic world (*see* BonJour, 1986).

#### GOODMAN'S "NEW RIDDLE OF INDUCTION"

Suppose that prior to some specified time *t* (perhaps the year 2000) we observe a large number of emeralds (property A) and find them all to be green (property B). We proceed to reason inductively and conclude that all emeralds are green. Goodman (*see* GOODMAN) points out, however, that we could have drawn a quite different conclusion from the same evidence. If we define the term "grue" to mean "green if examined before *t* and blue if examined after *t*", then all of our observed emeralds will also be grue, and a parallel inductive argument will yield the conclusion that all emeralds are grue, and hence that all those examined after the year 2000 will be blue. Presumably the first of these conclusions is genuinely supported by our observations and the second is not, but the problem is to say *why* this is so and to impose some further restriction upon inductive reasoning that will permit the first argument and exclude the second.

Goodman himself formulates the problem in terms of the notion of projectibility (*see* PROJECTION, PROJECTIBILITY): a generalization that receives genuine inductive support from observed instances is *projectible* on to unobserved cases. His suggestion is that projectibility is a matter of the history of the terms involved: a projectible generalization is one whose terms are well *entrenched*, where this means that they have been used frequently in



past generalizations of this sort. Thus it is because (and only because) the term “green” is better-entrenched than the term “grue” that the first inductive argument in the preceding paragraph is to be preferred to the second.

The obvious alternative suggestion is that “grue” and similar predicates do not correspond to genuine, purely qualitative properties in the way that “green” and “blue” do, and that this is why inductive arguments involving them are unacceptable. Goodman, however, claims to be unable to make clear sense of this suggestion, pointing out that the relations of formal definability are perfectly symmetrical: “grue” may be defined in terms of “green” and “blue”, but “green” can equally well be defined in terms of “grue” and “bleen” (blue if examined before  $t$ , and green if examined after  $t$ ).

See also GOODMAN; HUME.

#### BIBLIOGRAPHY

- BonJour, L.: “A Reconsideration of the Problem of Induction,” *Philosophical Topics* 14 (1986), 93–124.
- Goodman, N.: *Fact, Fiction, and Forecast* (Cambridge, MA: Harvard University Press, 1955).
- Hume, D.: *An Enquiry Concerning Human Understanding* (1748), ed. L. A. Selby-Bigge, revised by P. H. Niddich (Oxford: Oxford University Press, 1975).
- Popper, K.: “Conjectural Knowledge: My Solution to the Problem of Induction,” reprinted in his *Objective Knowledge* (Oxford: Clarendon Press, 1972).
- Reichenbach, H.: *Experience and Prediction* (Chicago, IL: University of Chicago Press, 1938), 339–63.
- Reichenbach, H.: *Theory of Probability* (Berkeley, CA: University of California Press, 1949), 469–82.
- Russell, B.: *The Problems of Philosophy* (Oxford: Oxford University Press, 1912), ch. 6.
- Salmon, W.: *The Foundations of Scientific Inference* (Pittsburgh, PA: Pittsburgh University Press, 1967).
- Skyrms, B.: *Choice and Chance: An Introduction to Inductive Logic*, 2nd edn (Encino, CA: Dickenson, 1975).
- Strawson, P. F.: *Introduction to Logical Theory* (London: Methuen, 1952), 248–63.

LAURENCE BONJOUR

**projection, projectibility** Projection from present to absent cases occurs in inductive, hypothetical and counterfactual reasoning. Goodman’s grue paradox reveals that valid projection depends not only on the constitution of the evidence class, but also on its characterization (see GOODMAN). If all known emeralds are green, we may infer:

H1: All emeralds are green

but not

H2: All emeralds are grue,

where something is grue if examined before some future time  $t$  and found to be green, or not so examined and blue. H1 is valid, for “green” is projectible. But even though all members of our evidence class are grue as well as green, H2 is invalid since “grue” is unprojectible. A valid ampliative inference must be supported by evidence, unviolated by counterevidence, unexhausted (else it would not be ampliative), and must be framed in terms of projectible predicates. For only if a predicate is projectible can it convey credibility from known to unknown cases. How to distinguish between projectible and unprojectible predicates is a critical question for epistemology.

See also PROBLEMS OF INDUCTION.

CATHERINE Z. ELGIN

**proof** A proof is a collection of considerations and reasonings that instill and sustain the conviction that some proposed theorem – the theorem proved – is not only true, but could not possibly be false. A perceptual observation may instill the conviction that the water is cold, but not that the water could not but be cold. But a proof that  $2 + 3 = 5$  must not only instill the conviction that it is true that  $2 + 3 = 5$ , but also that  $2 + 3$  could not be anything but 5.

No one has succeeded in replacing this largely psychological characterization of proofs by a more objective characterization. The representations or reconstructions of

proofs as mechanical and semiotical derivations in formal-logical systems all but completely fail to capture “proofs” as mathematicians are quite content to give them. For example, formal-logical derivations depend solely on the logical form of the considered propositions, whereas usually proofs depend in large measure on contents of propositions other than their logical form.

*See also* INTUITION AND DEDUCTION.

## BIBLIOGRAPHY

- Detlefsen, M., ed.: *Essays on Proof* (London: Routledge, 1991).  
 Tragesser, R. S.: “Three Little Noticed Aspects of Mathematical Proof,” in Detlefsen, 1991.

ROBERT S. TRAGESSER

**propositional knowledge** Propositional knowledge (PK) is the type of knowing whose instances are labelled by means of a phrase expressing some proposition, e.g. in English a phrase of the form “that  $h$ ”, where some complete declarative sentence is instantiable for “ $h$ ”.

Theories of PK differ over whether the proposition that  $h$  is involved in a more intimate fashion, such as serving as a way of picking out a propositional attitude required for knowing (e.g. believing that  $h$ , accepting that  $h$  or being sure that  $h$ ) (*see also* BELIEF; CERTAINTY; KNOWLEDGE AND BELIEF). For instance, the tripartite analysis of PK, sometimes called the traditional or standard analysis, treats PK as consisting in having a justified, true belief that  $h$  (*see* TRIPARTITE DEFINITION OF KNOWLEDGE). In contrast, we shall later consider theories that treat PK as the possession of specific abilities, capacities, or powers, and that view the proposition that  $h$  as needing to be expressed only in order to label a specific instance of PK.

Although most theories of PK purport to analyse it, philosophers disagree about the goal of a philosophical analysis. Theories of PK may differ over whether they aim to cover all species of PK and, if they do not have this goal, over whether they aim to reveal any unifying

link between the species that they investigate, e.g. empirical knowledge, and other species of knowing.

Very many accounts of PK have been inspired by the quest to add a fourth condition to the tripartite analysis so as to avoid Gettier-type counterexamples to it (*see* GETTIER PROBLEM), and by the resulting need to deal with more counterexamples provoked by these new analyses (*see* Shope, 1983, for a survey). Keith Lehrer (1965) originated a Gettier-type example that has been a fertile source of important variants. It is the case of Mr Nogot, who is in one's office and has provided some evidence,  $e$  in response to all of which one forms a justified belief that Mr Nogot is in the office and owns a Ford, thanks to which one arrives at the justified belief that  $h_1$ : “Someone in the office owns a Ford.” In the example,  $e$  consists of such things as Mr Nogot's presently showing one a certificate of Ford ownership while claiming to own a Ford and having been reliable in the past. Yet Mr Nogot has just now been shamming, and the only reason that it is true that  $h_1$  is because, unbeknown to oneself, a different person in the office owns a Ford.

Variants on this example continue to challenge efforts to analyse species of PK. For instance, Alan Goldman (1988) has proposed that when one has empirical knowledge that  $h$ , then the state of affairs (call it  $h^*$ ) expressed by the proposition that  $h$  figures prominently in an explanation of the occurrence of one's believing that  $h$ , where explanation is taken to involve one of a variety of probability relations concerning  $h^*$  and the belief state. But this account runs foul of a variant on the Nogot case akin to one that Lehrer (1979) has described. In Lehrer's variant, Mr Nogot has manifested a compulsion to trick people into justifiably believing truths yet falling short of knowledge by means of concocting Gettierized evidence for those truths. If we make the trickster's neurosis highly specific to the type of information contained in the proposition that  $h$ , we obtain a variant satisfying Goldman's requirement that the occurrence of  $h^*$  significantly raises the probability of one's believing that  $h$ . (Lehrer himself (1990, pp. 103–4) has criticized

Goldman by questioning whether, when one has ordinary perceptual knowledge that an object is present, the presence of the object is what explains one's believing it to be present.)

In grappling with Gettier-type examples, some analyses proscribe specific relations between falsehoods and the evidence or grounds that justify one's believing. A simple restriction of this type requires that one's reasoning to the belief that *h* does not crucially depend upon any false lemma (such as the false proposition that Mr Nogot is in the office and owns a Ford). However, Gettier-type examples have been constructed where one does not reason through any false belief (e.g. a variant of the Nogot case where one arrives at belief that *h*, by basing it upon a true existential generalization of one's evidence: "There is someone in the office who has provided evidence *e*"). In response to similar cases, Sosa (1991) has proposed that for PK the "basis" for the justification of one's belief that *h* must not involve one's being justified in believing or in "presupposing" any falsehood, even if one's reasoning to the belief does not employ that falsehood as a lemma. Alternatively, Roderick (1989) (*see* CHISHOLM) requires that if there is something that makes the proposition that *h* "evident" for one and yet makes something else that is false evident for one, then the proposition that *h* is implied by a conjunction of propositions, each of which is evident for one and is such that something that makes it evident for one makes no falsehood evident for one. (*See* Shope, 1983, for discussion of earlier, related analyses by Sosa and Chisholm.) Other types of analyses are concerned with the role of falsehoods within the justification of the proposition that *h* (versus the justification of one's believing that *h*). Such a theory may require that one's evidence bearing on this justification not already contain falsehoods. Or it may require that no falsehoods are involved at specified places in a special explanatory structure relating to the justification of the proposition that *h*. (*See* Shope, 1983, and forthcoming, for details.)

A frequently pursued line of research concerning a fourth condition of knowing seeks what is called a defeasibility (*see* DEFEASIBILITY) analysis of PK. Early versions character-

ized defeasibility by means of subjunctive conditionals of the form "If A were the case then B would be the case". But more recently the label has been applied to conditions about evidential or justificational relations that are not themselves characterized in terms of conditionals. Early versions of defeasibility theories advanced conditionals where A is a hypothetical situation concerning one's acquisition of a specified sort of epistemic status for specified propositions (e.g. one's acquiring justified belief in some further evidence or truths) and B concerns, for instance, the continued justified status of the proposition that *h* or of one's believing that *h*.

A unifying thread connecting the conditional and non-conditional approaches to defeasibility may lie in the following facts: (1) What is a reason for being in a propositional attitude is in part a consideration, instances of the thought of which have the power to affect relevant processes of propositional attitude formation; (2) Philosophers have often hoped to analyse power ascriptions by means of conditional statements; and (3) Arguments portraying evidential or justificational relations are abstractions from those processes of propositional attitude maintenance and formation that manifest rationality. So even when some circumstance, *R*, is a reason for believing or accepting that *h*, some other circumstance, *K*, may prevent an occasion from being present for a rational manifestation of the relevant power of the thought of *R*, and it will not be a good argument to base a conclusion that *h* on the premiss that *R* and *K* obtain. Whether *K* does play this interfering, "defeating", role will depend upon the total relevant situation.

Accordingly, one of the most sophisticated defeasibility accounts, which has been proposed by John Pollock (1986), requires that in order to know that *h*, one must believe that *h* on the basis of an argument whose force is not defeated in the above way, given the total set of circumstances described by all truths. More specifically, Pollock defines defeat as a situation where (1) one believes that *p* and it is logically possible for one to become justified in believing that *h* by believing that *p*, and (2) one actually has a further set of beliefs, *S*, logically

consistent with the proposition that  $h$ , such that it is not logically possible for one to become justified in believing that  $h$  by believing it on the basis of holding the set of beliefs which is the union of  $S$  with the belief that  $p$  (cf. Pollock, 1986, pp. 36, 38). Furthermore, Pollock requires for PK that the rational presumption in favour of one's believing that  $h$  created by one's believing that  $p$  is undefeated by the set of all truths, including considerations that one does not actually believe. Pollock offers no definition of what this requirement means. But he may intend roughly the following, where  $T$  is the set of all true propositions: (I) one believes that  $p$  and it is logically possible for one to become justified in believing that  $h$  by believing that  $p$ , and (II) there are logically possible situations in which one becomes justified in believing that  $h$  on the basis of having the belief that  $p$  and the beliefs in  $T$ . Thus, in the Nogot examples, since  $T$  includes the proposition that Mr Nogot does not own a Ford, one lacks knowledge because condition (II) is not satisfied.

But given such an interpretation, Pollock's account illustrates the fact that defeasibility theories typically have difficulty dealing with introspective knowledge of one's own beliefs. Suppose that some proposition, say, that  $f$ , is false, but one does not realize this and holds the belief that  $f$ . Condition (II) has no coherent application to one's introspective knowledge that  $h_2$ : "I believe that  $f$ ." At least, this is so if one's reason for believing that  $h_2$  includes the presence of the very condition of which one is aware, i.e. one's believing that  $f$ . It is incoherent to suppose that one retains the latter reason yet also believes the truth that *not-f*. This objection can be avoided, but at the cost of adopting what is a controversial view about introspective knowledge that  $h$ , namely, the view that one's belief that  $h$  is in such cases mediated by some mental state intervening between the mental state of which there is introspective knowledge and the belief that  $h$ , so that it is the mediating state rather than the introspected state that is included in one's reason for believing that  $h$ . In order to avoid adopting this controversial view, Paul Moser (1989) has proposed a disjunctive analysis of PK, which requires that

either one satisfies a defeasibility condition rather like Pollock's or else one believes that  $h$  by introspection. However, Moser leaves obscure exactly why beliefs arrived at by introspection count as knowledge (see GETTIER PROBLEM).

Early versions of defeasibility theories had difficulty allowing for the existence of evidence that is "merely misleading", as in the case where one does know that  $h_3$ , "Tom, Grabit stole a book from the library", thanks to having seen him steal it, yet where, unbeknown to oneself, "Tom's mother out of dementia has testified that Tom was far away from the library at the time of the theft". One's justifiably believing that she gave the testimony would destroy one's justification for believing that  $h_3$  if added by itself to one's present evidence.

At least some defeasibility theories cannot deal with the knowledge one has while dying that  $h_4$ : "In this life there is no time at which I believe that  $d$ ", where the proposition that  $d$  expresses the details regarding some erudite matter, e.g. the maximum number of blades of grass ever simultaneously growing on the earth. When it just so happens that it is true that  $d$ , defeasibility analyses typically consider the addition to one's dying thoughts of a belief that  $d$  in such a way as to improperly rule out actual knowledge that  $h_4$ . (See Shope, forthcoming, for further discussion.)

A quite different approach to knowledge, and one able to deal with some Gettier-type cases, involves developing some type of causal theory of PK (see CAUSAL THEORIES IN EPISTEMOLOGY). Such theories require that one or another specified relation holds that can be characterized by mention of some aspect of causation concerning one's belief that  $h$  (or one's acceptance of the proposition that  $h$ ) and its relation to state of affairs  $h^*$ , e.g.  $h^*$  causes the belief;  $h^*$  is causally sufficient for the belief;  $h^*$  and the belief have a common cause. Such simple versions of a causal theory are able to deal with the original Nogot case, since it involves no such causal relationship, but cannot explain why there is ignorance in the variants where Nogot is a neurotic trickster. Moreover, Fred Dretske and Berent Enç (1984) have pointed out that sometimes one

knows of  $x$  that it is  $\phi$  thanks to recognizing a feature merely correlated with the presence of  $\phi$ ness. Without endorsing a causal theory themselves, they suggest that it would need to be elaborated so as to allow that one's belief that  $x$  has  $\phi$  has been caused by a factor whose correlation with the presence of  $\phi$ ness has caused in oneself (e.g. by evolutionary adaption in one's ancestors) the disposition that one manifests in acquiring the belief in response to the correlated factor. Not only does this strain the unity of a causal theory by complicating it, but no causal theory without other shortcomings has been able to cover instances of a priori knowledge.

Causal theories of PK differ over whether they deviate from the tripartite analysis by dropping the requirement that one's believing (accepting) that  $h$  be justified. The same variation occurs regarding reliability theories (*see* RELIABILISM), which present the knower as reliable concerning the issue of whether or not  $h$ , in the sense that some of one's cognitive or epistemic states,  $\theta$ , are such that, given further characteristics of oneself – possibly including relations to factors external to one and of which one may not be aware (*see* EXTERNALISM/INTERNALISM) – it is nomologically necessary (or at least probable) that  $h$ . In some versions, the reliability is required to be “global” in so far as it must concern a nomological (probabilistic) relationship of states of type  $\theta$  to the acquisition of true beliefs about a wider range of issues than merely whether or not  $h$ . There is also controversy about how to delineate the limits of what constitutes a type of relevant personal state or characteristic. (For instance, in a case where Mr Nogot has not been shamming and one does know thereby that someone in the office owns a Ford, does  $\theta$  concern a way of forming beliefs about Ford owners in offices, or something broader, such as a way of forming beliefs about the properties of persons spatially close to one, or instead something narrower, such as a way of forming beliefs about Ford owners in offices partly upon the basis of their relevant testimony?)

One important variety of reliability theory is a conclusive reasons account, which includes a requirement that one's reasons for

believing that  $h$  be such that in one's circumstances, if  $h^*$  were not to occur then, e.g., one would not have the reasons one does for believing that  $h$ ; or, e.g., one would not believe that  $h$ . Roughly, the latter is demanded by theories that treat a knower as “tracking the truth”, theories which include the further demand that, roughly, if it were the case that  $h$ , then one would believe that  $h$ . A version of the tracking theory has been defended by Robert Nozick (1981), who adds that if what he calls a “method” has been used to arrive at the belief that  $h$ , then the antecedent clauses of the two conditionals that characterize tracking will need to include the hypothesis that one would employ the very same method.

But unless more conditions are added to Nozick's analysis, it will be too weak to explain why one lacks knowledge in a version of the last variant of the tricky Mr Nogot case described above, where we add the following details: (a) Mr Nogot's compulsion is not easily changed; (b) while in the office, Mr Nogot has no other easy trick of the relevant type to play on one; and (c) one arrives at one's belief that  $h_1$  not by reasoning through a false belief but by basing the belief that  $h_1$  upon a true existential generalization of one's evidence.

Nozick's analysis is also too strong to permit anyone ever to know that  $h_5$ ; “Some of my beliefs about beliefs might be otherwise, e.g., I might have rejected one of them”. If I know that  $h_5$ , then satisfaction of the antecedent of one of Nozick's conditionals would involve its being false that  $h_5$ , thereby thwarting satisfaction of the consequent's requirement that I not then believe that  $h_5$ . For the belief that  $h_5$  is itself one of my beliefs about beliefs (*see* Shope (1984) for further discussion).

Some philosophers think that the category of knowing for which true, justified believing (accepting) is a requirement constitutes only a species of PK construed as an even broader category. They have proposed various examples of PK that do not satisfy the belief and/or justification conditions of the tripartite analysis. Such cases are often recognized by analyses of PK in terms of powers, capacities, or abilities. For instance, Alan R. White (1982) treats PK as merely the ability to provide a



correct answer to a possible question. However, White may be equating "producing" knowledge in the sense of producing "the correct answer to a possible question" with "displaying" knowledge in the sense of manifesting knowledge (cf. White, 1982, pp. 119–20). The latter can be done even by very young children and some nonhuman animals independently of their being asked questions, understanding questions, or recognizing answers to questions. Indeed, an example that has been proposed as an instance of knowing that *h* without believing or accepting that *h* can be modified so as to illustrate this point. The example concerns an imaginary person who has no special training or information about horses or racing, but who in an experiment persistently and correctly picks the winner of upcoming horseraces. If the example is modified so that the hypothetical "seer" never picks winners but only muses over whether those horses might win, or only reports picturing their winning, this behaviour should be as much of a candidate for the person's manifesting knowledge that the horse in question will win as would the behaviour of picking it as a winner.

These considerations expose limitations in Edward Craig's analysis (1990) of the concept of knowing as the concept of a Le person's being a satisfactory informant in relation to an inquirer who wants to find out whether or not *h*. Craig realizes that counter-examples to his analysis appear to be constituted by knowers who are too recalcitrant to inform the inquirer, or too incapacitated to inform, or too discredited to be worth considering (as with the boy who cried "Wolf"). Craig admits that this might make preferable some alternative view of knowledge as a different state that helps to explain the presence of the state of being a suitable informant when the latter does obtain. I have proposed (Shope, forthcoming) such an alternative, which offers a recursive definition that concerns one's having the power to proceed in a way representing the state of affairs *h*\* and the capacity to have the thought of *h*\* causally involved in one's proceeding in way. When combined with a suitable analysis of representing, this theory of PK can be unified with a struc-

turally similar analysis of knowing how to do something.

*See also* EVIDENCE; GETTIER PROBLEM; KNOWLEDGE AND BELIEF.

## BIBLIOGRAPHY

- Chisholm, R. M.: *Theory of Knowledge* (Englewood Cliffs, NJ: Prentice-Hall, 1989).
- Craig, E.: *Knowledge and the State of Nature: An Essay in Conceptual Synthesis* (Oxford: Clarendon Press, 1990).
- Dretske, F. and Enc, B.: "Causal Theories of Knowledge," *Midwest Studies in Philosophy*, vol. IX, *Causation and Causal Theories*, ed. P. A. French, T. E. Uehling, Jr and H. K. Wettstein (Minneapolis, MN: University of Minnesota Press, 1984), 517–28.
- Goldman, A.: *Empirical Knowledge* (Berkeley and Los Angeles, CA: University of California Press, 1988).
- Goldman, A.: *Epistemology and Cognition* (Cambridge, MA: Harvard University Press, 1986).
- Lehrer, K.: "Knowledge, Truth, and Evidence," *Analysis*, 25 (1965), 168–75; reprinted in *Knowing: Essays in the Analysis of Knowledge*, ed. M. D. Roth and L. Galis (New York: Random House, 1970), 55–66.
- Lehrer, K.: "The Gettier Problem and the Analysis of Knowledge," *Justification and Knowledge*, ed. G. Pappas (Dordrecht, Boston, MA, and London: Reidel, 1979), 65–78.
- Lehrer, K.: *Theory of Knowledge* (Boulder, Colo., and San Francisco, CA: Westview, 1990).
- Moser, P.: *Knowledge and Evidence* (Cambridge: Cambridge University Press, 1989).
- Nozick, R.: *Philosophical Explanation* (Cambridge, MA: Harvard University Press, 1981).
- Pollock, J. L.: *Contemporary Theories of Knowledge* (Totowa, NJ: Rowman and Littlefield, 1986).
- Shope, R. K.: *The Analysis of Knowing* (Princeton, NJ: Princeton University Press, 1983).
- Shope, R. K.: "Cognitive Abilities, Conditionals, and Knowledge: A Response to Nozick," *The Journal of Philosophy*, 81 (1984), 29–48.
- Shope, R. K.: *Knowledge as Power: An Essay in Epistemological Metaphysics with an Application to Ethical Theory* (forthcoming).
- Sosa, E.: *Knowledge in Perspective: Selected Essays in Epistemology* (Cambridge: Cambridge University Press, 1991).
- White, A. R.: *The Nature of Knowledge* (Totowa, NJ: Rowman and Littlefield, 1982).

ROBERT K. SHOPE

**protocol sentences** “Protocol sentences” or “protocol statements” were defined by Carnap (*see* CARNAP) in *The Unity of Science* (1932) as belonging to the “direct record of a scientist’s . . . experience”. “A ‘primitive’ protocol . . . exclude[s] all statements obtained indirectly by induction or otherwise and postulates therefore a sharp (theoretical) distinction between the raw material of scientific investigation and its organisation.” The problem for the logical positivists (*see* LOGICAL POSITIVISM) was how such records of “private experience” could serve as foundational (*see* FOUNDATIONALISM) for the public language of a unified science; in effect, how their empiricism (*see* EMPIRICISM) and their physicalism could be reconciled. Carnap himself was undecided whether protocol sentences should take the form of phenomenalistic sense-datum statements (e.g. most crudely “Joy now”, “Here now blue”) (*see* PHENOMENALISM; SENSE-DATA), or whether they should be more like ordinary reports of observation (e.g. “A red cube is on the table”); he hoped to escape metaphysical commitment by his use of the distinction between “formal” and “material” modes of speech (*see* CARNAP). Neurath insisted that protocol sentences should make explicit the identity and location of the speaker, and the time of utterance; but later recognized that there is no reason to regard such sentences as foundational (*see* NEURATH; SCHLICK). Carnap later argued that which statements should serve as protocols was “a matter of decision” (i.e. convention); one may choose one’s own observation-statements, but that is only because “intersubjective testing of statements about observations (brain-processes) is relatively inconvenient . . .”.

ANDY HAMILTON

**psychologism** With respect to a given subject-matter psychologism is the theory that the subject-matter in question can be reduced to, or explained in terms of, psychological phenomena: mental acts, events, states, dispositions and the like. Psychologism was widespread in both Britain and Germany during the latter half of the nineteenth century. It fell into disrepute at the beginning of the

present century, however, largely as a result of the objections raised against it by two philosophers, Edmund Husserl (*see* HUSSERL) and Gottlob Frege (*see* FREGE) (*see* Husserl, 1970, *passim*; Frege, 1974, pp. i-xi). They attacked psychologism in logic and the foundations of mathematics by showing that a psychologistic theory could never account for the objectivity, the necessity, or the universality of the truths belonging to these disciplines. Ludwig Wittgenstein (*see* WITTGENSTEIN) continued this antipsychologistic tradition by arguing that even meaning and understanding should not be construed as species of mental act (*see* Wittgenstein, 1953, §§329, 541 and pp. 175–6).

#### BIBLIOGRAPHY

- Frege, G.: *The Foundations of Arithmetic*, trans. J. L. Austin (Oxford: Blackwell, 1974).  
 Husserl, E.: *Logical Investigations*, trans. J. N. Findlay, vol. 1, *Prolegomena to Pure Logic* (London: Routledge and Kegan Paul, 1970).  
 Wittgenstein, L.: *Philosophical Investigations*, trans. G. E. M. Anscombe (Oxford: Blackwell, 1953).

DAVID BELL

**psychology and epistemology** There has been a steady stream of two-way traffic between epistemology and psychology. Philosophers and psychologists have relied on novel epistemological doctrines and arguments to support psychological views; more recently, epistemologists have been drawn to psychology in an attempt to solve their own problems.

Many epistemological disagreements within psychology pertain in some way or other to disputes about behaviourism (*see* BEHAVIOURISM). Beginning in 1908, John Watson argued forcefully that the science of psychology had not advanced very far. His proposed remedy was to transform the discipline into a natural science by changing the subject-matter: behaviour, rather than consciousness or the mind, was to be the proper object of study by psychologists. In fact, all reference to the mental, he argued, should be excluded from

psychology. Sometimes, he supported this recommendation by arguing that the mental does not exist, or, more weakly, that there is no evidence of its existence. At other times, he appealed to what was later called “methodological behaviourism”, the view that all mentalistic explanations be rejected for methodological reasons. In defending this latter view, he and his behaviourist successors developed a series of distinctive epistemological arguments.

One of these arguments, used by B. F. Skinner, runs like this: We can avoid reference to mentalistic causes by going directly to the prior physical causes while bypassing intermediate feelings or other mental events. If all linkages are lawful, Skinner argues, nothing is lost by neglecting a mentalistic link. In saying that nothing is lost, Skinner presumably means nothing that would hamper our predicting or controlling behaviour. Suppose, however, that one of our goals is to *explain* someone's behaviour. Something significant, then, would be lost if a mental event caused, or partially caused, the behaviour in question. By neglecting that cause, we would be guaranteeing that our proposed explanation of the person's behaviour is incorrect. Some behaviourists are not convinced by this reply because they deny that explaining phenomena is a legitimate goal of science. This denial may be buttressed by saying that explanation is valuable only as a guide to understanding, but understanding is something mentalistic and hence should not be sought in a scientific psychology. To resort to this defence, however, is to assume what is at issue: that the behaviourist view of what constitutes a scientific psychology is correct. A behaviourist may have other reasons for denying that scientists should try to explain psychological phenomena, but if they are also unsatisfactory, then Skinner's argument is likely to fail.

A second argument used by Skinner concerns the failure to explain cognitive or other mentalistic causes. A disturbance in behaviour, he contends, is not explained by relating it to anxiety until the anxiety in turn is explained. The postulation of mentalistic events, however, discourages the tracing the causal sequence; investigators, Skinner

claims, simply stop with the anxiety and fail to ask what caused *it*. One problem with this argument is the assumption that non-behaviourists generally end inquiry with the postulation of a mentalistic cause. Some do and some do not. For example, a Freudian psychologist who explains a client's behaviour in terms of anxiety might well go on and try to explain the anxiety in terms of a repressed wish. A more serious problem, however, lies with Skinner's initial assumption that even if anxiety causes a certain behaviour, citing it is not explanatory unless it is in turn explained. It is not generally true that postulating B as the cause of A explains A's occurrence only if, and when, B's occurrence is also explained. If B had to be explained in terms of C, and C in terms of D, and so on *ad infinitum*, then all causal explanation would be impossible. It could be replied that we have a satisfactory explanation once B's occurrence is explained; there is no need to go any further. However, stopping at precisely that point seems arbitrary. Why does explanation require citing the cause of the cause of a phenomenon but not the next link in the chain of causes? Perhaps what is not generally true of explanation is true only of mentalistic explanation; only in giving the latter type are we obligated to give the cause of a cause. However, this too seems arbitrary. What is the difference between mentalistic and non-mentalistic explanation that would justify imposing more stringent restrictions on the former?

The epistemological argument most widely used by behaviourists turns on the alleged unobservability of mental events or states. If cognitions are unobservable in principle, the argument runs, we have no warrant for believing that they exist and, hence, no warrant for accepting cognitive explanations. The same argument applies to non-cognitive mental states, such as sensations or emotions. Opponents of behaviourism sometimes reply that mental states can be observed: Each of us, through introspection (*see* INTROSPECTION), can observe at least some mental states, namely our own (at least those of which we are conscious). To this point, behaviourists have made several replies. Some (e.g. Zuriff, 1985) argue that introspection is

too unreliable for introspective reports to qualify as firm scientific evidence. Others have replied that, even if introspective reports were reliable, what we introspect is private and that this fact alone renders introspective data unsuitable as evidence in a science of behaviour. A more radical reply, advanced by certain philosophers, is that introspection is not a form of observation, but rather a kind of theorizing. More precisely, when we report on the basis of introspection that we have a painful sensation, a thought, a mental image, etc., we are theorizing about what is present. The resulting theory may or may not be correct, but, on this view, the fact that we introspect does not show that any mental states are observable (see PERCEPTUAL KNOWLEDGE; PRIVATE LANGUAGE ARGUMENT).

Whether or not introspection is reliable or is a form of observation, there remains the problem of determining the content of, or even the existence of, other people's minds. The traditional solution to this problem, the argument from analogy (see ARGUMENT FROM ANALOGY), is now generally thought to be a failure. An inference to the best explanation (see INFERENCE TO THE BEST EXPLANATION) might serve instead, but some philosophers (e.g. Van Fraassen, 1980) have challenged the validity of this mode of inference. If that challenge can be sustained, then behaviourists may be able to defend at least one of the central planks in their programme: the rejection of all mentalistic explanations.

Many cognitivists and behaviourists agree about the desirability of making psychology a rigorous, natural science; what they often disagree about is what that requires. Some philosophers and psychologists, however, have serious reservations about rendering psychology "scientific", at least if that means employing the same sort of epistemological standards as are used in biology, chemistry and physics. These latter sciences, some in the Hermeneutic tradition argue, are primarily causal sciences, i.e. they involve both the search for causal laws and the widespread use of causal explanations. In psychology, however, we should be explaining human actions, some argue, in non-causal terms, perhaps in terms of meaning or motives (see

REASONS/CAUSES). Karl Jaspers, the German philosopher-psychiatrist, appears to be an early exponent of this view. In the 1922 edition of his *General Psychopathology* (1963, p. 539), he criticizes Freud as follows: "The falseness of the Freudian claim lies in the mistaking of meaningful connections for causal connections." Earlier in the same work, he discusses the sort of psychology that would replace the Freudian sort: a science of meaningful connections.

Some contemporary Hermeneuticians (e.g. Ricoeur, 1981) agree with Jaspers about meaning connections, but unlike him they defend Freudian theory. They take the view that the theory really is about meanings rather than causes, but that Freud sometimes misinterpreted his own theory. Properly interpreted, at least some central parts of the theory, they argue, are true. Freud was right, moreover, in denying the need for experimental confirmation of his various hypotheses. Meeting such a requirement, some Hermeneuticians argue, is reasonable only if we use the same evidential standards as physics or chemistry, but doing that is inappropriate. Other Hermeneuticians (e.g. Taylor, 1971) would not restrict their epistemological point about standards to Freudian theory; more generally, any psychological theory that purports to explain human actions, they argue, is to be evaluated by criteria not applicable to theories in the natural sciences (see HERMENEUTICS).

Some who reject the hermeneutical viewpoint argue that meanings, in so far as they are explanatory, are causes. For example, suppose that I vote for a certain candidate and the meaning of my vote is that I am taking a stand against abortion. That is, I see myself as objecting to the views of the rival candidate who favors abortion. Did my seeing what I am doing in this way make any difference to how I voted? If it did, it is argued, then the meaning of my action was a contributing cause; if it did not, then talking about the meaning of my action does not explain why it occurred. The same sort of point is often made about motives (Grünbaum, 1984). I may have been motivated not to promote Tom because I do not trust him, but that may



have made no difference to my promoting someone else instead. If it did make a difference, citing my motive helps explain my action; but, in that case, the motive is also a cause. The general suggestion, then, is this: where meanings and motives make no difference to what a human being does, then citing them is not explanatory; where they do make a difference, they are causes. A problem, then, for some Hermeneutical views is to explain how psychologists can explain human actions while simultaneously rejecting all causal explanations of such actions.

The idea that psychology requires different evidential standards from the natural sciences (and hence is not itself a natural science) need not be tied to the view that its explanations should be non-causal. Some who promote the first idea reject the second; they agree that motives and meanings can be causes, but still argue for the use of standards other than those of natural science. Those who argue for this thesis, however, must face the following question: What are these different, and presumably lower, evidential standards that are to be employed in psychology? Karl Jaspers gave an answer to this question (1963, p. 303). In assessing claims of meaning connections, he argued, we must be willing to accept what is, on his view, self-evident. As an example of self-evidence, he cites Nietzsche's claim of a general connection between awareness of one's wretchedness and the development of a slave morality. Many later Hermeneuticians have also appealed to the idea of self-evidence or, what is very similar, the idea of what is intuitively obvious. It is difficult to see, however, how either can be an acceptable standard for any area in psychology in which rival hypotheses are relatively equal in plausibility given our current evidence. In fact, even where we can think of only one hypothesis that appears self-evident we may still have no rational grounds for believing it. At one time, it seemed self-evident to most observers that some people acted strangely because they were possessed by the devil; yet, that hypothesis may have had no evidential support at all. Of course, one can draw a distinction between hypotheses that only appear to be

self-evident and those that truly are, but does this help if we are not given any way to tell the difference?

In the cases discussed so far, philosophers and psychologists have employed novel and controversial epistemological arguments to support their psychological views. In a related development, some philosophers have engaged in what might be termed "applied epistemology". That is, they have relied on epistemological assumptions *plus* empirical data to criticize or defend such items as Freudian theory (Grünbaum, 1984), the foundations of behaviour therapy (Erwin, 1978), behaviourism and various theories in cognitive psychology. What is mainly new in these applied studies is the greater attention paid to the empirical details of psychology compared to earlier attempts to rest psychological views primarily on abstract epistemological doctrines.

The appeal to empirical data is also evident in the movement to "naturalize" epistemology (see NATURALIZED EPISTEMOLOGY). Here the direction is reversed: psychology is used to support epistemology instead of the other way around. In one respect, this development is not new; earlier philosophers, such as David Hume (see HUME) and Immanuel Kant (see KANT), often tried to buttress their epistemologies by appeal to what they believed to be facts about the human mind. In the immediate post-World War II period, however, many epistemologists in English speaking countries placed great emphasis on conceptual analysis. The result was that epistemology was often practiced as if it were primarily, or solely, an *a priori* discipline. Relatively few attempts were made by epistemologists to make use of empirical studies in psychology (or, for that matter, in any other science).

One reason for naturalizing epistemology is that the traditional quest for a foundation of knowledge is said to have ended in failure. This appears to be the main reason given by Quine (see QUINE) in his classic paper "Epistemology Naturalized".

To this point, traditional epistemologists have made several replies. First, it is argued that there are new variants of "foundationalism" (see FOUNDATIONALISM) and that one of these might turn out to be right. Second, coherentists



(see COHERENTISM) argue that they can explain how knowledge claims can be ultimately justified without invoking foundationalism of any sort. Third, it is argued that there is more to traditional epistemology than providing a foundation for knowledge. For example, epistemologists have been interested in analysing the concept of knowledge, developing theories of evidence and justification, and justifying non-demonstrative rules of inference. The pursuit of these projects might be warranted even if the traditional foundational problem cannot be solved. The alleged failure to solve that problem, however, is not the only reason cited by naturalists. Both in epistemology and in its sister discipline, the philosophy of science, there have been complaints about the lack of interesting, positive results. A related reason is that traditional epistemology has relied too heavily, it is argued, on a priori claims. Some naturalists argue that either there is no a priori knowledge at all (see A PRIORI KNOWLEDGE in Part I) or that there is no such knowledge of non-trivial propositions. To get firm, interesting positive results, as opposed merely to finding more and more counterexamples to false theories or generating more triviality, an epistemologist must, it is argued, appeal to empirical results of psychology and other sciences. Naturalists disagree among themselves, however, about the nature of this appeal.

One naturalist view, associated with Quine (1985), is that we should replace traditional epistemological questions with questions answerable by empirical studies in psychology. For example, he suggests that the traditional question about the foundations of knowledge be replaced by one about how sensory stimulations result in the storing of information. Some philosophers are likely to reply, however, that by substituting psychological questions for epistemological ones, we are not naturalizing epistemology; we are simply changing the subject.

A second view is that we should abandon a priori arguments altogether and restrict ourselves to appeals to empirical evidence in answering epistemological questions. The key issue here, itself partly empirical, is whether any interesting results will (could?) emerge

from this approach. A more modest view is that epistemologists should continue to use a priori arguments as before but, where possible, to appeal to empirical results as well. A possible example concerns the dispute between some experimental psychologists and their opponents about the epistemological value of clinical case studies. Some argue that the data from case studies generally have only heuristic rather than evidential value, although they may occasionally refute some psychological theory. To confirm causal hypotheses, it is argued, we generally need experimental evidence rather than evidence from case studies.

Other psychologists, however, contend that case studies can often confirm as well as disconfirm causal hypotheses. An epistemologist, in commenting on this dispute, might appeal partly to abstract, a priori considerations about the nature of evidence and confirmation, but might also have to appeal to empirical data about the presence or absence of competing, plausible alternatives to the hypothesis being tested. For example, it might turn out that in certain areas in psychology, case studies can be confirmatory because the hypotheses being tested often have no plausible competitors; in other areas, experimentation may generally be needed to adjudicate between plausible rivals.

The above example concerns an epistemological dispute within psychology. Whether empirical data from psychology are likely to be helpful in resolving issues within epistemology itself is still controversial. Nevertheless, recent work in epistemology does indicate a greater willingness among epistemologists, even among those not describing themselves as 'naturalists,' to at least consider empirical data from psychology to be relevant to their concerns.

See also NATURAL SCIENCE; REASONS/CAUSES; SOCIAL SCIENCES.

#### BIBLIOGRAPHY

Erwin, E.: *Behavior Therapy: Scientific Philosophical and Moral Foundations* (New York: Cambridge University Press, 1978).

- Goldman, A.: *Epistemology and Cognition* (Cambridge, MA: Harvard University Press, 1984).
- Grünbaum, A.: *The Foundations of Psychoanalysis: A Philosophical Critique* (Berkeley, CA: University of California Press, 1984).
- Jaspers, K.: *General Psychopathology* (Chicago, CA: University of Chicago Press, 1963).
- Quine, W. V.: "Epistemology Naturalized," in his *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969).
- Ricoeur, P.: *Hermeneutics and the Human Sciences*, trans. J. B. Thompson (New York: Cambridge University Press, 1981).
- Taylor, C.: "Interpretation and the Sciences of Man," *The Review of Metaphysics* 25 (1971), 1–51.
- Van Fraassen, B.: *The Scientific Image* (Oxford: Oxford University Press, 1980).
- Zuriff, G. E.: *Behaviorism: A Conceptual Reconstruction* (New York: Columbia University Press, 1985).

EDWARD ERWIN

**Putnam, Hilary (1926–)** One of the most influential philosophers of the latter half of the twentieth century, Putnam (Professor of Mathematical Logic at Harvard University) has discussed epistemology only late in his career. But he has recently made some major contributions to this field: his arguments that truth is ultimately an epistemic concept, i.e., that truth and "ideal rational acceptability" (or ideal justification) are *interdependent* concepts and his criticisms of radical or evidence-transcendent scepticism. The two themes are tied together by Putnam's defence of what he calls "internal realism".

In the early 1970s Putnam was a staunch realist with respect to science and language. In both the philosophies of language and science, he argued for realism (*see* REALISM) against verificationism (*see* VERIFICATIONISM) (e.g. 1978, parts I and 3), and for the priority of reference over meaning (i.e. that most terms refer directly without mediating ideas, senses, or properties; cf e.g. Putnam, 1975). Against antirealism, Putnam argued truth couldn't be warranted assertibility since we can recognize the possibility that a proposition might be warrantably assertible but false (1978, part 3). However, about 1976, Putnam's philosophy took a dramatic turn. He

abandoned his view that truth was essentially non-epistemic and instead argued that truth was *idealized* warranted assertibility (1978, part 4; 1981).

A key step in his development was Putnam's realization that, on his own view, evidence-transcendent scepticism was self-refuting. One naïvely assumes that if we were all brains in a vat (assuming we're not now), then the sentence "we are brains in a vat" would continue to *mean* just what it does now and so would be true in the vat case. But this is to assume meanings "are in the head", an assumption repudiated by Putnam's causal or direct theory of reference. In fact, in a world in which we were all brains in a vat, there would be a "referential shift" and the sentence "we are brains in a vat" would mean something very different. While "brains" now refers directly to certain organs, in the vat world "brains" would refer directly to aspects of the computer program, perhaps, controlling the subjects' brain-images; similarly for "vats". But then in that case, the brains' assertion "we are brains in a vat" would be *false*, since they are not aspects of a computer program. Consequently, the *utterance* "we are brains in a vat" is always false, regardless of who says it and so radical scepticism is refuted (Putnam, 1981). It is useful to compare Putnam's analysis to Skolem's paradox where, because of a similar referential shift, the sentence "the real numbers are uncountable" is true not only for us in an uncountable world but even *true in a countable model of set theory!* (Tymoczko, 1990).

The residual feeling that we could somehow step outside the world (or model) and discover that we really were brains in a vat is an expression of what Putnam calls "metaphysical realism" (Putnam, 1978, part 4; 1981). He points out that this view is, strictly speaking, false or incoherent. It is false if expressed in our language (since "we are brains in a vat" is false). Moreover, to the response "Yes, but if we could adopt a God's eye point of view we'd see we really *are* brains in a vat". Putnam could reply that this requires using a language *other* than the language we now use – in order to say "we really are brains in a vat" and make it come out true. In that case, the

quoted sentence is meaningless. (In analogous fashion, the sentence “the reals are countable” is either trivially false by Cantor’s Theorem, or else a sentence of a language that by hypothesis, we don’t understand.)

For Putnam, “metaphysical realism” is the view that even an epistemically ideal theory could be false (in other words, that truth is non-epistemic). His recent argument purports to show this view has no content. Nevertheless, if we abandon metaphysical realism, we should still hold to the internal or pragmatic realism suggested by Peirce (see PEIRCE), according to Putnam. Internal realism is realism about science and language, but only as an empirical theory internal to science. It is stronger than verificationism (because true beliefs are not justified beliefs but only *ideally* justified beliefs) and it still maintains the priority of reference over meaning, and in this sense is realist. On the other hand, reference is seen as dependent on use and on what can be ideally verified, and since truth is tied to reference, truth too is an epistemic concept. Crudely: the only criterion for what is a fact is what it is (ideally) rational to accept (and so bivalence might not be preserved since, for certain *p*, it might not be ideally rational either to accept *p* or to reject it). Thus, truth and justification are two separate, but interdependent notions. (Putnam 1981, 1989).

(Putnam’s argument can be formulated more abstractly in model theoretic terms, as in Putnam, 1983, 1981. For a concise summary of internal realism, see Putnam, 1989.)

See also ANALYTICITY; PRAGMATISM; REALISM; SCEPTICISM, CONTEMPORARY; TRUTH.

#### WRITINGS

“The Meaning of ‘Meaning,’” reprinted in his *Mind, Language and Reality: Philosophical Papers*, vol. 2 (Cambridge: Cambridge University Press, 1975).

*Meaning and the Moral Sciences* (London: Routledge and Kegan Paul, 1978), esp. part 3, “Reference and Understanding,” and part 4, “Realism and Reason.”

*Reason, Truth and History* (Cambridge: Cambridge University Press, 1981).

*Realism and Reason* (Cambridge: Cambridge University Press, 1983), esp. introduction and ch. 1. *The Many Faces of Realism* (LaSalle, IL: Open Court, 1987).

*Representation and Reality* (Cambridge, MA: MIT Press, 1989), ch. 7.

#### BIBLIOGRAPHY

Tymoczko, T.: “Brains Don’t Lie,” in *Doubting: Contemporary Perspectives on Scepticism*, ed. M. Roth and G. Ross (Dordrecht: Kluwer, 1990).

THOMAS TYMOCZKO

**Pyrrhonism** A sceptical school in ancient Greek philosophy, which traced its origins to Pyrrho of Elis (c.365–270 bc). Pyrrhonists together with Academic sceptics account for a strong tradition of scepticism in the Hellenistic period of Greek philosophy. Pyrrho himself wrote nothing, but through the satirical poetry of his pupil Timon of Phlius he became legendary for his disdainful attitude toward speculative philosophy and for the detachment and simplicity of his way of life. This anti-theoretical orientation, and the practical view of philosophy as a way of life leading to ataraxia (see ATARAXIA), became two important characteristics of Pyrrhonian scepticism. After the death of Timon, Pyrrho’s sceptical ideas seem not to have been revived until the first century bc, when they were adopted by a radical sceptic from the Academy under Philo of Larissa. Aenesidemus, dissatisfied with the scepticism of the Fourth Academy, which had become increasingly dogmatic and influenced by Stoic views (see STOIC EPISTEMOLOGY), broke away from it to form his own school of scepticism, claiming Pyrrho as its founder. Aenesidemus built upon the anti-speculative foundation of Pyrrhonian scepticism by formulating arguments against all dogmatic philosophies, while giving particular attention to refuting the Stoics. He also compiled all the sceptical arguments known to Greek philosophers and grouped them into “modes”, patterns of argument leading to suspension of judgment. The sceptical arguments produce oppositions of appearances equal in weight and credibility (see

ISOSTHENEIA). The Pyrrhonian sceptic finds no good reason to prefer this appearance over that, so he suspends judgement. He is able to say how things appear to him but not how things really are. It was no doubt due to the influence of Aenesidemus that Pyrrhonian scepticism evolved to produce a sophisticated philosophical methodology. Agrippa, a Pyrrhonist probably from the first century AD, also formulated a set of "modes" classifying different argument forms. Sextus Empiricus, who lived in the late second century AD, is the last major representative of Pyrrhonian scepticism in Greek philosophy. Besides being a sceptic himself Sextus is the most important source of our knowledge of ancient Pyrrhonism. After the time of Sextus, Pyrrhonian scepticism diminished in importance and finally died out in later antiquity. The writings of Sextus were rediscovered in the sixteenth century and were influential in shaping modern philosophical scepticism (Popkin, 1979).

*See also* SCEPTICISM; SEXTUS EMPIRICUS.

## BIBLIOGRAPHY

- Annas, J. and Barnes, J.: *The Modes of Scepticism* (Cambridge: Cambridge University Press, 1985).
- Burnyeat, M.: "Can the Sceptic Live His Scepticism?," in *Doubt and Dogmatism*, ed. M. Schofield, M. Burnyeat and J. Barnes (Oxford: Clarendon Press, 1980), 20–53.
- Flintoff, E.: "Pyrrho and India," *Phronesis* 25 (1980), 88–108.
- Popkin, R.: *The History of Skepticism from Erasmus to Spinoza* (Berkeley, CA: University of California Press, 1979).
- Sedley, D.: "The Protagonists," in *Doubt and Dogmatism*, ed. M. Schofield, M. Burnyeat and J. Barnes (Oxford: Clarendon Press, 1980), 1–19.
- Sedley, D.: "The Motivation of Greek Scepticism," in *The Sceptical Tradition*, ed. M. Burnyeat (Berkeley, CA: University of California Press, 1983), 9–29.
- Stough, C.: *Greek Scepticism* (Berkeley, CA: University of California Press, 1969).

CHARLOTTE STOUGH

# Q

---

## **Quine, Willard Van Orman (1908–2000)**

Born in Akron, Ohio, Quine's professional career was spent in the philosophy department at Harvard University. A visit to the Vienna Circle during the 1930s, when he encountered logical empiricists (see LOGICAL EMPIRICISM) such as Rudolf Carnap, shaped his work: although he challenged some of their most fundamental doctrines, he shares their commitment to empiricism and their view that philosophy should be pursued as part of science. His earliest work was in mathematical logic and the foundations of set theory. But several of the papers in *From a Logical Point of View* (1953) have had a major influence upon epistemology, and his major work, *Word and Object* (1960), together with many subsequent writings, have been instrumental in the development of naturalized epistemology (see NATURALIZED EPISTEMOLOGY).

"Two Dogmas of Empiricism", contained in the former book, identified and rejected two assumptions of traditional empiricism. The first of these was the distinction between analytic statements such as "All bachelors are male" which are true by virtue of meaning, and other synthetic truths which are grounded in "fact" (p. 20). The second was "radical reductionism": "every meaningful statement is held to be translatable into a statement (true or false) about immediate experience" (p. 38). Quine held that the analytic/synthetic distinction could be sustained only if radical reductionism was true. The latter doctrine gains support from a verificationist theory of meaning: by translating each statement into a statement about immediate experience, we show its true meaning by showing what experiences would be required to verify it. And the analytic statements are those which are "vacuously confirmed . . . come what may" (p. 41) (see ANALYTICITY).

As well as explaining the distinctive epistemological and modal status of disciplines like mathematics and logic, these "dogmas" provided a theory of rationality: analytic truths, linking other statements to ones about immediate experience, determine how we should revise our beliefs in the face of surprising observations. Quine rejected reductionism and the epistemological picture that went with it. He did this primarily by presenting an alternative picture of how our beliefs relate to experience, likening the whole of our knowledge to "a man-made fabric which impinges on experience only along the edges" (p. 42). Although experiential surprise requires us to revise our beliefs, "there is much latitude of choice as to what statements to reevaluate in the light of any single contrary experience" (pp. 2–3): "any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system"; and "no statement is immune from revision" (p. 43). The resulting holistic picture (see HOLISM) allows that we may make revisions in logic or mathematics in order to restore harmony between our beliefs and our experience. It is a psychological matter, a "natural tendency to disturb the total system as little as possible" which explains our treating some beliefs as answerable to experience while other, more embedded ones seem secure from empirical falsification (see UNDERDETERMINATION OF THEORY).

In these relatively early writings, Quine espoused a "thorough pragmatism" (p. 46): our scientific knowledge is a "tool" for effective prediction; physical objects are introduced as "convenient intermediaries", or as "cultural posits" (p. 44). He says that the epistemological superiority of the "myth" of physical objects means only that it is "more efficacious than other myths as a device for working a



manageable structure into the flux of experience". This suggests an anti-realist or instrumentalist view of physical object talk and scientific theories – although even in these papers Quine admitted to *believing in* physical objects but not in the gods of Homer. In later writings, his holism is expressed much more moderately and this pragmatism (see PRAGMATISM) is in eclipse. His focus is on the naturalistic idea that epistemology, like all philosophy, is continuous with, or part of, natural science; and science is interpreted realistically. These later writings convert the metaphors and suggestions of "Two Dogmas of Empiricism" into a scientific account of how experience and theory are related.

In a 1968 lecture, "Epistemology Naturalized", Quine stated his mature view of epistemology, a discipline he takes to be concerned with "the foundations of science" (1969, ch. 3). He holds that the stimulation of sensory receptors is all a scientist has to go on in constructing and defending theories. Arguing, once again, that radical reductionism fails, he proposes that we study how this construction actually proceeds: epistemology "falls into place as a chapter of psychology and hence of natural science". This study answers to the traditional concern of epistemology: "to see how evidence relates to theory, and in what ways one's theory of nature transcends any available evidence" (pp. 82–3).

Some have objected that Quine has changed the subject: for the tradition, it would have been circular to use science in order to ground the legitimacy of science, or to defeat scepticism. Quine's response is to challenge the aspiration to ground science as a whole. Turning his back on the Cartesian search for certainty, he insists that the desire to understand science has always been

something that arose within science itself. Unless we are given good scientific reasons for doubting the legitimacy of science, we are warranted in standing firm on our scientific view of the world in order to understand its undoubted legitimacy (see FIRST PHILOSOPHY). Others accuse Quine of a naturalistic fallacy: epistemology is concerned with understanding the normative standards that guide our inquiries, but psychology can only tell us how we actually do defend our beliefs. His response is that once we abandon the Cartesian desire to ground science as a whole, normative epistemology becomes applied science: we can use our scientific understanding of ourselves and our surroundings to debate the best ways of answering particular questions.

See also EMPIRICISM; INDETERMINACY OF TRANSLATION; NATURALIZED EPISTEMOLOGY; ONTOLOGICAL COMMITMENT; ONTOLOGICAL RELATIVITY; PHILOSOPHICAL KNOWLEDGE; PRAGMATISM; PSYCHOLOGY AND EPISTEMOLOGY.

#### WRITINGS

- From a Logical Point of View* (Cambridge, MA: Harvard University Press, 1953).
- Word and Object* (Cambridge, MA: MIT Press, 1960).
- Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969).
- The Roots of Reference* (La Salle, IL: Open Court, 1973).
- "The Nature of Natural Knowledge," in *Mind and Language*, ed. S. Guttenplan (Oxford: Oxford University Press, 1975), 57–81.
- Theories and Things* (Cambridge, MA: Harvard University Press, 1981).
- Pursuit of Truth* (Cambridge, MA: Harvard University Press, 1990).

CHRISTOPHER HOOKWAY

# R

---

**rationalism** “Rationalism” is a multiply ambiguous term whose meaning varies greatly according to the context. The common thread running through its various uses seems to be that the philosopher classified as a rationalist gives undue weight to reason at the expense of something else: in religion, that may be revelation or faith; in politics, tradition; in morals, feeling or sentiment; in epistemology, experience, etc. This apparent commonality is deceptive, however, since “reason” tends to bear different meanings in the different contexts, referring to a faculty of a priori knowledge in epistemology, but being construed much more broadly in religion, morals or politics. The term does generally seem to carry a negative connotation; it is one philosophers typically apply to those with whom they disagree, not to themselves.

The most significant use of the concept of epistemological rationalism is to organize the textual data of the period from Descartes to Kant so that they tell a coherent story with an edifying moral. Modern philosophy, it is often said, begins with a rationalist reaction against scholastic Aristotelianism, a reaction that privileges mathematics as a model of human knowledge. Ideally our knowledge of ourselves, of God and of the world ought to be organized into a deductive system, in which all truths are derived from a relatively small number of axioms and definitions, whose truth is guaranteed by their self-evidence (*see* AXIOMATIZATION, AXIOMATICS). Only if our starting points are absolutely certain, and we proceed by careful, certainty-preserving, deductive steps from them, can we achieve knowledge, for genuine knowledge requires certainty. On this picture of knowledge, experience is essentially irrelevant; it is not needed and cannot provide the certainty we require. This is the textbook rationalist pro-

gramme in epistemology, to which Descartes, Spinoza and Leibniz are all supposed to subscribe (*see* DESCARTES; SPINOZA; LEIBNIZ).

So understood, rationalism is an exercise in extravagant optimism, as might be argued either by considering the mutually inconsistent (and often bizarre) metaphysical systems the rationalists advocated, or by noting the crucial role arguments from experience play in the development of the sciences. It was only natural, the story then goes, that there should develop in opposition to rationalism a school of philosophy which would (over-)emphasize the importance of experience, not only in verifying claims to knowledge, but also in acquiring the concepts those claims employ. All real knowledge depends ultimately on experience. That is the textbook empiricist programme in epistemology, to which Locke, Berkeley and Hume are all supposed, with varying degrees of ideological purity, to subscribe (*see* LOCKE; BERKELEY; HUME). But strict empiricism leads inevitably to radical scepticism and cannot account adequately for the a priori knowledge we do possess. What is required is a theory of knowledge which synthesizes these two opposed tendencies, rescuing what is true in each, but avoiding their exaggerations. We can then see the Kantian system as reconciling the competing claims of reason and experience, giving each its due.

This way of telling the story of philosophy in the seventeenth and eighteenth centuries is aesthetically pleasing, and gratifies our desire to think of philosophy as a progressive discipline, whose later practitioners have a better grasp of its problems than their predecessors had. Those are two reasons for its influence on the construction of university curricula, anthologies and general histories of philosophy (e.g. Copleston, 1958, ch. 1). It probably

represents fairly enough Kant's view (see KANT) of his relation to his predecessors (cf. Kant, 1781, A854/B882). It also has the advantage that it is not entirely false. With judicious attention to the right evidence, it can even be made to seem a fairly accurate picture of the way philosophy developed in this period. To some extent it may represent the way Kant's predecessors thought of themselves (cf. Leibniz, 1765, Preface). Still, many historians specializing in the early modern period now regard this story as a gross distortion of what happened. The remainder of this article will sketch the reasons why the picture is appealing to some people and sharply rejected by others.

Descartes did much to encourage reading him as a rationalist when he wrote in the *Discourse* (1637, VI/19) of his admiration for "those long chains composed of very simple and easy reasonings", which geometers use to demonstrate their most difficult propositions. This led him to suppose that "all the things which can fall under human knowledge are interconnected in the same way". All we need to do, it seems, is withhold our assent from anything we do not see clearly and distinctly to be true, and then proceed in a careful and orderly fashion from our initial certainties. The method the *Discourse* advocates would seem from this description to be wholly a priori (see A PRIORI/A POSTERIORI).

The reader might then be surprised by the scientific essays which accompany the *Discourse*, and illustrate what Descartes has accomplished with his method. Descriptions of what can only be characterized as experiments seem to be crucial to its application (there are particularly striking examples in the Fifth Discourse of the *Optics*, or the Eighth of the *Meteorology*). Surprise is unwarranted, however, since an appeal to experience is an acknowledged part of the method in the *Discourse* itself. In the end it appears that what can be demonstrated a priori are only the fundamental principles of physics. Presumably these include the basic laws of motion deduced in the unpublished *Le monde* from God's immutability; what else they might include is obscure (but see Curley 1978, ch. 8). "The further we advance in our knowledge, the

more necessary observations become" (VI/63). The first principles are so simple and general that effects can be deduced from them in a variety of ways. To learn which way an effect really depends on first principles, we must conduct experiments whose outcomes will vary depending on which hypothesis is correct (VI/65).

If Descartes' scientific theory and practice are less a priori than the rationalist paradigm would lead us to expect, the paradigm also has deficiencies in explaining his approach to metaphysical issues. Descartes claimed to have written his *Meditations* in geometrical order (VII/155), but his work does not have the external trappings of a mathematical treatise. Unlike Spinoza's *Ethics*, it does not begin with definitions and axioms, and proceed from them by formal deductions of theorems. Descartes distinguished between the geometric *order* and geometric *manner* of writing. The essence of the geometric order was that no proposition should be advanced which could not be known solely by consideration of what had preceded it in the argument, without the aid of anything later. Descartes tried to adhere to that order, but regarded the geometric manner of writing as inappropriate in metaphysics, since it presupposes that its first principles are already clear and distinct. This presupposition holds in mathematics, but not in metaphysics, where the main problem is to get a clear and distinct perception of the primary notions (VII/157). Achieving that kind of perception requires use of the analytic method; to judge from the *Meditations* themselves, which exemplify it, the analytic method involves a systematic reflection on our past beliefs and suspense of judgment whenever one belief suggests ground for doubting another (Curley, 1986a). Though Descartes' earliest work on method (the unpublished and unfinished *Rules for the Direction of the Mind*) seemed to assume uncritically that we possess an infallible faculty of intuiting first principles, the Descartes of the *Meditations* is more ambitious and more interesting; he is prepared to doubt (provisionally, at least) even the simplest propositions of mathematics, and undertakes a dialectical defence of his first principles

and the reliability of the faculty by which we know them (Curley, 1978, chs 2–5).

If Descartes does not fit the rationalist profile, it may be thought, Spinoza must. After all, he does organize the definitive exposition of his philosophy in the geometric manner. Doesn't this imply that he regards the axioms and definitions of his *Ethics* as self-evident truths, which neither require nor admit of argument, but are necessarily seen to be true as soon as we understand the terms in which they are framed? Doesn't it imply that the propositions of the *Ethics* derive their certainty from that of the axioms and definitions, and not conversely?

Oldenburg put the first question to Spinoza when they corresponded about an early draft of the *Ethics* (1677, IV/10–11). Spinoza's reply was that he did not care whether Oldenburg regarded these principles as self-evident, so long as he would allow that they were true (IV/13). Significantly, in the final version of the *Ethics* Spinoza treats most of the principles Oldenburg questioned, not as axioms, but as propositions. What does Spinoza think the status of his axioms is? No doubt he thought some were self-evident. Others seem to be propositions whose warrant is empirical, e.g., "We feel that a certain body is affected in many ways" (II/86). The view seems to be that there are some empirical propositions so fundamental to our thought about the world that we cannot rationally doubt them. And the system does not make its claim on our belief solely because its axioms are absolutely certain and its deductions valid. When Spinoza reaches conclusions he knows his readers will find hard to accept, he asks them "to continue on with me slowly, step by step, and to make no judgement on these matters until they have read through them all" (II/95). We are to judge the system as a whole, partly on the basis of the coherence of its conclusions with our experience (on this see Bennett, 1984, pp. 16–25; and Curley, 1986b). The contrast we find in Spinoza between reason and experience may be misleading, since it seems that "experience" often refers only to a casual, uncritical use of experience, as opposed to a systematic, critical use of it (cf. II/10–11 with II/36–7), and

since reason does not appear to be conceived as a wholly a priori faculty (cf. Curley, 1973). Researchers have not sufficiently noted the absence of a theory of innate ideas (see INNATE IDEAS) in Spinoza, or his account of our knowledge of the common notions which are the basis of reason (II/118–22).

Turning from metaphysics to the sciences, we encounter similar difficulties in making Spinoza fit preconceived notions of rationalism. Some regard his correspondence with Boyle as a classic confrontation between a rationalist and an empiricist (Hall and Hall, 1964). But this seems a very unsympathetic reading of the correspondence. Spinoza's principal complaints about Boyle are that some of his experiments were unnecessary, because the conclusions drawn from them had already been established adequately by the simpler experiments of Bacon (see BACON) and Descartes (IV/25), and that other conclusions required further experiment to justify the conclusion drawn (IV/16–17). Some invoke Spinoza's (supposed) rationalism to explain his (supposed) failure to make any substantial contribution to the sciences (Maull, 1986). But if biblical hermeneutics is a science, the *explanandum* should not be granted. Spinoza played an important role in the development of that discipline; his success resulted from his extending to textual interpretation a model of scientific procedure which emphasizes the importance of the careful, systematic collection of relevant data and rejects any attempt to determine the meaning of the text by applying principles known by the light of reason (Savan, 1986, commenting on Spinoza, 1670).

If neither Descartes nor Spinoza fits the rationalist profile, perhaps there is hope for Leibniz. Doesn't he maintain that every true proposition is either an identity or reducible to an identity by the analysis of concepts, thereby making possible an a priori proof of every truth (Leibniz, 1903, p. 267)? Doesn't he claim to deduce from that theory of truth factual propositions which could only be known by experience, e.g. that in nature there cannot be two individuals which differ only numerically (1903, p. 268)? Doesn't he

defend innate ideas and principles against the Lockean attack (1765, Book I)?

The answer to all the above questions is yes. Yet Leibniz can be quite Baconian about the need to compile "histories of nature", cataloguing, classifying and verifying the experiments already performed; he also advocates setting up laboratories at public expense to perform new experiments (Leibniz, 1956, pp. 281–2). When Locke expresses pessimism about the possibility of attaining scientific knowledge however far "human industry may advance... experimental philosophy in physical things", Leibniz replies that "considerable progress will eventually be made in explaining various phenomena... because the great number of experiments which are within our reach can supply us with more than sufficient data" (1765, p. 389). The distinction between truths of reason and truths of fact (*see* TRUTHS OF REASON/TRUTHS OF FACT) is as fundamental in Leibniz's philosophy as that between relations of ideas and matters of fact in Hume's. If even truths of fact are ultimately analytic (*see* ANALYTICITY) in Leibniz's system, still, only God can perceive the a priori demonstration of truths about individuals (1857, p. 264). Human use of the a priori method in science is "not entirely impossible", so long as we don't descend to particular things, but even at the level of general truths only superior geniuses can hope for anything from it in this life. Most of us must rely on what Leibniz calls "the conjectural a priori method", which we might nowadays call "the hypothetico-deductive method" (1956, p. 283, cf. Descartes, 1644, VIII-1/327–8, Spinoza, 1663, 1/226–9). Even in mathematics, where we deal with truths of reason, Leibniz is reluctant to rely on the supposed self-evidence of the axioms, urging that mathematicians demonstrate "the secondary axioms we ordinarily use" by reducing them to identities (1765, p. 408).

None of the three great philosophers commonly counted as paradigms of epistemological rationalism was as dismissive of experience or as trusting of a priori intuition as traditional accounts of rationalism imply. Such difficulties have made specialists increasingly reluctant to accept that classification.

(For further elaboration, beyond the works already cited, *see* Clarke, 1982; Rescher, 1986. For an attempt to rescue something of the traditional picture in the face of acknowledged difficulties, *see* Cottingham, 1988.) Determining the exact relation between reason and experience in the three paradigm "rationalists" is difficult; an accurate account would be too complicated to make for a readable general history, and would probably not conclude that they had any significant epistemological programme in common, or that they constituted a school of thought to which empiricism could usefully be opposed.

Nevertheless, the traditional way of structuring that history will continue to be influential so long as historians do not provide a more attractive alternative paradigm. Loeb 1981 is a lively attempt to do just that, arguing that the figures should be classified according to their metaphysical, not their epistemological, doctrines, and grouping Malebranche and Berkeley with Descartes, Spinoza and Leibniz, because of their common willingness to deny generally accepted causal relations. Another alternative, possibly more fruitful, would be to group the figures by their approach to the conflict between religion and science, depending on whether they thought (1) that we can reconcile traditional religion with the new science (Descartes, Malebranche, Leibniz, Locke); or (2) that we should abandon or radically revise traditional religion because of its inconsistency with a scientific worldview (Hobbes (*see* HOBBS), Spinoza, Hume); or (3) that we should revise the worldview presupposed by the new science to make it compatible with traditional religion (Berkeley). Accepting this principle of classification would require delicate judgements about how free philosophers felt they were to openly advocate views radically opposed to those of their culture, and what they might have said had they felt completely free. How historians of early modern philosophy sort out these matters of interpretation and classification will do much to determine whether the next generation of reference works such as this will still count rationalism as a major "movement" in the history of philosophy.



See also A PRIORI KNOWLEDGE in Part I; EMPIRICISM; GEOMETRY.

## BIBLIOGRAPHY

Dates used in the text are those of the first publication; pagination is that of the standard modern edition, when that is given in the margins of the standard translation, in which case "II/86" means "volume II, p. 86", otherwise the pagination is that of the translation cited.

- Bennett, Jonathan: *A Study of Spinoza's Ethics* (Totowa, NJ: Hackett, 1984).
- Clarke, Desmond: *Descartes' Philosophy of Science* (Manchester: Manchester University Press, 1982).
- Copleston, F.: *A History of Philosophy* (Westminster, MD: Newman Press, 1958), vol. IV.
- Cottingham, John: *The Rationalists* (Oxford: Oxford University Press, 1988).
- Curley, Edwin: "Experience in Spinoza's Theory of Knowledge," in *Spinoza: A Collection of Critical Essays*, ed. Marjorie Grene (Garden City, NY: Anchor, 1973).
- Curley, Edwin: *Descartes against the Skeptics* (Cambridge, MA: Harvard University Press, 1978).
- Curley, Edwin: "Analysis in the *Meditations*: The Quest for Clear and Distinct Ideas," in *Essays on Descartes' Meditations*, ed. A. Rorty (Berkeley, CA: University of California Press, 1986a).
- Curley Edwin: "Spinoza's Geometric Method," *Studia Spinozana* 2 (1986b), 151–68.
- Descartes, René: *Discours de la méthode et les essais* (Leiden, 1637); *Meditationes de prima philosophiae* (Paris, 1641); *Principia philosophiae* (Amsterdam, 1644); trans. J. Cottingham, R. Stoothoff and D. Murdoch as *Philosophical Writings of Descartes*, 2 vols (Cambridge: Cambridge University Press, 1984).
- Hall, A. R. and Hall, M. B.: "Philosophy and Natural Philosophy: Boyle and Spinoza," in *Mélanges Alexandre Koyré* (Paris: Hermann, 1964), vol. II, pp. 241–6.
- Kant, Immanuel: *Kritik der reinen Vernunft* (Riga, 1781); trans. N. Kemp Smith as *Critique of Pure Reason* (New York: St Martin's Press, 1965).
- Leibniz, G. W.: "On Freedom" (Paris, 1857), "First Truths" (Paris, 1903), "An Introduction on the Value and Method of Natural Science" (Chicago, IL, 1956); trans. in L. Loemker, *Philosophical Papers and Letters of G. W. Leibniz* (first edition, Chicago, IL: University of Chicago Press, 1956;

page references to the second edition, Dordrecht: Reidel, 1969).

- Leibniz, G. W.: *Nouveaux essais sur l'entendement humain* (Berlin, 1765); trans. P. Remnant and J. Bennett as *New Essays on Human Understanding* (Cambridge: Cambridge University Press, 1981).
- Loeb, Louis: *From Descartes to Hume* (Ithaca, NY: Cornell University Press, 1981).
- Maull, Nancy: "Spinoza in the Century of Science," *Spinoza and the Sciences*, ed. Marjorie Grene and Debra Nails (Dordrecht: Reidel, 1986).
- Rescher, Nicholas: *Leibniz: An Introduction to His Philosophy* (Lanham, MD: University Press of America, 1986), ch. 12.
- Savan, David: "Spinoza: Scientist and Theorist of Scientific Method," *Spinoza and the Sciences*, 95–123.
- Spinoza, Benedict: *Renati Descartes Principiorum philosophiae pars I & II* (Amsterdam, 1663); *Opera posthuma* (Ethics and correspondence, Amsterdam, 1677); texts from these works cited in this article are translated in E. Curley, *The Collected Works of Spinoza* (Princeton, NJ: Princeton University Press, 1985), vol. I. The translation of the *Tractatus theologicophilosophicus* (Hamburg, 1670) is to appear in vol. II, forthcoming.

EDWIN CURLEY

**rationality** To be rational is to be guided by legitimate reasoning. But different criteria of legitimacy are normally considered appropriate for different types of reasoning and so at least nine types of rationality, or roles for the faculty of reason, seem to be commonly recognized in Western culture.

First, there is the rationality that consists in conformity with the laws of deductive logic. Thus it would be termed rational (as an instance of the law termed "*modus ponens*") to infer "The streets are wet" from the two premisses "It is raining" and "If it is raining, the streets are wet", whereas it would be irrational to infer "The streets are not wet" from those premisses. The exact extent of such laws is, however, controversial. For example, deductive logic may or may not be conceived to include laws not only about inferences depending on propositional connectives like "if" and "or", or on quantifiers like "some" and "all", but also about those

depending on modalities such as “possibly” and “necessarily”.

A second form of rationality is exhibited by correct mathematical calculations. Thus it is rational to infer “ $x > 12$ ” from the premisses “ $x$  is a prime number” and “ $x > 11$ ”. Of course, this has nothing to do with the technical sense in which numbers, like the square root of 2, that are not equivalent to the ratio of one integer to another, are termed “irrational” and others “rational”.

A third form of rationality is exhibited by reasoning that depends for its correctness solely on the meanings of words that belong neither to our logical nor to our mathematical vocabulary. Thus it would be rational to infer that, if two people are first cousins of one another, they share at least one grandparent.

A fourth form of rationality is exhibited by ampliative inductions that conform to appropriate criteria, as in an inference from experimental data to a general theory that explains them. For example, a hypothesis about the cause of a phenomenon needs to be tested in a relevant variety of controlled conditions in order to eliminate other possible explanations of the phenomenon, and it would be irrational to judge the hypothesis to be well supported unless it had survived a suitable set of such tests. Similarly, it would be irrational to suppose a rule of common law to be well-supported if it had not been implicit in a suitable range of judicial decisions.

A fifth form of rationality is exhibited by correct assessments of mathematical probability. For example, it would be irrational to infer that a particular person, John Smith, has a  $2/5$  probability of dying before the age of 70 just because he is a lorry-driver and  $2/5$  of all lorry-drivers die before the age of 70: he might have many other attributes relevant to his prospects of survival. Of course, the rationality of a probability judgement may also be affected by its mathematical relationship to other judgements of probability. For example, if the mathematical probability of a particular outcome is at issue it would be irrational to assess both the occurrence and the non-occurrence of the outcome at values greater than .5.

A sixth form of rationality is exhibited by inferences that are licensed by an accepted factual generalization. For example, if the pavement is wet, it would normally be rational to infer that it has recently rained and irrational to infer that it has not.

A seventh form of rationality, unlike the previous six, is concerned with practice rather than cognition. It is exhibited by actions that further the purposes or interests of the agent. Thus, if a model of economic behaviour is said to assume the rationality of those operating a free market, what is implied is that each agent conducts his transactions in the way that will maximize his own profits. And, where the agent cannot be certain about what will promote his purposes or interests, rationality is often assumed to require him to prefer that course of action which will maximize the arithmetical product of the probability and utility.

An eighth form of rationality that is sometimes supposed to exist is also concerned with practice but with the choice of ends rather than of means to ends. Thus it might be considered rational to treat other human beings as ends in themselves rather than as means to your own ends. The ascription of such rationality would constitute a fundamental moral judgement.

A ninth form of rationality is also concerned with practice but with linguistic communication rather than with action and decision. Successful communication is a cooperative endeavour that has to engage the participants' reasoning powers over and above the comprehension of literal meanings. In particular there is a presumption, at least within Western culture, that persons conversing with one another will conform to certain tacit maxims prescribing relevance, informativeness, brevity, orderliness, etc. Consequently, as Grice (1975) (*see* GRICE) argued, it is rational to draw certain inferences from what they say over and above whatever inferences are licensed by standard principles of deductive or inductive reasoning. For example, it may be rational to suppose, in an appropriate context, that a professor who is described just as being a good bicyclist is thereby implied to possess no other skills or accomplishments.

Philosophical theories about rationality may be classified as either affirmative or negative. Affirmative theories endorse one or more culturally accepted roles for rationality, or propose some further role, while negative theories are sceptical about the validity of one or more such roles.

Thus the tripartite analysis of the soul that Plato (*see* PLATO) propounds in his *Republic* guarantees a dominant role for the faculty of reason both in the cognitive task of discovering the explanation of how things are and the justification of how they ought to be and also in the practical task of correctly managing a person's life. Reason, according to Plato, is like an eye of the soul and, after a due process of education and study, it can discern ultimate reality. Knowledge and virtue are therefore so bound up together in Plato's conception of rationality that no one who knows the right way to act can act wrongly. Aristotle (*see* ARISTOTLE), however, insisted on the independence of cognitive and practical rationality, and thus allowed for the possibility of a person's doing what he knows to be wrong. Both Plato and Aristotle may be said to have held affirmative theories of rationality. But, while Aristotle's theory endorses culturally accepted norms in allowing room for an agent to feel remorse, Plato's theory attributes greater potential to the faculty of reason than ordinary norms of rationality assume.

The rise of modern science gave new impetus to affirmative theorising about rationality. It was probably at least in part because of the important part played by mathematics in the new mechanics of Kepler, Galileo and Newton, that some philosophers thought it plausible to suppose that rationality was just as much the touchstone of scientific truth as of mathematical truth. At any rate that supposition seems to underlie the epistemologies of Descartes and Spinoza, for example, in which observation and experiment are assigned relatively little importance compared with the role of reason. Correspondingly it was widely held that knowledge of right and wrong is knowledge of necessary truths that are to be discovered by rational intuition in much the same way as it was believed that the fundamental principles of arithmetic and

geometry are discovered. For example, Richard Price argued that "a rational agent void of all moral judgment . . . is not possible to be imagined" (1787, p. 72).

In contrast with this thoroughgoing rationalism, Kant's critical philosophy (*see* KANT) may be seen as attempting a more restrictively affirmative account. On his view reason supplies the human mind with regulative ideals, not constitutive ideas. Such principles guide us in the systematization of knowledge but generate irresolvable antinomies if interpreted as representations of features of reality. In ethics, too, on Kant's view, reason guides our choice of maxims on which to act, but does not actually supply those maxims itself.

Negative theorizing about rationality generates a sceptical challenge to one or more culturally accepted principles. Even the fundamental laws of deductive logic have been exposed to such challenges. For example, the principle that every proposition is either true or false but not both was called into question in the fourth century BC by the paradox of Eubulides. (Suppose only one sentence is written on a piece of paper, viz. "The sentence on this piece of paper is false": then that sentence is arguably both false if it is true and true if it is false.)

But in modern philosophy the most influential sceptical challenge to everyday beliefs about rationality was originated by Hume (*see* HUME). Hume argued the impossibility of reasoning from the past to the future or from knowledge about some instances of a particular kind of situation to knowledge about all instances of that kind. There would be nothing contradictory, he claimed, in supposing both that the sun had always risen in the past and that it would not rise tomorrow. In effect therefore Hume assumed the only valid standards of cognitive rationality were those of the first three kinds listed above – viz. deductive, mathematical or semantical. Induction was not a rational procedure, on his view, because it could not be reduced to the exercise of reason in one or another of these three roles (*see* PROBLEMS OF INDUCTION).

Hume's argument here is often criticized for begging the question on the ground that induction should be held to be a valid process

in its own right and with its own criteria of good and bad reasoning. But this response to Hume seems just to beg the question in the opposite direction. What is needed instead, perhaps, is to demonstrate a continuity between inductive and deductive reasoning, with the latter exhibited as a limiting case of the former (Cohen, 1989, pp. 186–7). Even so, Hume's is not the only challenge that defenders of inductive rationality need to rebuff. Popper (*see* POPPER) has also denied the possibility of inductive reasoning, and much-discussed paradoxes about inductive reasoning have been proposed by Goodman and Hempel (*see* GOODMAN; HEMPEL; HEMPEL'S PARADOX).

Hume also argued, as against philosophers like Richard Price (1787), that it was impossible for any reasoning to demonstrate the moral rightness or wrongness of a particular action. There would be nothing self-contradictory in preferring the destruction of the whole world to the scratching of one's little finger. The only role for reason in decision-making was to determine the means to desired ends. Nevertheless Price's kind of ethical rationalism has been revived in more recent times by W. D. Ross (1930) and others. Perhaps Hume's argument here was again based on question-begging assumptions, and it may be more cogent to point out that ethical rationalism implies a unity of moral standards that is not found to exist in the real world.

Probabilistic reasoning is another area in which the possibility of attaining fully rational results has sometimes been queried, as in the lottery paradox (*see* LOTTERY PARADOX). And serious doubts have also been raised (Sen, 1982) about the concept of a rational agent that is required by classical models of economic behaviour. No doubt a successful piece of embezzlement may in certain circumstances further the purposes or interests of an accountant, and need not be an irrational action. But is it entitled to the accolade of rationality? And how should its immorality be weighed against its utility in the scales of practical reasoning? Or is honesty always the rationally preferable policy?

These philosophical challenges to rationality have been directed against the very

possibility of there existing valid standards of reasoning in this or that area of enquiry. They have thus been concerned with the integrity of the concept of rationality rather than with the extent to which that concept is in fact instantiated by the actual thoughts, procedures and actions of human beings. Indeed the latter issue seems at first sight to be a matter for psychological, rather than philosophical, research. Some of this research will no doubt be concerned with the circumstances under which people fail to perform in accordance with valid principles that they have nevertheless developed or adopted, as when they make occasional mistakes in their arithmetical calculations. But there also seems to be room for research into the content of the principles that particular categories of the population have developed or adopted. Some of this would be research into the success with which the relevant principles have been taught, as when students are educated in formal logic or statistical theory. Some would be research into the extent to which those who have not had any relevant education are, or are not, prone to any systematic patterns of error in their reasoning. And it is this last type of research that has claimed results with "bleak implications for human rationality" (Nisbett and Borgida, 1975).

One robust result here is when (Wason, 1966) logically untutored subjects are presented with four cards showing, respectively, "A", "D", "4" and "7" and they know that every card has a letter on one side and a number on the other. They are then given the rule "If a card has a vowel on one side, it has an even number on the other", and told that their task is to say which of the cards they need to turn over in order to find out whether the rule is true or false. The most frequent answers are "A and 4" and "Only A" which are both wrong, while the right answer "A and 7" is given spontaneously by very few subjects. Wason interpreted this result as demonstrating that most subjects have a systematic bias towards seeking verification rather than falsification in testing the rule, and he regarded this bias as a fallacy of the same kind as Popper claimed to have discerned in

the belief that induction could be a valid form of human reasoning.

Some of these results concern probabilistic reasoning. For example, in an experiment (Kahneman and Tversky, 1972) on statistically untutored students the subjects are told that in a certain town blue and green cabs operate in a ratio of 85 to 15 respectively. A witness identifies the cab in an accident as green and the court is told that in the relevant circumstances he says that a cab is blue when it is blue, or that a cab is green when it is green, in 80 per cent of cases. When asked the probability that the cab involved in the accident was blue subjects tend to say 20 per cent. The experimenters have claimed that this robust result shows the prevalence of a systematic fallacy in ordinary people's probabilistic reasoning, viz. a failure to pay proper attention to prior probabilities. And it has been argued (Saks and Kidd, 1980–1, p. 134) that the existence of several such results demonstrates the inherent unsoundness of mandating lay juries to decide issues of fact in law courts.

However, it is by no means clear that these psychological experimenters have interpreted their data correctly or that the implications for human rationality are as bleak as they suppose (Cohen, 1981, 1982). For example, it might be argued that Wason's experiment merely shows the difficulty that people have in applying the familiar rule of contraposition to artificial conditional relationships that lack any basis in causality or in any other explanatory system. And as for the cabs, it might well be disputed whether the size of the fleet to which a cab belongs should be accepted as determining a prior probability that can count against a posterior probability founded on the causal relation between a witness's mental powers and his courtroom testimony. To count against such a posterior probability one would need a prior one that was also rooted in causality, such as the ratio in which cabs from the blue fleet and cabs from the green fleet (which may have different policies about vehicle maintenance and driver training) are involved in accidents of the kind in question. In other words the subjects may interpret the question to concern proba-

bilities conceived as causal propensities, not probabilities conceived as relative frequencies that may be accidental. Indeed it is always necessary to consider whether the dominant responses given by subjects in such experiments should be taken, on the assumption that they are correct, as indicating how the task is generally understood – instead of as indicating, on the assumption that the task is understood exactly in the way intended, what errors are being made.

Finally, there is an obvious paradox in supposing that untutored human intuitions may be systematically erroneous over a wide range of issues in human reasoning. On what non-circular basis other than such intuitions can philosophers ultimately found their theories about the correct norms of deductive or probabilistic reasoning? No doubt an occasional intuition may have to be sacrificed in order to construct an adequately comprehensive system of norms. But empirical data seem in principle incapable of showing that the untutored human mind is deficient in rationality, since we need to assume the existence of this rationality – in most situations – in order to provide a basis for those normative theories in terms of which we feel confident in criticizing occasional errors of performance in arithmetical calculations, etc.

*See also* PSYCHOLOGY AND EPISTEMOLOGY.

#### BIBLIOGRAPHY

- Cohen, L. J.: "Can human irrationality be experimentally demonstrated?," *The Behavioral and Brain Sciences* 4 (1981), 317–31 and 359–67.
- Cohen, L. J.: "Are People Programmed to Commit Fallacies? Further Thoughts about the Interpretation of Experimental Data on Probability Judgment," *Journal for the Theory of Social Behaviour* 12 (1982), 251–74.
- Cohen, L. J.: *An Introduction to the Philosophy of Induction and Probability* (Oxford: Clarendon Press, 1989).
- Grice, H. P.: "Logic and Conversation," in *The Logic of Grammar*, ed. D. Davidson and G. Harman (Encino, CA: Dickinson, 1975), 64–75.
- Kahneman, D. and Tversky, A.: "On the Psychology of Prediction," *Oregon Research Institute Bulletin* 12: 4 (1972).



- Nisbett, R. E. and Borgida, E.: "Attribution and the Psychology of Prediction," *Journal of Personal and Social Psychology* 32 (1975), 932–43.
- Price, R.: *A Review of the Principal Questions in Morals* (London: T. Cadell, 1787).
- Ross, W. D.: *The Right and the Good* (Oxford: Clarendon Press, 1930).
- Saks, M. J. and Kidd, R. F.: "Human Information-processing and Adjudication: Trial by Heuristics," *Law and Society Review* 15 (1980–1), 123–60.
- Sen, A.: *Choice, Welfare and Measurement* (Oxford: Blackwell, 1982), 84–106.
- Wason, P. C.: "Reasoning," in *New Horizons in Psychology*, ed. B. Foss (Harmondsworth: Penguin, 1966), vol. 1.

L. JONATHAN COHEN

**realism** Realism in any area of thought is the doctrine that certain entities allegedly associated with that area are indeed real. Common sense realism – sometimes called "realism", without qualification – says that ordinary things like chairs and trees and people are real. Scientific realism says that theoretical posits like electrons and fields of force and quarks are equally real. And psychological realism says mental states like pains and beliefs are real. Realism can be upheld – and opposed – in all such areas, as it can with differently or more finely drawn provinces of discourse: for example, with discourse about colours, about the past, about possibility and necessity, or about matters of moral right and wrong. The realist in any such area insists on the reality of the entities in question in the discourse.

If realism itself can be given a fairly quick characterization, it is more difficult to chart the various forms of opposition, for they are legion. Some opponents deny that there are any distinctive posits associated with the area of discourse under dispute; a good example is the emotivist doctrine that moral discourse does not posit values but serves only, like applause and exclamation, to express feelings. Other opponents deny that the entities posited by the relevant discourse exist or at least exist independently of our thinking about them; here the standard example is idealism (see IDEALISM). And others again insist that the

entities associated with the discourse in question are tailored to our human capacities and interests and, to that extent, are as much a product of invention as a matter of discovery.

The variety of the opposition shows that realism about any area of discourse, though apparently simple, actually involves a number of distinct claims. I distinguish three, which I call respectively descriptivism, objectivism and cosmocentrism.

#### THE DESCRIPTIVIST THESIS

Participants in the discourse necessarily posit the existence of distinctive items, believing and asserting things about them: the utterances fail to come off, as an understanding of them reveals, if there are no such entities. The entities posited are distinctive in the sense that, for all that participants are in a position to know, the entities need not be identifiable with, or otherwise replaceable by, entities independently posited. Although realists about any discourse agree that it posits such entities, they may differ about what sorts of things are involved. Berkeley (see BERKELEY) differs from the rest of us about what common sense posits and, less dramatically, colour realists differ about the nature of colours, mental realists about the status of psychological states, modal realists about the locus of possibility, and moral realists about the place of value.

#### THE OBJECTIVIST THESIS

The objects posited exist and have their character fixed independently of the dispositions of participants in the discourse to assert and believe things about them. Thus the epistemic states of the participants have no causal influence on the existence or character of those objects, nor are the objects supervenient in any way (see EPISTEMIC SUPERVENIENCE) – that is, non-causally dependent – on such epistemic states. In short, the entities posited in the discourse enjoy a substantial kind of objectivity.

## THE COSMOCENTRIC THESIS

In order to avoid error and ignorance with regard to substantive propositions of the discourse – in order to get at the truth – participants have to make suitable contact with the objects of the discourse and there is no guarantee that they will succeed in doing so. The human search for truth is a matter of discovery, not invention, and discovery is a matter of contingent success. Ignorance is possible, because normally it is possible that human subjects lack contact with certain regions of the independent reality in question. Error is possible, because normally it is possible that human subjects are only imperfectly attuned to the regions with which they do make contact.

The realist's first thesis puts him in conflict with at least three sorts of opponent: the reductivist, the instrumentalist and theorists who explore sophisticated variations on instrumentalism. The reductivist says of the area of discourse in question, that it can be reduced to some other area and does not therefore introduce any distinctive posits. Thus she may say that commonsense discourse about physical objects or scientific discourse about unobservable entities reduces to talk about the purely phenomenal level, that moral discourse reduces to talk about the attitudinal or that mental discourse reduces to talk about the purely behavioural level. The instrumentalist says of the discourse that it does not involve talk about anything at all, let alone novel existential posits, because it is not assertoric: it does not involve assertions, only utterances with the force of imperatives, exclamations, or whatever. Thus she says that theoretical discourse in science is really just a way of generating appropriate laboratory dispositions – "That's fragile" has the imperative force of "Be careful" – and that moral discourse is just a way of expressing emotions, a way of making exclamations of approval or disapproval: emotivism, on this account, is a variety of moral instrumentalism. There are two currently influential variants on instrumentalism, projectivism and constructive empiricism. The projectivist holds that the discourse in

question serves the sort of role ascribed to it by the instrumentalist, and does not involve distinctive posits, but that it still has the marks of assertoric talk that impress – and mislead – the realist. (Blackburn, 1984) The constructive empiricist – a sort of fictionalist – holds that while the discourse represents assertoric talk about the relevant sorts of objects, accepting what is said – participating in the discourse – does not mean positing those objects; it may only mean treating the propositions involved as empirically adequate, treating them as adequate for the practical purposes on which instrumentalists focus (Van Fraassen, 1980).

The realist's second thesis puts him in conflict with two main sorts of opponent: the error theorist and the idealist. The error theorist denies that there are any objects of the kind that the discourse in question posits (Mackie, 1977). While admitting that modal discourse posits the existence of possibilities, and moral discourse posits that of values, she denies that there are any such things; thus she says that assertions and beliefs within the area of discourse inevitably fail to be true. Unlike the error theorist, the idealist admits that the objects posited do exist, as Berkeley admits the existence of the items he takes common sense to posit. What she denies is that the objects are independent of people's dispositions to believe and assert things about them. Such objects are held to depend in some way on people's dispositions; the dispositions invoked may be individual or shared, depending on whether the idealism involved is of the subjective or objective variety (*see* IDEALISM). Berkeley is a subjective idealist, whereas someone like Hegel (*see* HEGEL) is usually counted as an idealist of an objective – better, an intersubjective – kind.

The realist's first two theses in any area of discourse can be run together into a straightforward claim, made within the discourse itself, that there are such and such entities and they are independent of our epistemic influence. On this representation, the realist about commonsense says that there are independent chairs and tables and other such objects, the realist about science says that there are independent protons and electrons

and things of that ordinarily unobservable kind (Devitt, 1984). This is a perfectly accurate way of expressing the realist's first two claims, though it fails to make clear that there are very different ways of rejecting his position: the ways that correspond with the denial of the descriptivist and objectivist theses, respectively.

The third, cosmocentric thesis is made central to realism by some writers (Smart, 1982; Papineau, 1987) but not by all (Devitt, 1984). It puts the realist in conflict with an opponent that we can describe as the anthropocentric. The anthropocentric says that with certain substantive propositions within the discourse in question – with a certain number or with certain specific cases – there is no possibility that specified individuals or groups could be in ignorance or error. The anthropocentric may deny the possibility of a certain error or ignorance by taking the interpretationist line that the objects posited by a discourse are whatever objects participants are mostly right about; this will put limits on error (Davidson, 1984; Rorty, 1980). She may do it by going the verificationist or anti-realist way of refusing to acknowledge that propositions for which we lack adjudication procedures have a determinate truth-value; this will put limits on ignorance (Tennant, 1987, ch. 2). She may do it through becoming a relativist and increasing a group's chances of hitting the truth by moving the target nearer: by defining truth, in the relevant sense, as truth relative to that group. Or she may take any of a variety of other approaches (Goodman, 1978; Putnam, 1981; Price, 1988; Pettit, 1990). Whatever form the anthropocentric claim takes, however, the realist will deny it. He says of any discourse he judges favourably that error and ignorance are always possible with regard to the substantive propositions of the discourse. It is possible that participants are wrong about all and every substantive claim in the discourse.

It may at first seem that there is going to be an inconsistency involved in agreeing to the first two realist theses and then, in the anthropocentric style, denying the third. If there were, that would mean that anthropocentrism was not really an independent

way of rejecting realism. But there is no inconsistency involved and anthropocentrism does represent an independent mode of opposition to realism. It may be entailed by certain ways of denying the first two theses but it does not require the denial of either. Consistently with thinking that a discourse introduces distinctive posits, and that the posited objects are suitably independent of people's epistemic states, a philosopher might hold that the posited objects are tailored to human interests and capacities. Consider the popular view of colour according to which it is *a priori* true that something is red if and only if it looks red to normal observers under normal conditions. This view allows that the redness property is posited in colour discourse and it may identify that property with a certain physical feature: say, a certain surface reflectance. Thus it may be compatible with descriptivism and objectivism about colour. But under the view in question, cosmocentrism fails, for there are limits on human error: normal observers in normal conditions cannot be wrong about the colour of something. Objective though it is on this approach, colour is nevertheless an anthropocentric property, a property that reflects an element of invention as well as discovery.

There are three things that need to be said in further commentary on the realist's cosmocentric thesis. The first is that while it invokes the notion of truth, the notion involved is just that which is given by the disquotational schema, “‘*p*’ is true if and only if *p*”. I assume that the notion of assertion is given, so that we understand what is involved in asserting that *p*, for any arbitrary “*p*”; for example, we understand that it is inconsistent with asserting that not-*p*, that it is equivalent to denying that not-*p*, and that it combines with the assertion that if *p*, then *q* to license the assertion that *q*. Given an understanding of assertion, the disquotational schema is sufficient to communicate an understanding of truth in the sense in which the realist's third thesis – or the anthropocentric's counter-thesis – invokes the notion.

The second thing that needs to be said about the realist's third thesis bears on the question of what truths are sufficiently substantive to

be relevant to the thesis. The realist says that error and ignorance are possible with regard to the substantive propositions in any area of discourse. So which propositions, if any, are non-substantive? My answer is brief: if a proposition is such that just to count as a proper participant in the discourse in question, just to count as someone who understands what is going on, you must accept the proposition or you must reject it, then it is non-substantive; otherwise it is substantive. By many accounts, there are truths in every area of discourse whose acceptance or whose rejection is criterial for counting as a proper participant there: you must accept them – they are so obviously true – or you must reject them – they are so obviously false – if you are going to be held as someone who genuinely asserts and believes things in the discourse, as someone who understands enough not to be seen as a mere mouther of words. If a realist accepts such an account, then he will naturally deny that error and ignorance are possible for proper participants in the discourse with such propositions. But that denial will not come of any faltering in his realist commitments; it will merely give expression to his view of what proper participation in the discourse presupposes. The realist will have to regard it as a non-substantive proposition of a discourse that there are the entities associated with the discourse since, by the descriptivist thesis, participants necessarily posit such items and by the objectivist thesis they cannot be wrong to do so. Otherwise he can be uncommitted: he may or may not acknowledge further non-substantive propositions. If further non-substantive propositions are countenanced, they will presumably be the platitudes and the howlers whose acceptance and rejection, respectively, are generally taken to reveal little more than an understanding of an area of discourse; these will overlap with the traditional analytic truths and falsehoods but the two categories may not be co-terminous.

The third thing I need to say about the realist's cosmocentric thesis is that it may come in any of a variety of strengths, depending on whether it is maintained *vis-à-vis* individuals or groups – at the limit, the

**Table 1**

	<i>Individual judgement</i>	<i>Community consensus</i>
In actual	1	4
Or normal	2	5
Or ideal circumstances	3	6

community as a whole – and depending on how the circumstances of those individuals and groups are specified. It is one thing to say that an individual may fall into error or ignorance, it is something much stronger to say that the community as a whole may do so. It is one thing to say that individual or community may, in their actual circumstances, fall into error or ignorance; it is something much stronger to say that they may do so in normal or even in ideal circumstances. Normal circumstances will be ones in which certain obstacles are lacking, ideal circumstances will be ones in which certain desirable facilities are present: say, all the relevant evidence is available. The strongest version of the realist thesis says that ignorance and error are possible for any of the epistemic combinations represented in the six boxes in Table 1.

In conclusion, a summary. The realist about any area of discourse asserts three theses, setting himself against three different kinds of opponent. Marking his opposition to reductivists, instrumentalists and the like, he asserts that the discourse introduces distinctive posits; this is the descriptivist thesis. Marking his opposition to error theorists and idealists, he holds that the objects posited exist and are independent of people's dispositions to assert and believe things about them; this is the objectivist thesis. Finally, taking his stand against the many varieties of anthropocentric, he maintains the cosmocentric thesis that participants may be in error or ignorance with regard to any or all substantive propositions in the discourse.

*See also* IDEALISM; NOUMENAL/PHENOMENAL; OBJECTIVITY; PRIMARY AND SECONDARY QUALITIES; TRUTH.

## BIBLIOGRAPHY

- Blackburn, S.: *Spreading the Word* (Oxford: Oxford University Press, 1984).
- Davidson, D.: *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984).
- Devitt, M.: *Realism and Truth* (Oxford: Blackwell, 1984).
- Goodman, N.: *Ways of Worldmaking* (Brighton: Harvester, 1978).
- Mackie, J. L.: *Ethics: Inventing Right and Wrong* (Harmondsworth: Penguin, 1977).
- Papineau, D.: *Reality and Representation* (Oxford: Blackwell, 1987).
- Pettit, P.: "The Reality of Rule-following," *Mind* 99 (1990), 1–21.
- Price, H.: *Facts and the Function of Truth* (Oxford: Blackwell, 1988).
- Putman, H.: *Reason, Truth and History* (Cambridge: Cambridge University Press, 1981).
- Rorty, R.: *Philosophy and the Mirror of Nature* (Oxford: Blackwell, 1980).
- Smart, J. J. C.: "Metaphysical realism," *Analysis* 42 (1982), 1–3.
- Tennant, N.: *Anti-realism and Logic* (Oxford: Oxford University Press, 1987).
- Van Fraassen, B.: *The Scientific Image* (Oxford: Oxford University Press, 1980).

PHILIP PETTIT

**reasons/causes** The distinction between reasons and causes is motivated in good part by a desire to separate the rational from the natural order. Historically, it probably traces back at least to Aristotle's similar (but not identical) distinction between final and efficient cause. Recently, the contrast has been drawn primarily in the domain of actions and, secondarily, elsewhere.

Many who have insisted on distinguishing reasons from causes have failed to distinguish two kinds of reason. Consider my reason for sending a letter by express mail. Asked why I did so, I might say I wanted to get it there in a day, or simply: to get it there in a day. Strictly, the *reason* is expressed by "to get it there in a day". But what this expresses is *my* reason only because I am suitably motivated; I am in a *reason state*: wanting to get the letter there in a day. It is reason states – especially wants, beliefs and intentions – and not reasons strictly so called, that are candidates

for causes. The latter are abstract contents of propositional attitudes; the former are psychological elements that play motivational roles.

If reason states can motivate, however, why (apart from confusing them with reasons proper) deny that they are causes? For one thing, they are not events, at least in the usual sense entailing change; they are dispositional states (this contrasts them with occurrences, but does not imply that they admit of dispositional *analysis*). It has also seemed to those who deny that reasons are causes that the former *justify* as well as explain the actions for which they are reasons, whereas the role of causes is at most to explain. Another claim is that the relation between reasons (and here reason states are often cited explicitly) and the actions they explain is non-contingent, whereas the relation of causes to their effects is contingent. The "logical connection argument" proceeds from this claim to the conclusion that reasons are not causes.

These arguments are inconclusive. First, even if causes are events, *sustaining causation* may explain, as where the (state of) standing of a broken table is explained by the (condition of) support of stacked boards replacing its missing legs. Second, the "because" in "I sent it by express because I wanted to get it there in a day" is in *some* sense causal – indeed, where it is not so taken, this purported explanation would at best be construed as only rationalizing, rather than justifying, my action. And third, *if* any non-contingent connection can be established between, say, my wanting something and the action it explains, there are close causal analogues, such as the connection between bringing a magnet to iron filings and their gravitating to it: this is, after all, a "definitive" connection, expressing part of what it *is* to be magnetic, yet the magnet *causes* the filings to move.

There is, then, a clear distinction between reasons proper and causes, and even between reason states and *event* causes; but the distinction cannot be used to show that the relation between reasons and the actions they justify is in no way causal. Precisely parallel points hold in the epistemic domain (and indeed for all the propositional attitudes, since they all similarly admit of justification,



and explanation, by reasons). Suppose my reason for believing that you received my letter today is that I sent it by express yesterday. My reason, strictly speaking, is *that* I sent it by express yesterday; my reason *state* is my *believing* this. Arguably, my reason justifies the further *proposition* I believe for which it is my reason, and my reason state – my evidence belief – both explains and justifies my *belief* that you received the letter today. I can *say* that what justifies that belief is (the fact) that I sent the letter by express yesterday; but this statement *expresses* my believing that evidence proposition, and indeed if I do not believe it then my belief that you received the letter is not justified: it is not justified by the mere truth of the proposition (and can be justified even if that proposition is false).

Similarly, there are, for belief as for action, at least five main kinds of reason: (1) *normative reasons*, reasons (objective grounds) there are to believe (say, to believe that there is a greenhouse effect); (2) *person-relative normative reasons*, reasons for (say) me to believe; (3) *subjective reasons*, reasons I have to believe; (4) *explanatory reasons*, reasons *why* I believe; and (5) *motivating reasons*, reasons *for which* I believe. (1) and (2) are propositions and thus not serious candidates to be causal factors. The states corresponding to (3) may or may not be causal elements. Reasons why, case (4), are always (sustaining) explainers, though not necessarily even *prima facie* justifiers, since a belief can be causally sustained by factors with no evidential value. Motivating reasons are both explanatory and possess whatever minimal *prima facie* (see PRIMA FACIE REASONS) justificatory power (if any) a reason must have to be a basis of belief.

Current discussion of the reasons-causes issue has shifted from the question whether reason states can causally explain to the perhaps deeper questions whether they can justify *without* so explaining, and what *kind* of causal chain non-waywardly connects reason states with actions and beliefs they do explain. Reliabilists (see RELIABILISM) tend to take a belief as justified *by* a reason only if it is held at least in part *for* that reason, in a sense implying, but not entailed by, being causally based on that reason. Internalists

often deny this, perhaps thinking we lack internal access to the relevant causal connections (see EXTERNALISM/INTERNALISM). But internalists need not deny it, particularly if they require only internal access to *what* justifies – say, the reason state – and not to the (perhaps quite complex) relations it bears to the belief it justifies, by virtue of which it does so. Many questions also remain concerning the very nature of causation, reasonhood, explanation and justification.

See also NATURAL SCIENCE; SOCIAL SCIENCES.

#### BIBLIOGRAPHY

- Alston, W. P.: "Wants, Actions, and Causal Explanations," in *Intentionality, Minds, and Perception*, ed. H.-N. Castaneda (Detroit, MI: Wayne State University Press, 1967).
- Aristotle: *Nicomachean Ethics*.
- Audi, R.: "Acting for Reasons," *Philosophical Review* 95 (1986), 511–46.
- Audi, R.: *Practical Reasoning* (London: Routledge, 1989).
- Brandt, R. B. and Kim, J.: "Wants as Explanations of Actions," *Journal of Philosophy* 60 (1963), 425–35.
- Bratman, M. E.: *Intention, Plans, and Practical Reason* (Cambridge, MA: Harvard University Press, 1987).
- Castaneda, H.-N.: *Thinking and Doing* (Dordrecht: Reidel, 1975).
- Davidson, D.: "Actions, Reasons, and Causes," in his *Essays on Actions and Events* (Oxford: Clarendon Press, 1980).
- Goldman, A. I.: *A Theory of Human Action* (Englewood Cliffs, NJ: Prentice-Hall, 1970).
- Melden, A. I.: *Free Action* (London: Routledge and Kegan Paul, 1961).

ROBERT AUDI

#### Reichenbach, Hans (1891–1953)

Reichenbach was born in Hamburg, Germany, and died in Los Angeles, California. He was a leading figure in the movement known as logical empiricism (see LOGICAL EMPIRICISM).

Reichenbach began his career as a Kantian, but soon rejected that position. In a series of works, the most important being

(1920, 1928, 1956), he argued that Kant's doctrine of the synthetic a priori character of space, time and causality (see KANT) is incompatible with modern science, especially the theory of relativity. His principal argument was that it is possible to cling to Euclidean geometry, but only at the cost of relinquishing normal causality, or to retain normal causality, but only at the cost of relinquishing Euclidean geometry. General relativity precludes the retention of both on any a priori grounds (see GEOMETRY).

Having abandoned Kantianism, Reichenbach remained a steadfast empiricist for the rest of his life. He adopted a verifiability criterion of cognitive meaningfulness, but unlike the logical positivists (see LOGICAL POSITIVISM) of the Vienna Circle he held that probabilistic (as opposed to conclusive) verifiability is sufficient.

Reichenbach's views differed from those of the positivists in other ways as well. He rejected phenomenalism (see PHENOMENALISM), maintaining instead that our knowledge of the world is based on perceptions of physical things, even though such experiences cannot be certified as always veridical. Moreover, he was a realist (see REALISM) with regard to unobservable entities, maintaining that it is possible to have probabilistic knowledge of them. These views were stated in his (1938), a work he regarded as a refutation of logical positivism.

A great deal of Reichenbach's work was devoted to probability and induction. In his major work on the subject (1935, revised in 1949), he presented a mathematical calculus of probability and advocated a limiting frequency interpretation, showing that the axioms of the calculus become arithmetical truisms if that interpretation is adopted. He advanced his "rule of induction" (essentially induction by enumeration) as a method of positing values of limiting frequencies.

Reichenbach took seriously Hume's challenge to induction (see HUME; PROBLEMS OF INDUCTION) and, in answer thereto, offered a pragmatic justification of his rule. Roughly stated, his justification maintained that, although we cannot know whether nature will continue to be uniform, we have everything

to gain and nothing to lose by using induction, for if any method works, induction will. More precisely, he argued that, if sequences of events have limiting frequencies, his rule of induction is bound, sooner or later, to lead to posits that become and remain accurate to any desired degree of accuracy. That result follows immediately from the definition of a limit and the character of his rule. He realized, however, that an infinite class of "asymptotic rules" have the same convergence property, but he never successfully showed how to justify the selection of his rule in preference to any other asymptotic rule.

Reichenbach's epistemology was totally empiricist and fallibilist (see EMPIRICISM; FALLIBILISM). Our factual knowledge is based entirely on experience, and experience does not provide any certain knowledge of the world.

#### WRITINGS

- Relativitätstheorie und Erkenntnis Apriori* (Berlin: 1920); trans. M. Reichenbach as *The Theory of Relativity and A Priori Knowledge* (Berkeley, CA: University of California Press, 1965).  
*Philosophie der Raum-Zeit-Lehre* (Berlin and Leipzig: 1928); trans. M. Reichenbach and J. Freund as *The Philosophy of Space and Time* (New York: Dover, 1958).  
*Wahrscheinlichkeitslehre* (Leiden: 1935); trans. E. H. Hutten and M. Reichenbach as *The Theory of Probability* 2nd edn revised (Berkeley, CA: University of California Press, 1949).  
*Experience and Prediction* (Chicago, IL: University of Chicago Press, 1938).  
*The Direction of Time* (Berkeley, CA: University of California Press, 1956).

#### BIBLIOGRAPHY

- Salmon, W. C.: *Hans Reichenbach: Logical Empiricist* (Dordrecht: Reidel, 1979).

WESLEY C. SALMON

**Reid, Thomas (1710–96)** Thomas Reid was born, worked and died in Scotland. Beginning in 1751, he was a regent and lecturer at King's College in Old Aberdeen

where he founded the Aberdeen Philosophical Society with John Gregory, until, in 1764, he succeeded Adam Smith in the Chair of Moral Philosophy in the Old College at Glasgow. In the same year, 1764, his first book, *Inquiry into the Human Mind on the Principles of Common Sense*, which emerged from his lectures to the Aberdeen Philosophical Society, was published. His other two works, his *Essays on the Intellectual Powers*, 1785, and his *Essays on the Active Powers*, 1788, which resulted from his Glasgow lectures, were published after his retirement.

Thomas Reid is most famous as a defender of commonsense against the scepticism of David Hume (see HUME). Reid's defence of commonsense is based on a sophisticated theory of conception and justification which combines naturalistic nativism and normative epistemology. It is this positive theory rather than his criticism of Hume that constitutes his major contribution. Reid claimed that our most basic conceptions of qualities, for example, movement, figure and colour, result from original faculties of the mind. Faculties are innate powers of the mind. These faculties give rise to conceptions of external objects in immediate response to sensory stimulation and the sensations that accompany such stimulation. Sensations give rise to our conceptions of the qualities of objects of our commonsense world.

Reid denied that such qualities of objects were reducible to sensations, or, for that matter, to the impressions of Hume. A conception is necessarily a conception of something, of some immanent or intentional object. We can conceive of things that do not exist, centaurs, for example, but conception is always about something, even if a non-existent thing. Sensations are only modifications of the mind that have no object.

Conviction as well as conception arises from original principles and irreducible operations of the faculties of the mind. For example, a sensation acts as input to our innate faculty of perception, which gives rise according to the first principles of the faculty to the conception of a moving object and to the immediate and irresistible conviction of the existence of such an object as output. This nativist theory of

conception and belief is joined to Reid's epistemology by a simple connecting premise, namely, that beliefs resulting from first principles are justified beliefs supported by evidence. The justification and evidence of such beliefs is immediate, which is his reply to Hume. Evidence, Reid says, is the ground of belief, and beliefs arising immediately from first principles receive their evidence as a birthright.

This conception of evidence is a form of naturalized foundationalism (see FOUNDATIONALISM), but there is also a coherentist component (see COHERENTISM) to Reid. Suppose we ask why we should trust this evidence arising from the first principles of our faculties. Reid has two answers. The first is that if we do not trust our faculties our situation is desperate and without remedy. We shall be led to total scepticism. Reason, for example, is a faculty, and if we do not trust reason, then it cannot provide us relief from scepticism, and of course the same is true of perception, consciousness and all the rest of our faculties. If, on the other hand, we do trust our faculties, it appears that we should treat them equally, at least initially, and trust them all.

However, it is not just desperation that leads us to trust our faculties. We trust our faculties as a result of a first principle of our faculties, a sort of metaprinciple, to the effect that our faculties are not fallacious but are, instead, trustworthy. Other first principles tell us that the things we distinctly perceive really do exist, that the events we clearly remember did really happen, but the metaprinciple, which is also a first principle, tells us that all the first principles of our faculties, which includes the metaprinciple itself, are trustworthy. The metaprinciple is a source of evidence for the other principles and for itself. Though all first principles yield convictions that are immediately evident, that is, evident without reasoning, the metaprinciple supports the others, and, when the convictions they yield turn out to be true and not fallacious, the metaprinciple is also supported.

Reid's system thus provides us with an epistemology featuring elements of coherence, foundationalism and a natural connection with truth. Nature bestows faculties containing first principles of conception and belief upon

us which are, contrary to Hume, also principles of evidence. We must trust the evidence of our faculties in our practical and scientific concerns or fall defeated into the coalpit of scepticism.

This epistemology is applied to diverse areas by Reid, and one interesting application is to moral epistemology (see MORAL EPISTEMOLOGY). Reid holds that there is a moral faculty. Reid contends that children have moral convictions, such as that of being treated unjustly, before they are taught any moral conceptions. So there is an innate capacity to form moral judgements and, contrary to Hume, judgements of justice and injustice. The faculty responds to our conceptions of actions with judgements as to whether those actions are just or unjust. Differences in moral judgement, Reid contends, are the result of differences in the way in which we conceive of the action. These in turn result from differences in moral education. For any two people, given exactly the same conception of an action as input, the same judgement will result as output.

Reid has important contributions to the theory of human action and liberty as well as to perception, epistemology and morals. His system is based on his fundamental premise of the trustworthiness of the innate faculties of the human mind. A detailed account of his system is contained in Lehrer (1989).

See also COMMONSENSISM AND CRITICAL COGNITIVISM; PRINCIPLE OF CREDULITY; TESTIMONY.

## WRITINGS

*Essays on the Intellectual Powers of Man* (1785), ed. B. Brody (Cambridge, MA: MIT Press, 1969).  
*Thomas Reid's Inquiry and Essays*, ed. K. Lehrer and R. E. Beanblossom (Indianapolis, IN: Bobbs-Merrill, 1975).

## BIBLIOGRAPHY

De Bary, P.: *Thomas Reid and Scepticism: His Reliabilist Response* (London: Routledge, 2002).  
 Lehrer, K.: *Thomas Reid* (London: Routledge, 1989).  
 Nichols, R.: *Thomas Reid's Theory of Perception* (Oxford: Oxford University Press, 2007).

Wolterstorff, N.: *Thomas Reid and the Story of Epistemology* (Cambridge: Cambridge University Press, 2001).

KEITH LEHRER

**reification, hypostatization** We interpret our world as composed of things or objects. As well as everyday external things and the posits of scientific theory, there are more controversial objects like universals, sense-data, classes and numbers. When we *reify* objects we accept their existence as distinct things in the world, and this requires us to be able to count them and quantify over them; it also requires us to understand identity statements involving objects of the appropriate kinds. Claims about the existence of particular kinds of objects can be challenged by arguing that we lack criteria for counting them or for evaluating identity statements, and by denying that they contribute to our understanding of our surroundings.

See also EXISTENCE.

## BIBLIOGRAPHY

Quine, W. V.: *From a Logical Point of View* (Cambridge, MA: Harvard University Press, 1953), ch. 4.  
 Quine, W. V.: *Ontological Relativity and Other Essays* (New York: Columbia University Press, 1969), ch. 1.  
 Quine, W. V.: *Theories and Things* (Cambridge, MA: Harvard University Press, 1981), ch. 1.  
 Quine, W. V.: *Pursuit of Truth* (Cambridge, MA: Harvard University Press, 1990), ch. 2.

CHRISTOPHER HOOKWAY

**relativism** Epistemological relativism may be defined as the view that knowledge (and/or truth) is relative – to time, to place, to society, to culture, to historical epoch, to conceptual scheme or framework, or to personal training or conviction – so that what counts as knowledge depends upon the value of one or more of these variables. If knowledge and truth are relative in this way, this will be because different cultures, societies, etc.,

accept different sets of background principles and standards of evaluation for knowledge-claims, and there is no neutral way of choosing between these alternative sets of standards. So the relativist's basic claim is that the truth and rational justifiability of knowledge-claims are relative to the standards used in evaluating such claims, and that there is no overarching meta-standard from the perspective of which such standards can themselves be fairly judged. (For a more technical definition, see Siegel, 1987, p. 6.)

The doctrine of relativism is usually traced to Protagoras, who is portrayed in Plato's *Theaetetus* as holding that "man is the measure of all things" ("homo mensura"), and that any given thing "is to me such as it appears to me, and is to you such as it appears to you" (152a). Plato's Socrates characterizes Protagorean relativism as consisting in the view that "what seems true to anyone is true for him to whom it seems so" (*Theaetetus*, 170a). This view is a form of relativism in our sense, since for Protagoras there is no standard higher than the individual with reference to which claims to truth and knowledge can be adjudicated. But relativism as defined above is more general than Protagorean relativism, for it places the source of relativism at the level of standards rather than at the level of personal opinion or perception, and as such aptly characterizes more recent versions of relativism.

Opponents of relativism have made many criticisms of the doctrine; by far the most fundamental is the charge that relativism is *self-referentially incoherent*, in that defending the doctrine requires one to give it up. There are several versions of the incoherence charge. The most powerful (for others, see Siegel, 1987, 2004) is that relativism precludes the possibility of determining the truth, warrant or epistemic merit of contentious claims and doctrines – including itself – since according to relativism no claim or doctrine can fail any test of epistemic adequacy or be judged unjustified, false or unwarranted. Take Protagorean relativism as an example. If "what seems true [or warranted] to anyone is true [or warranted] for him to whom it seems so", then no sincere claim can fail any test of

epistemic adequacy or be judged unjustified or false. But if there is no possibility that a claim or doctrine can fail a test of epistemic adequacy or rightness, then the distinction between adequacy and inadequacy, rightness and wrongness is given up. If so, then the very notions of rightness, truth and warrantedness are undermined. But if this is so, then relativism itself cannot be right, true or warranted. In short: relativism is incoherent because, if it is right, the very notion of rightness is undermined, in which case relativism itself cannot be right. The assertion *and defence* of relativism requires one to presuppose neutral standards in accordance with which contentious claims and doctrines can be assessed; but relativism denies the possibility of evaluation in accordance with such neutral standards. Thus, the doctrine of relativism cannot be coherently defended – it can be defended only by being given up. Relativism is thus impotent to defend itself, and falls to this fundamental reflexive difficulty (Siegel, 1987, ch. 1; 2004).

A further difficulty worth noting is that concerning the notion of *relative truth*. Many versions of relativism rely on such a notion, but it is very difficult to make sense of it. An assertion that a proposition is "true for me" (or "true for members of my culture") is more readily understood as a claim about what I (or members of my culture) *believe* than it is as a claim ascribing to that proposition some peculiar form of truth. Moreover, even if this notion could be made sense of, the doctrine would still fall to the incoherence argument above (Siegel, 1987, ch. 1).

Despite these ancient and powerful responses to relativism, the last several decades have witnessed a resurgence of the doctrine. This is at least in part due to the difficulty of formulating a defensible conception of non-relativism. Many relativists argue for relativism on the grounds that any non-relativistic alternative will require repugnant epistemological commitments, e.g. to certainty (see CERTAINTY), privileged frameworks, or dogmatism (see DOGMATISM). The challenge to opponents of relativism ("absolutists") is to develop a non-relativistic epistemology which includes an acceptable account of



rationality (see RATIONALITY) and rational justification, which is fallibilistic and non-dogmatic, which rejects any notion of a privileged framework in which knowledge-claims must be couched, and which is self-referentially coherent (see Siegel, 1987, ch. 8; 2004) (see TRANSCENDENTAL ARGUMENTS; PROBLEM OF THE CRITERION).

Contemporary versions of relativism occur in a wide variety of philosophical contexts and enjoy an equally wide variety of philosophical pedigrees. Chief among them are versions of relativism spawned by Wittgensteinian considerations concerning language use, conceptual schemes or frameworks, and “forms of life” (Wilson, 1970) (see WITTGENSTEIN); proponents of the strong programme (see STRONG PROGRAMME) in the sociology of knowledge (see Barnes and Bloor, in Hollis and Lukes, 1982) (see SOCIOLOGY OF KNOWLEDGE); a variety of quite different positions which might be grouped together under the heading of “contemporary neo-Pragmatism” (e.g. Rorty, 1979, 1982; see also Goodman, 1978; Putnam, 1981) (see DEATH OF EPISTEMOLOGY); and, perhaps most surprisingly, work in philosophy of science (Kuhn, 1970; Feyerabend, 1975). These and other contemporary versions of relativism make clear that the doctrine is alive and well, and is the subject of intense philosophical debate, as philosophers sympathetic to relativism attempt to develop versions of the doctrine which are immune to the standard criticisms. Of course, philosophers who are unsympathetic to the doctrine continue to press traditional and more recently developed objections to it. The current scene is, then, one in which interest in relativism remains high.

See also ONTOLOGICAL RELATIVITY; SUBJECTIVISM.

# BIBLIOGRAPHY

- Baghramian, M.: *Relativism* (London: Routledge, 2004).
- Feyerabend, P.: *Against Method: Outline of an Anarchist Theory of Knowledge* (London: Verso, 1975).
- Goodman, N.: *Ways of Worldmaking* (Indianapolis, IN: Hackett, 1978).

- Hollis, M. and Lukes, S., eds: *Rationality and Relativism* (Cambridge, MA: MIT Press, 1982).
- Krausz, M., ed.: *Relativism: Interpretation and Confrontation* (Notre Dame, IN: University of Notre Dame Press, 1989).
- Kuhn, T. S.: *The Structure of Scientific Revolutions*, 2nd edn (Chicago, IL: University of Chicago Press, 1970).
- Meiland, J. W. and Krausz, M., eds: *Relativism: Cognitive and Moral* (Notre Dame, IN: University of Notre Dame Press, 1982).
- Mosteller, T.: *Relativism in Contemporary American Philosophy* (London: Continuum, 2006).
- Plato: *Theaetetus*, trans. F. M. Cornford, in *The Collected Dialogues of Plato*, ed. E. Hamilton and H. Cairns (New York: Pantheon, 1961), 845–919.
- Putnam, H.: *Reason, Truth and History* (Cambridge: Cambridge University Press, 1981).
- Rorty, R.: *Philosophy and the Mirror of Nature* (Princeton, NJ: Princeton University Press, 1979).
- Rorty, R.: *Consequences of Pragmatism* (Minneapolis, MN: University of Minnesota Press, 1982).
- Siegel, H.: *Relativism Refuted: A Critique of Contemporary Epistemological Relativism* (Dordrecht: Reidel, 1987).
- Siegel, H.: “Relativism,” in *Handbook of Epistemology*, ed. I. Niiniluoto, M. Sintonen and J. Woleński (Dordrecht: Kluwer, 2004), 747–80.
- Wilson, B. R., ed.: *Rationality* (London: Blackwell, 1970).

HARVEY SIEGEL

**relevant alternatives** The theory of relevant alternatives is best viewed as an attempt to accommodate two opposing strands in our thinking about knowledge. The first is that knowledge is an *absolute* concept. On one interpretation, this means that the justification or evidence one must have in order to know a proposition *p* must be sufficient to eliminate all the alternatives to *p* (where an alternative to a proposition *p* is a proposition incompatible with *p*). That is, one’s justification or evidence for *p* must be sufficient for one to know that every alternative to *p* is false. This element of our thinking about knowledge is exploited by sceptical arguments. These arguments call our attention to alternatives that our evidence can not eliminate. For example (Dretske, 1970), when we are at the zoo, we might claim to

know that we see a zebra on the basis of certain visual evidence, viz. a zebra-like appearance. The sceptic inquires how we know that we are not seeing a cleverly disguised mule. While we do have some evidence against the likelihood of such a deception, intuitively it is not strong enough for us to *know* that we are not so deceived. By pointing out alternatives of this nature that we cannot eliminate, as well as others with more general application (dreams, hallucinations, etc.), the sceptic appears to show that this requirement that our evidence eliminate every alternative is seldom, if ever, satisfied (see SCEPTICISM).

This conclusion conflicts with another strand in our thinking about knowledge, viz. that we know many things. Thus there is a tension in our ordinary thinking about knowledge – we believe that knowledge is, in the sense indicated, an absolute concept and yet we also believe that there are many instances of that concept.

There would seem to be two options for removing this tension, each involving the denial of one of the components on the basis of the other. If one finds absoluteness to be too central a component of our concept of knowledge to be relinquished, one could argue from the absolute character of knowledge to a sceptical conclusion (Unger, 1975). Most philosophers, however, have taken the other course, choosing to respond to the conflict by giving up, perhaps reluctantly, the absoluteness criterion. This latter response holds as sacrosanct our commonsense belief that we know many things (Pollock, 1974; Chisholm, 1977). Each approach is subject to the criticism that it preserves one aspect of our ordinary thinking about knowledge at the expense of denying another (see CHISHOLM; COMMONSENSISM AND CRITICAL COGNITIVISM).

The theory of relevant alternatives can be viewed as an attempt to provide a more satisfactory response to this tension in our thinking about knowledge. It attempts to characterize knowledge in a way that preserves both our belief that knowledge is an absolute concept and our belief that we have knowledge.

According to the theory, we need to qualify rather than deny the absolute character

of knowledge. We should view knowledge as absolute, relative to a certain standard (Dretske, 1981; Cohen, 1988). That is to say, in order to know a proposition, our evidence need not eliminate all the alternatives to that proposition. Rather we can know when our evidence eliminates all the *relevant* alternatives, where the set of relevant alternatives (a proper subset of the set of all alternatives) is determined by some standard. Moreover, according to the relevant alternatives view, the standards determine that the alternatives raised by the sceptic are not relevant. If this is correct, then the fact that our evidence can not eliminate the sceptic's alternatives does not lead to a sceptical result. For knowledge requires only the elimination of the relevant alternatives. So the relevant alternatives view preserves both strands in our thinking about knowledge. Knowledge is an absolute concept, but because the absoluteness is relative to a standard, we can know many things.

The relevant alternatives account of knowledge can be motivated by noting that other concepts exhibit the same logical structure. Two examples of this are the concept *flat* and the concept *empty* (Dretske, 1981). Both appear to be absolute concepts – a space is empty only if it does not contain anything and a surface is flat only if it does not have any bumps. However, the absolute character of these concepts is relative to a standard. In the case of *flat*, there is a standard for what counts as a bump and in the case of *empty*, there is a standard for what counts as a thing. We would not deny that a table is flat because a microscope reveals irregularities in its surface. Nor would we deny that a warehouse is empty because it contains particles of dust. To be flat is to be free of any relevant bumps. To be empty is to be devoid of all relevant things. Analogously, the relevant alternatives theory says that to know a proposition is (*inter alia*) to have evidence that eliminates all relevant alternatives.

Some philosophers (Dretske, 1970) have argued that the relevant alternatives theory of knowledge entails the falsity of the principle that the set of known (by S) propositions is closed under known (by S) entailment; although others have disputed this (Stine,

1976; Cohen, 1988). This principle affirms the following conditional (the *closure principle*):

If S knows *p* and S knows that *p* entails *q*, then S knows *q*.

According to the theory of relevant alternatives, we can know a proposition *p*, without knowing that some (non-relevant) alternative to *p* is false. But since an alternative *h* to *p* is incompatible with *p*, then *p* will trivially entail not-*h*. So it will be possible to know some proposition without knowing another proposition trivially entailed by it. For example, we can know that we see a zebra without knowing that it is not the case that we see a cleverly disguised mule (on the assumption that "we see a cleverly disguised mule" is not a relevant alternative). This will involve a violation of the closure principle. This is an interesting consequence of the theory because the closure principle seems to many to be quite intuitive. In fact, we can view sceptical arguments as employing the closure principle as a premiss, along with the premiss that we do not know that the alternatives raised by the sceptic are false. From these two premisses, it follows (on the assumption that we see that the propositions we believe entail the falsity of sceptical alternatives) that we do not know the propositions we believe. For example, it follows from the closure principle and the fact that we do not know that we do not see a cleverly disguised mule, that we do not know that we see a zebra. We can view the relevant alternatives theory as replying to the sceptical arguments by denying the closure principle.

What makes an alternative relevant? What standard do the alternatives raised by the sceptic fail to meet? These questions have been notoriously difficult to answer with any degree of precision or generality. This difficulty has led critics to view the theory as *ad hoc* or obscure (Sosa, 1988). The problem can be illustrated through an example (Goldman, 1976). Suppose Smith sees a barn and believes that he does, on the basis of very good perceptual evidence. When is the alternative that Smith sees a papier-mâché replica relevant? If there are many such replicas in the immediate area, then this alternative can be relevant. In

these circumstances, Smith fails to know that he sees a barn unless he knows that it is not the case that he sees a barn replica. Where no such replicas exist, this alternative will not be relevant (*ceteris paribus*). Smith can know that he sees a barn without knowing that he does not see a barn replica.

This suggests that a criterion of relevance is something like probability conditional on Smith's evidence and certain features of the circumstances. But which circumstances in particular do we count? Consider a case where we want the result that the barn replica alternative is clearly relevant, e.g. a case where the circumstances are such that there are numerous barn replicas in the area. Does the suggested criterion give us the result we wanted? The probability that Smith sees a barn replica given his evidence and his location in an area where there are many barn replicas is high. However, that same probability conditional on his evidence and his particular visual orientation toward a real barn is quite low. We want the probability to be conditional on features of the circumstances like the former but not on features of the circumstances like the latter. But how do we capture the difference in a general formulation?

How significant a problem is this for the theory of relevant alternatives? This depends on how we construe the theory. If the theory is supposed to provide us with an analysis of knowledge, then the lack of precise criteria of relevance surely constitutes a serious problem. However, if the theory is viewed instead as providing a response to sceptical arguments, then it can be argued that the difficulty has little significance for the overall success of the theory (Cohen, 1988; forthcoming).

See also CAUSAL THEORIES IN EPISTEMOLOGY; GETTIER PROBLEM; RELIABILISM; CONTEXTUALISM in Part I.

## BIBLIOGRAPHY

- Chisholm, R. M.: *Theory of Knowledge*, 2nd edn (Englewood Cliffs, NJ: Prentice-Hall, 1977).  
Cohen, S.: "How to Be a Fallibilist," *Philosophical Perspectives* 2 (1988), 91–123.

- Cohen, S.: "Scepticism, Relevance, and Relativity," in *Critical Essays on the Philosophy of Fred Dretske*, ed. B. McGlaughlin (forthcoming).
- Dretske, F.: "Epistemic Operators," *Journal of Philosophy* 69 (1970), 1007–23.
- Dretske, F.: "The Pragmatic Dimension of Knowledge," *Philosophical Studies* 40 (1981), 363–78.
- Goldman, A.: "Discrimination and Perceptual Knowledge," *Journal of Philosophy* 73 (1976), 771–91.
- Pollock, J.: *Knowledge and Justification* (Princeton, NJ: Princeton University Press, 1974).
- Sosa, E.: "Knowledge in Context, Scepticism in Doubt," *Philosophical Perspectives* 2 (1988), 139–56.
- Stine, G. C.: "Scepticism, Relevant Alternatives, and Deductive Closure," *Philosophical Studies* 29 (1976), 249–61.
- Unger, P.: *Ignorance* (Oxford: Clarendon Press, 1975).

STEWART COHEN

**reliabilism** An approach to epistemology that explains key epistemic concepts in terms of the truth-conduciveness of a person's reasons, belief-forming processes, methods, faculties, or the like. An item's truth-conduciveness is its tendency to generate true beliefs and avoid erroneous ones. There is reliabilism about knowledge and reliabilism about justification; there are reliable-indicator theories and reliable-process theories. However, the most common referent of the term "reliabilism" is the process-reliabilist theory of justification, so it will be the primary topic of this article.

## 1. KNOWLEDGE AND RELIABILITY

We begin with reliabilism about knowledge. Epistemologists generally agree that S knows that p only if S believes that p and p is true. Reliabilism is a distinctive approach to knowledge that adds a reliability requirement on top of true belief. This can be done in three ways. First, one might include a condition to the effect that how the belief is formed – the process, method, or faculty used in producing it – must be *globally* reliable, where *global* reliability refers to the overall reliability of

the process, method, or faculty. Second, one might include a condition to the effect that the belief, or the method that produces it, is reliable in a *situation-specific* sense, where situation-specific reliability refers to getting truth or avoiding error in possible situations closely related to the actual one. Third, a reliability theory of knowing might include both of the foregoing kinds of conditions: global and local reliability.

F. P. Ramsey (1931) gave the earliest formulation of a reliability theory of knowing, saying that a belief is knowledge if it is true, certain, and obtained by a reliable process. Presumably he meant by "reliable" what we are calling generic reliability. D. M. Armstrong's (1973) theory of knowledge analyzed (non-inferential) knowledge in terms of a nomological connection between some properties of the target belief and its truth. This theory had a definite element of generality, but this element wasn't located in the belief-generating process or method. Alvin Goldman (1986) offered a theory of knowing that featured requirements of both global and situation-specific reliability. The majority of reliability theories of knowing, however, focus on situation-specific reliability. What I am calling situation-specific reliability theories include sensitivity theories, safety theories, and no-relevant-alternatives theories, which all highlight subjunctive or counterfactual conditionals relating belief and truth, or belief and error-avoidance.

A sensitivity theory says that a person S knows that p only if S wouldn't believe p if p were false. In other words, in a situation closest to the actual one in which p is false, S would not believe p (hence would not believe it falsely). Robert Nozick (1981) endorsed such a condition in his "tracking" theory of knowing. Nozick also offered a more refined version of the sensitivity condition according to which S knows that p only if S wouldn't believe p (if p were false) via the same method of belief-formation used in the actual case. A safety condition on knowing has been endorsed by Ernest Sosa (1996) and Timothy Williamson (2000). Safety can be formulated as "if S believes that p, then p could not easily have been false", or "if S believes that p, then

p isn't false in close possible worlds". (This is not equivalent to sensitivity.) A more careful safety condition can be given by adding a qualification about the basis for belief: S knows that p on basis B only if p isn't false in close possible worlds where S believes p on basis B (Williamson, 2000: 128). Under a basis-qualified safety condition, knowing is compatible with having false belief in a close case on a very different basis for belief than the one used in the actual world. Similarly, a no-relevant-alternatives condition says (roughly) that S knows that p only if there is no relevant situation close to the actual one in which p is false but S believes that p on the same basis as he believes it in the actual world (Goldman, 1976). Fred Dretske (1971) formulated the earliest version of a "reliable basis" theory, saying that a person knows that p only if the reasons on which he bases his belief that p in the actual situation are reasons he would not have if p were false. Others who have defended "reliable indicator" theories of knowledge or justification include Marshall Swain (1981) and William Alston (1988).

Let us illustrate how these reliability theories would work when applied to a familiar adequacy test for theories of knowing: a Gettier case. Edmund Gettier (1963) presented a case in which Smith has excellent evidence that Jones owns a Ford. Smith believes this proposition and deduces from it the additional, disjunctive proposition p: either Jones owns a Ford or Brown is in Barcelona. As it happens, Jones does not own a Ford. By coincidence, however, Brown is in Barcelona (though Smith has no information about this), so the disjunctive proposition p is true. Does Smith know p? Intuitively, he doesn't know it, so any adequate theory must explain why not. The tracking theory explains it by saying that the sensitivity condition is violated. In one of the closest possible worlds in which the disjunctive proposition p is false, Brown isn't in Barcelona; but Smith still believes p because he has the same (misleading) evidence about Jones. So it's not the case that Smith *wouldn't* believe p (on the same basis) if p were false. The safety theory explains it by saying that Smith could easily

have believed p falsely because there is a close possible world in which he believes it but it is false, namely one in which Brown isn't in Barcelona and Jones doesn't own a Ford. The no-relevant-alternatives theory explains it by saying that there is a relevant alternative situation in which p is false but Smith has the same basis for believing p, a situation in which his evidence about Jones stays the same but Brown isn't in Barcelona. All these reliability theories acquit themselves pretty well on the Gettier test.

This isn't to say that these theories are free of difficulties. We shall not pause to examine these difficulties but simply note that reliability theories derive their attractiveness in part from their ability to handle Gettier-type cases successfully.

## 2. RELIABILISM ABOUT JUSTIFICATION: A FIRST PROPOSAL

The most discussed version of justificational reliabilism is the reliable-process approach, first formulated by Alvin Goldman in "What Is Justified Belief?" (1979). In a lead-up to his account, Goldman proposes some constraints or ground rules for theories of justification. The proposal is that such theories should specify non-epistemic conditions for the epistemic status of being justified. This is analogous to the aim of normative ethics to provide non-moral conditions for an action to be morally right. In epistemology one doesn't want a theory of justifiedness to rely (ultimately) on epistemic properties or concepts such as "knowing", "being reasonable", or "being rational". Appealing to justification in an account of justification would obviously be circular. But an appeal to other epistemic concepts would not be much better, because the entire family of epistemic concepts needs illumination. This still leaves many types of concepts available to feature in the analysis: (1) psychological concepts such as belief and experience; (2) metaphysical concepts such as causation; and (3) relations between propositions such as logical deducibility, probabilistic coherence, and degrees of confirmation or support.



In the decades leading up to the 1970s, most approaches to epistemology concentrated on the first and third categories just delineated. In foundationalist epistemologies, perceptual experience and memorial experiences played a crucial role in accounts of directly justified propositions (Chisholm, 1966, 1977), and theories of logical confirmation aimed to show how indirectly justified propositions can be confirmed or supported by directly justified ones (Carnap, 1950; Hempel, 1945). Coherentist theories, including subjective Bayesianism, aimed to show how justifiedness or rationality is a function of logical and/or probabilistic relations between the propositional contents of doxastic states – either categorical beliefs or degrees of belief (Lehrer, 1974; Bonjour, 1985; Jeffrey, 1983). None of these kinds of theories assigned any substantial role to causal relations between an agent's beliefs, or to causal processes responsible for belief production. Indeed, epistemologists of a positivist stripe explicitly opposed the idea that the causal provenance of a belief has anything to do with its justificational status. Hans Reichenbach (1938) drew a sharp distinction between the “context of discovery” and the “context of justification”. Discovery was a proper subject-matter for psychology, but could shed no light on justification. Justification was presumed to be concerned with reasons and logic, which had nothing to do with psychological causes.

In “What Is Justified Belief?”, however, Goldman argued that the psychological processes that cause a belief are critical to its justificational status. This position was a natural successor to the causal theory of knowing that Goldman (1967) had advanced a decade earlier. The rationale for the view about justification can be illustrated as follows. Suppose Reginald begins with a large corpus of justified beliefs including belief in the following four propositions: (i)  $p \rightarrow q$ , (ii)  $q$ , (iii)  $r \rightarrow p$ , and (iv)  $r$ . Reginald thereupon adopts a belief in  $p$ . Is he justified in holding this belief? A belief in  $p$  indeed seems justified for Reginald. After all,  $p$  is deducible from a pair of other propositions that he already justifiably believes, namely,  $r \rightarrow p$  and  $r$ .

However, suppose the story specifies that Reginald fails to retrieve from memory either the belief that  $r \rightarrow p$  or the belief that  $r$ ; nor does he notice that those propositions jointly entail  $p$ . No psychological state of his concerning  $r$  makes a causal contribution to his coming to believe  $p$ . Reginald does take notice, however, of what he presumes to be a support relation between the conjunction of  $q$  and  $p \rightarrow q$  and the proposition  $p$ . Reginald's psychological equipment includes a pseudo-deductive operation of “affirming the consequent”. He uses this defective operation to “derive”  $p$  from the premises  $q$  and  $p \rightarrow q$ . This is *how* he comes to believe  $p$ , or *why* he believes it. Intuitively, then, Reginald's belief in  $p$  is not justified. At least it's not justified as long as this is the only process that causally sustains his believing that  $p$ . This shows that the causal processes involved in belief production are critical determinants of justifiedness.

Other examples lend further support to this thesis. If one forms beliefs by defective processes like wishful thinking, hasty generalization, or mere hunch or guesswork, these beliefs are not justified. If one hears somebody utter a sentence  $H$  and comes to believe what  $H$  expresses merely because one likes the sound of  $H$  (Kornblith, 1980), this belief is not justified. Thus, the mode of belief-formation (i.e. belief-causation) is crucial to its justificational status. This is now usually called the “basing relation”.

Suppose, then, that the relevance of causal processes to justification is granted. What property or properties of a belief-producing process makes it epistemically good or bad? Goldman (1979) hypothesized that the truth-conduciveness of a process accounts for its justification-conferring or non-justification-conferring power. A belief-generating process that is highly reliable – that is, has a high truth-ratio – confers justifiedness on its outputs; a belief-generating process that isn't highly reliable confers unjustifiedness on its outputs. For example, perceptual processes are generally reliable and they generally confer justifiedness. Memory processes are pretty reliable and they generally confer justifiedness. By contrast, all the belief-forming processes mentioned in

the preceding paragraph are quite unreliable, and their belief outputs, as indicated, are intuitively unjustified.

When it comes to reasoning processes, matters are a bit more complex. We shouldn't say that a reasoning process is justification-conferring only if its total belief outputs have a high ratio of truths. Think of a reasoning process as a function that maps input beliefs into output beliefs. Suppose that a large number of the process's input beliefs are false and so a largish proportion of its output beliefs are false as well. This shouldn't count against the justification-conferring power of such a process. What we need here is the notion of *conditional reliability*. A reasoning process is epistemically good if it is conditionally reliable: when applied to input beliefs that are all true, the beliefs it outputs are usually true. Reliabilism is strengthened by including both conditional and unconditional notions of reliability. (How to extend these notions to degrees of belief, or subjective probabilities, is an open question.)

The discussion thus far has focused on the justificational status of beliefs actually held by a cognizer. But epistemologists often use a second notion of justification. They want to say that a person's epistemic situation makes her justified in believing a proposition even if she doesn't actually adopt an attitude of belief towards it. This is the *propositional* sense of justification as contrasted with the *doxastic* sense we have been pursuing until now. Can process reliabilism capture the propositional sense of justification? The *prima facie* problem is that, in the propositional sense, there can be justification without actual belief, and if there is no actual belief there is no question of what process causes it. So how does process reliabilism get any traction on the question?

"What Is Justified Belief?" addressed this problem by talking about the hypothetical use of a belief-forming operation. Suppose some reliable belief-forming operation *O* is available to *S* such that if *S* applied *O* to her total cognitive state at time *t*, then *O* would believe *p* at time *t*-plus delta and this belief would be *doxastically* justified. Then *S*'s total cognitive state at *t* makes *S* *propositionally*

justified in believing *p*. If no reliable operation is available to *S* at *t* that meets these conditions, then *S* isn't propositionally justified in believing *p*.

At this point it is instructive to consider some advantages of process reliabilism. One advantage is that it deftly circumvents traditional problems associated with justifying external world beliefs. The traditional problematic is how one can make defensible inferences from propositions about one's own mental states to conclusions about the external world. Under reliabilism, however, no such inferences are required in order to have justified external-world beliefs. All that is needed to have justified beliefs of this sort is to use reliable perception-based processes to arrive at such beliefs.

Another advantage of process reliabilism is its ability to account for sticky issues concerning which propositions, exactly, can be justifiably believed immediately, or non-inferentially. Richard Feldman (2003) discusses an example in which you enter a room and see an ordinary table, which happens to be exactly twelve years old. Presumably, you are (propositionally) justified in believing that it's a table but not justified in believing that it's twelve years old. Although Feldman tries to account for this difference, he seems to recognize that his own, non-reliabilist resources cannot adequately account for it. By contrast, process reliabilism handles it pretty straightforwardly. You will have in your repertoire a reliable belief-forming operation that, if applied to your visual experience, would yield a belief that that's a table. But no reliable belief-forming operation in your possession would yield a belief that that's a twelve-year-old table (for detailed argumentation, see Goldman, 2008). So process reliabilism does a good job in resolving this epistemological puzzle.

### 3. PROBLEMS FOR EARLY PROCESS RELIABILISM

Objections were raised against process reliabilism from the start. In the early going three

problems were highlighted. The first is that reliability isn't necessary for justifiedness. The second is that reliability isn't sufficient for justifiedness. The third is that it's unclear what the reliability of a token belief-forming process is unless and until we are told which process *type* – of the many available types – represents the token's reliability. Without a procedure for type-selection, the theory's verdict on any token belief's justificational status is indeterminate. But no such procedure has been provided.

The most popular illustration of the non-necessity problem is the evil-demon case. Descartes' original evil-demon case posed a problem of skepticism. The present case aims to show that a belief can be justified despite being caused by an unreliable process. Suppose someone inhabits a world in which an evil demon continually deceives him with perceptual appearances that don't correspond to real things in the physical world. The demon similarly deceives other denizens of the world (if any). Then the ratio of true beliefs to total beliefs generated by the perception-based belief-forming processes in that world will be very low. According to reliabilism, it would seem, the perception-based beliefs in that world are all unjustified. Intuitively, though, they seem to be justified (Cohen, 1984). Suppose that one person's experiences in the demon world are "qualitatively" exactly like your experiences in the actual world. Surely that person will be as justified as you are in holding the appropriate beliefs. So says the critic. But this ostensibly conflicts with reliabilism's verdict.

This counter-example assumes that, under reliabilism, ordinary perceptual belief-forming processes are assigned different truth-ratios in different worlds: low truth-ratios in the demon world and high truth-ratios in the actual world. Reliabilism is not committed to this position, however. It isn't clear which domain of evaluation should be chosen for fixing a truth-ratio. In assigning justificational status to a hypothetical belief, it may not be the truth-ratio of the process in the believer's world that is relevant. Reliabilism might instead choose the world of the justification attributor (the actual world)

as the domain of evaluation. A process that is actual-world reliable may confer justification not only in the actual world but also in every possible world. In other words, justification-conferral might be *rigidified*. A rigidified version of reliabilism would handle the demon-world case straightforwardly.

The non-sufficiency criticism of reliabilism is supported by examples intended to show that not all reliable belief-forming processes confer justification. Laurence Bonjour (1980) gives an example of a reliable clairvoyance faculty, which produces true beliefs about remote states of affairs unaccompanied by any perceptual states. Bonjour considers individuals with this clairvoyant faculty who never independently verify the accuracy of their beliefs, and have no reason to believe that anyone, themselves included, has such a faculty. These clairvoyant beliefs aren't justified, says Bonjour, despite being caused by a reliable process. Keith Lehrer (1990) gives a similar example. A perfectly reliable thermometer is implanted in Mr Truetemp's brain. The device detects the ambient temperature and sends messages to his brain causing him to have accurate temperature beliefs. Since Truetemp is unaware of the implant (performed without his knowledge), Lehrer says that his temperature beliefs do not constitute knowledge, and therefore (presumably) aren't justified, despite being reliably caused.

What would be needed to make Truetemp's temperature beliefs justified? Lehrer provides the following hint: "[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 ignorance of this" (1990: 162). This suggests the view that justification would accrue to Truetemp's temperature beliefs only if they were accompanied by a "meta-belief" to the effect that they are caused by a reliable process. Can the problem be solved by requiring the specified meta-belief? Here we must ask whether the presence of a meta-belief would create justification in the first-order beliefs no matter how the meta-belief is caused and no matter whether it is justified. Suppose Truetemp has the requisite meta-belief but it, too, is caused by the brain implant. Would this

make the temperature beliefs justified? That seems unmotivated. If reliable causation doesn't suffice for justifiedness, the meta-belief itself would be unjustified. Why would an unjustified meta-belief make the lower-level belief justified?

Critics of reliabilism might respond: "OK, we have to require not merely that there is a meta-belief, but that the meta-belief is *justified*." This requirement, however, would violate the ground rules sketched earlier for an admissible theory of justification. Under these ground rules, conditions for justifiedness should not include, or presuppose, justifiedness. Let us make the admissibility constraints a bit more explicit (following the proposals of "What Is Justified Belief?"). A theory can avoid circularity by taking a recursive form. Such a theory would contain one or more base clauses, each having the form "If —, then S's belief in p is justified". The antecedent of a base clause must be filled by a condition that makes no appeal to justification (or anything definable in terms of justification). In addition to base clauses, recursive clauses are also admissible. These would be of the form "If S has justified beliefs in q, r, s, etc. . . ., then S's belief in p is justified". This form would be suitable for principles of inferential justification. Finally, the theory would contain a closure clause. For present purposes, the relevant issue concerns base clauses. If the (first-order) beliefs of the clairvoyant or Mr Truetemp are to be justified, those beliefs would presumably be justified in virtue of satisfying some base clause. But under the proposal introduced at the beginning of this paragraph the antecedent of an appropriate base clause will contain the condition that a *justified* meta-belief accompanies the target belief. Such a requirement violates the specifications for a base clause and is therefore inadmissible. Thus, the critics' epistemic ascent strategy doesn't work. In section 5 we shall revisit the problem of non-sufficiency and consider other possible fixes.

The third longstanding problem for reliabilism is the generality problem. Originally noted in "What Is Justified Belief?", the problem has been pressed and elaborated by

Earl Conee and Richard Feldman. When a reliabilist speaks about the reliability of a belief-forming process, she must be referring to a process *type*, because only a type can have a reliability number – at least one that differs from 1 and 0. But global reliabilism certainly expects many belief-forming processes to have intermediate reliability numbers, and these can only attach, in the first instance, to process types. A process token can be assigned a unique reliability number only in virtue of an association with process types. Given that each token instantiates indefinitely many types, of varying generality, how is the appropriate type to be selected?

Conee and Feldman (1998) lay down three requirements for solving the generality problem. A solution must be "principled", must make defensible epistemic classifications, and must remain true to the spirit of reliabilism. They argue that nobody has provided a solution of this sort – appealing either to common-sense types or to scientific types – and that the prospects for such a solution seem bleak.

A fourth problem for reliabilism is the problem of "bootstrapping", or "easy knowledge", pressed by Richard Fumerton (1995), Jonathan Vogel (2000), and Stewart Cohen (2002). These theorists typically formulate it as a problem about knowledge, so our exposition will follow this formulation. The problem is that it is too easy, under reliabilism, to acquire knowledge about the reliability of one's sources. Suppose, for example, that my color vision process is reliable. Then reliabilism implies that I can come to know that the table is red simply by using my color vision, even if I don't know that my color vision is reliable. By combining the fact that my belief was caused by color vision with the knowledge that the table is red, I can conclude that my color vision got matters right on this occasion. And if I repeat this process enough times I can amass enough evidence to know that my color vision is generally reliable. But surely it is an illegitimate procedure to use color vision to establish that color vision is reliable (Cohen, 2002).

A fifth problem for reliabilism is the "extra value of knowledge" problem. It has been proposed as an adequacy test for a theory of

knowledge that it should be capable of explaining why knowing that *p* is more valuable than truly believing that *p*. A number of writers urge that process reliabilism fails this adequacy test (Jones, 1997; Swinburne, 1999; Zagzebski, 1996, 2003; Riggs, 2002; Kvanvig, 2003). Process reliabilism implies that the extra value of knowledge must come from the reliability of the belief-causing process. How can this be? Linda Zagzebski (2003) illustrates the problem via an espresso analogy. If a cup of espresso tastes good, it is valuable; but it gains no additional value by having been produced by a reliable espresso machine. Similarly, if a belief is true, it makes no difference to its value if it stems from a reliable belief-producing source.

#### 4. RESPONSES TO OBJECTIONS AND REVISIONS OF PROCESS RELIABILISM

To meet the foregoing problems, a number of rejoinders and revisions of reliabilism have been proposed (we return here to accounts of justification). In seeking to resolve the demon-world objection to simple reliabilism, Goldman (1986) proposed the “normal worlds” theory. Approaching justification in terms of a right rule system, he sought to delineate a set of possible worlds consistent with our general beliefs about the actual world, beliefs about the kinds of things that, “realistically”, do or can happen. Such worlds were called “normal worlds”. The proposal was that a rule of justification is right just in case it has a sufficiently high truth-ratio in normal worlds, and beliefs (in any world) that comply with such a rule are justified. This idea meshes with assessments of reliability in other contexts. In saying that a car is reliable, one needn’t imply that it starts and runs smoothly in *all* weather conditions (not at –50 degrees Fahrenheit, for example), only that it starts and runs smoothly in “normal” conditions (Heller, 1995). Similarly, a perception-based belief-forming process can be treated as right, and hence as conferring justification, as long as it is reliable in normal worlds; it doesn’t have to be reliable in (abnormal) demon worlds.

The normal-worlds approach is problematic, however, on several counts (see Goldman, 1988). A more recent proposal of Juan Comesaña (2002), called “indexical reliabilism”, may be more promising – an approach he embeds within the semantic theory of two-dimensionalism. As we saw earlier, to say that a belief’s justifiedness hinges on its being produced by a reliable process invites the question: reliable *where*? We might answer, “reliable in the actual world”. But “actual”, argues Comesaña, is itself an indexical notion, and can be interpreted in two different ways. When attributing reliabilist justification to beliefs of merely possible epistemic agents, one might mean they are produced by processes that are reliable in their world or one might mean they are produced by processes that are reliable in the speaker’s world. The latter interpretation allows victims of an evil demon to have justified beliefs even if the processes that produce those beliefs are unreliable in their world.

We turn next to the second problem for reliabilism, represented by the clairvoyance and Truetemp cases. Although one way to strengthen a simple reliability condition is to add a requirement of epistemic ascent (considered but rejected in section 3 above), a weaker supplementary condition might do the job. This is the requirement that a cognizer must *not* have reason to believe that her first-order belief is *unreliably* caused. How would this non-undermining, or non-defeat, condition help with the clairvoyance and Truetemp cases? Plausibly, the agents in such cases *do* have reasons to believe that their first-order beliefs are unreliably caused. BonJour’s clairvoyants (like everybody else) have reasons to think that beliefs that pop out of the blue – as far as one can tell – are unreliably caused. Similarly, Truetemp has reason to think that purely spontaneous beliefs about the precise ambient temperature (beliefs that don’t result from reading thermometers or similar instruments) are unreliably caused. So a modified reliabilist account that incorporates the non-undermining condition would declare these first-order beliefs to be unjustified, just as critics of simple reliabilism have urged.



However, for this maneuver to help reliabilism, “having reason to believe” must be an admissible concept for a base clause. On the surface, the phrase “have reason to believe” seems to mean “be justified in believing”. If so, the current proposal would include a covert justification condition, which runs afoul of the admissibility constraints for a theory of justification. Thus, the maneuver in question needs to be formulated somewhat differently to avoid this problem.

Another version of process reliabilism might also cope with the non-sufficiency problem, though in an entirely different fashion. “Epistemic Folkways and Scientific Epistemology” (Goldman, 1992) advanced a theory that aimed to explain people’s judgments about justifiedness. Two distinct stages were distinguished. The first stage consisted in the construction of mental lists of “good” and “bad” ways to form beliefs. Processes are put on the “virtues” list if they are thought to be reliable, on the “vices” list if they are thought to be unreliable. At the second stage, attributions of justifiedness and unjustifiedness are made by appeal to these lists. The attributor tries to match the belief-forming process in a target case to some category on one of the lists. If the process matches a “virtue”, the target belief is classed as justified; if it matches a “vice”, the belief is classed as unjustified. How does this two-stage theory account for the clairvoyance cases? Even though clairvoyance may not be on many people’s lists of virtues or vices, there is a class of similar putative faculties that may be on the “vices” list: mental telepathy, ESP, telekinesis, and so forth. Plausibly, evaluators judge clairvoyance to be similar to such vices and therefore classify BonJour’s clairvoyants as having unjustified beliefs.

Let us turn from the first two problems facing reliabilism to the last three. The first in this trio is the generality problem. Attempts to solve the generality problem have been made by Alston (1995), Beebe (2004), Comesana (2006), Heller (1995), and Wunderlich (2003), among others. The two most promising tacks may be those of Wunderlich and Comesana. Wunderlich (2003) takes the novel step of denying that reliabilism must

provide a method for identifying a uniquely relevant type for each process token. Instead of a single reliability number for each token, Wunderlich would have us assign to each token a reliability *vector* and assess a belief’s justificatory status as a function of that vector. The vector would reflect the reliability numbers of all of the types that the token instantiates. Comesana’s (2006) main proposal is to concede that epistemologists haven’t managed to solve the generality problem but to insist that the problem is not peculiar to reliabilism; it is everybody’s problem. Everybody should concede that beliefs are justified in virtue (at least in part) of their provenance. This opens the door to worries about which ways of coming to believe make positive or negative contributions to justification, and which ones are critical. This is likely to produce analogues of the generality problem for other approaches.

A similar response may be in order for the problem of bootstrapping, or easy knowledge. Although Cohen (2002) regards easy knowledge as problematic for reliabilism, he doesn’t consider the problem unique to reliabilism. All views with “basic knowledge structure”, he contends, face the same difficulty. Moreover, James Van Cleve (2003) argues persuasively that we should not make a blanket prohibition against bootstrapping or easy knowledge on pain of inviting skepticism in through the back door. Externalist theories like reliabilism are the only alternatives to skepticism, and the latter is the worse option.

Turning to the final problem, the extra-value-of-knowledge problem, some responses are made by Goldman and Olsson (2008). They proffer two different solutions. One says that a reliably produced true belief possesses a valuable property not shared by an unreliably produced true belief, namely, making it likely that one’s future beliefs of a similar kind will also be true. The second solution argues that when tokens of a type *T* regularly cause tokens of another type, *T'*, then type *T* tends to inherit value from type *T'*. The inherited value accruing to type *T* is also assigned to each token of *T*, whether or not such a token causes one of *T'*. Moreover, things that initially have merely instrumental

value can acquire “autonomous” value status. This autonomous value can be added to the value of an outcome state, which may transpire when the value of a reliable process is added to the value of a true belief it causes.

## 5. OFFSPRING AND COUSINS OF RELIABILISM

If process reliabilism is correct for either knowledge or justification, this naturally invites a further agenda for epistemology, one that several authors have pursued. If reliable processes or methods are critical for knowledge or justifiedness, it makes sense for epistemology to address the question: “Which belief-forming processes or methods are reliable?” Since belief-forming processes are processes of the mind/brain, it is also natural to conclude that this type of question can best be addressed – at least in part – with the help of psychology. If so, it is not unreasonable to think of epistemology as an interdisciplinary subject, one that includes empirical inquiries such as psychology.

Precisely this program was undertaken in *Epistemology and Cognition* (Goldman, 1986). The first part of the book defended process reliabilism, and the second examined an array of cognitive processes in terms of their truth-conducive properties. A book with roughly similar motivations is Michael Bishop and J. D. Trout’s *Epistemology and the Psychology of Human Judgment* (2005). Adopting a variant of reliabilism called “strategic reliabilism”, they hold that “[e]pistemic excellence involves the efficient allocation of cognitive resources to robustly reliable reasoning strategies applied to significant problems” (2005: 71). Psychology is identified as the discipline that has made the most progress in identifying reliable strategies. Yet another book with broadly similar motivations is Gilbert Harman and Sanjeev Kulkarni’s *Reliable Reasoning: Induction and Statistical Learning Theory* (2007). Working in a broadly reliabilist framework, they argue that a branch of computer science – machine learning theory – can advance our understanding of reliable inductive methods.

The foregoing spinoffs of reliabilist epistemology comport with the popular association of reliabilism with epistemic naturalism. How exactly are these two programs related? This partly depends on how one defines “epistemic naturalism”. Goldman (1994) distinguishes several conceptions of epistemic naturalism, two of which are *substantive epistemic naturalism* and *methodological naturalism*. Substantive epistemic naturalism is epitomized by the treatment of cognitive agents as actors in a causal network and epistemic activities as products of the natural order. Thus, psychologically and biologically oriented concepts of the epistemic are considered naturalistic. A clear instance of this kind of naturalism is found in William Lycan (1988), who proposes to “reduce the evaluative notions of epistemology to the teleological notions of the theory of organ systems”. Similarly, Hilary Kornblith’s (2002) argues that knowledge is a natural kind, Paul Churchland (1979) conceives of the epistemic subject as an “epistemic engine”, John Pollock (1986) conceives of the epistemic agent as a “cognitive machine”, and Fred Dretske (1981) develops an information-theoretic approach that resonates with an engineering-based image of the knower. However, this type of epistemic naturalism has no intimate link to reliabilism. Neither Lycan, Churchland, nor Pollock, for example, endorses reliabilism.

Under the methodological conception of epistemic naturalism, a naturalist advocates an empirical, or partly empirical, methodology for epistemology. This methodological conception is present in Goldman’s *Epistemology and Cognition*, Pollock’s *How to Build a Person* (1989), Kitcher’s *The Advancement of Science* (1993), Kornblith’s *Knowledge and Its Place in Nature* (2002), and Bishop and Trout’s *Epistemology and the Psychology of Human Judgment*. However, reliabilists *qua* reliabilists aren’t required to incorporate empirical elements into their methodology. They might deny, for example, that the task of identifying reliable belief-forming processes belongs to epistemology proper. Only the task of determining the criterion or analysis of justifiedness belongs to epistemology proper, they

might hold, and that task isn't empirical. Or they might endorse a form of reliabilism that requires justification-conferring methods to be *necessarily* reliable, and such methods would not be empirically determinable. Bealer's (1998) modal reliabilism would be an example of this.

Apart from epistemic naturalism, reliabilism is allied to other epistemological movements, such as virtue reliabilism. Virtue reliabilism is championed by Ernest Sosa (1991, 2007) and John Greco (2000) (see VIRTUE EPISTEMOLOGY in Part I). Apart from terminological differences, the main distinctive features of Sosa's virtue reliabilism are (A) a distinction between two types of knowledge, and (B) an added causal element in the account of knowledge. Sosa compares a belief to an archer's shot. Like such a shot, it can be accurate, it can manifest epistemic virtue or competence (roughly reliability), and it can be accurate *because* of its competence. He calls these properties, respectively, accuracy, adroitness, and aptness. All three conditions are required for knowledge; that is, knowledge must be true belief, reliably produced, and true *because* reliably produced. The notion of "reflective" knowledge, which he distinguishes from "animal" knowledge, introduces the element of coherence on top of reliability. Greco's agent reliabilism is motivated by two alleged problems for simple reliabilism, problems that lead him to require a reliable process to be part of a stable disposition or faculty that belongs to the agent's cognitive character. On his approach, the creditworthiness of an agent in forming a belief is an important test of justifiedness and knowledge.

Another epistemological current to which reliabilism is allied is social epistemology, at least the "veritistic" form of social epistemology advocated in Goldman's *Knowledge in a Social World* (1999). Veritism is a generalization of reliabilism that assigns degrees of truth-possession to fine-grained doxastic attitudes – e.g. degrees of belief or subjective probabilities – in addition to belief, suspension, and rejection. Veritistic social epistemology would evaluate all sorts of social practices in terms of their veritistic consequences, just as individual epistemology from a reliabilist

perspective would evaluate psychological belief-forming practices in terms of their success in obtaining truth and avoiding error. For more details, see SOCIAL EPISTEMOLOGY in Part I.

See also CAUSAL THEORIES IN EPISTEMOLOGY; EXTERNALISM/INTERNALISM; NATURALIZED EPISTEMOLOGY; RECENT WORK ON THE INTERNALISM–EXTERNALISM CONTROVERSY in Part I.

## BIBLIOGRAPHY

- Alston, W. P.: "An Internalist Externalism," *Synthese* 74 (1988), 265–83. Reprinted in Alston, *Epistemic Justification* (Ithaca, NY: Cornell University Press (1989).
- Alston, W. P.: "How to Think about Reliability," *Philosophical Topics* 23 (1995), 1–29.
- Armstrong, D. M.: *Belief, Truth and Knowledge*. (Cambridge: Cambridge University Press, 1973).
- Bealer, G.: "Intuition and the Autonomy of Philosophy," in *Rethinking Intuition*, ed. M. DePaul and W. Ramsey (Lanham, MD: Rowman and Littlefield, 1998), 201–39.
- Beebe, J.: "The Generality Problem, Statistical Relevance and the Tri-level Hypothesis," *Nous* 38 (2004), 177–95.
- Bishop, M. and Trout, J. D.: *Epistemology and the Psychology of Human Judgment* (Oxford: Oxford University Press, 2005).
- BonJour, L.: "Externalist Theories of Empirical Knowledge," *Midwest Studies in Philosophy* 5 (1980), 53–73.
- BonJour, L.: *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1985).
- Carnap, R.: *Logical Foundations of Probability* (Chicago, IL: University of Chicago Press, 1950).
- Chisholm, R.: *Theory of Knowledge* (Englewood-Cliffs, NJ: Prentice-Hall, 1966).
- Chisholm, R.: *Theory of Knowledge*, 2nd edn (Englewood-Cliffs, NJ: Prentice-Hall, 1977).
- Churchland, P.: *Scientific Realism and the Plasticity of Mind* (Cambridge: Cambridge University Press, 1979).
- Cohen, S.: "Justification and Truth," *Philosophical Studies* 46 (1984), 279–95.
- Cohen, S.: "Basic Knowledge and the Problem of Easy Knowledge," *Philosophy and Phenomenological Research* 65 (2002), 309–29.
- Comesana, J.: "The Diagonal and the Demon," *Philosophical Studies* 110 (2002), 249–66.

- Comesana, J.: "A Well-founded Solution to the Generality Problem," *Philosophical Studies* 129 (2006), 27–47.
- Conee, E. and Feldman, R.: "The Generality Problem for Reliabilism," *Philosophical Studies* 89 (1998), 1–29.
- Dretske, F.: "Conclusive Reasons," *Australasian Journal of Philosophy* 49 (1971), 1–22.
- Dretske, F.: *Knowledge and the Flow of Information* (Cambridge, MA: MIT Press, 1981).
- Feldman, R.: *Epistemology* (Upper Saddle River, NJ: Prentice-Hall, 2003).
- Goldman, A. I.: "A Causal Theory of Knowing," *Journal of Philosophy* 64 (1967), 357–72.
- Goldman, A. I.: "Discrimination and Perceptual Knowledge," *Journal of Philosophy* 73 (1976), 771–91.
- Goldman, A. I.: "What Is Justified Belief?," in *Justification and Knowledge*, ed. G. Pappas (Dordrecht: Reidel, 1979), 1–23. Reprinted in Goldman, *Liaisons: Philosophy Meets the Cognitive and Social Sciences* (Cambridge, MA: MIT Press, 1992).
- Goldman, A. I.: *Epistemology and Cognition* (Cambridge, MA: Harvard University Press, 1986).
- Goldman, A. I.: "Strong and Weak Justification," in *Philosophical Perspectives*, vol. 13, ed. J. Tomberlin (Atascadero, CA: Ridgeview, 1988). Reprinted in Goldman, *Liaisons: Philosophy Meets the Cognitive and Social Sciences* (Cambridge, MA: MIT Press, 1992).
- Goldman, A. I.: "Epistemic Folkways and Scientific Epistemology," in Goldman, *Liaisons: Philosophy Meets the Cognitive and Social Sciences* (Cambridge, MA: MIT Press, 1992).
- Goldman, A. I.: "Naturalistic Epistemology and Reliabilism," *Midwest Studies in Philosophy* 19 (1994), 301–20.
- Goldman, A. I.: *Knowledge in a Social World* (Oxford: Oxford University Press, 1999).
- Goldman, A. I.: "Immediate Justification and Process Reliabilism," in *Epistemology: New Essays*, ed. Q. Smith (Oxford: Oxford University Press, 2008).
- Goldman, A. I. and Olsson, E. J.: "Reliabilism and the Value of Knowledge," in *Epistemic Value*, ed. D. Pritchard, A. Millar and A. Haddock (Oxford: Oxford University Press, 2008).
- Greco, J.: *Putting Skeptics in Their Place* (Cambridge: Cambridge University Press, 2000).
- Harman, G. and Kulkarni, S.: *Reliable Reasoning: Induction and Statistical Learning Theory* (Cambridge, MA: MIT Press, 2007).
- Heller, M.: "The Simple Solution to the Generality Problem," *Nous* 29 (1995), 501–15.
- Hempel, C. G.: "Studies in the Logic of Confirmation," *Mind* 54 (1945), 1–26, 97–121.
- Jeffrey, R.: *The Logic of Decision*, 2nd edn (Chicago, IL: University of Chicago Press, 1983).
- Jones, W. E.: "Why Do We Value Knowledge?," *American Philosophical Quarterly* 34 (1997), 423–39.
- Kornblith, H.: "Beyond Foundationalism and the Coherence Theory," *Journal of Philosophy* 77 (1980), 597–612.
- Kornblith, H.: *Knowledge and Its Place in Nature* (Oxford: Oxford University Press, 2002).
- Kvanvig, J. L.: *The Value of Knowledge and the Pursuit of Understanding* (Cambridge: Cambridge University Press, 2003).
- Lehrer, K.: *Knowledge* (Oxford: Oxford University Press, 1974).
- Lehrer, K.: *Theory of Knowledge* (Boulder, Colo.: Westview Press, 1990).
- Lycan, W.: *Judgment and Justification* (Cambridge: Cambridge University Press, 1988).
- Nozick, R.: *Philosophical Explanation* (Cambridge, MA: Harvard University Press, 1981).
- Pollock, J.: *Contemporary Theories of Knowledge* (Lanham, MD: Rowman and Littlefield, 1986).
- Pollock, J.: *How to Build a Person* (Cambridge, MA: MIT Press, 1989).
- Ramsey, F. P.: "Knowledge," in his *Foundations of Mathematics and Other Essays*, ed. R. B. Braithwaite (New York: Harcourt Brace, 1931).
- Reichenbach, H.: *Experience and Prediction* (Chicago, IL: University of Chicago Press, 1938).
- Riggs, W. D.: "Reliability and the Value of Knowledge," *Philosophy and Phenomenological Research* 64 (2002), 79–96.
- Sosa, E.: *Knowledge in Perspective* (Cambridge: Cambridge University Press, 1991).
- Sosa, E.: "Postscript to 'Proper Functionalism and Virtue Epistemology,'" in *Warrant in Contemporary Epistemology*, ed. J. Kvanvig (Lanham, MD: Rowman and Littlefield, 1996).
- Sosa, E.: *A Virtue Epistemology* (Oxford: Oxford University Press, 2007).
- Swain, M.: *Reasons and Knowledge* (Ithaca, NY: Cornell University Press, 1981).
- Van Cleve, J.: "Is Knowledge Easy – or Impossible? Externalism as the Only Alternative to Skepticism," in *The Skeptics*, ed. S. Luper (Aldershot: Ashgate, 2003).
- Vogel, J.: "Reliabilism Leveled," *Journal of Philosophy* 97 (2000), 602–23.
- Williamson, T.: *Knowledge and Its Limits* (Oxford: Oxford University Press, 2000).
- Wunderlich, M.: "Vector Reliability: A New Approach to Epistemic Justification," *Synthese* 136 (2003), 237–62.



Zagzebski, L.: "The Search for the Source of Epistemic Good," *Metaphilosophy* 34 (2003), 12–28.

ALVIN GOLDMAN

**religious belief, epistemology of** At least two large sets of questions are properly treated under this heading. First, there is a set of broadly theological questions about the relationship between faith and reason, between what one knows by way of reason, broadly construed, and what one knows by way of faith. I call these questions *theological* because, of course, one will find them of interest only if one thinks that in fact there is such a thing as faith, and that we do know something by way of it. Secondly, there is a whole set of questions having to do with whether and to what degree religious beliefs have *warrant*, or *justification*, or *positive epistemic status*. I shall concentrate upon the second of these two sets of questions.

Epistemology, so we are told, is *theory of knowledge*; its aim is to discern and explain that quality or quantity enough of which distinguishes knowledge from mere true belief. We need a name for this quality or quantity, whatever precisely it is: call it "warrant". From this point of view, the epistemology of religious belief should centre on the question whether religious belief has warrant, and if it does, how much it has and how it gets it. As a matter of fact, however, epistemological discussion of religious belief, at least since the Enlightenment (and in the Western world, especially the English-speaking Western world) has tended to focus, not on the question whether religious belief has warrant, but whether it is *justified*. More precisely, it has tended to focus on the question whether those properties are enjoyed by *theistic belief* – the belief that there exists a person like the God of traditional Christianity, Judaism, and Islam: an almighty, all-knowing, wholly benevolent and loving spiritual person who has created the world. The chief question, therefore, has been whether theistic belief is justified; the same question is often put by asking whether theistic belief is *rational* or rationally acceptable. Still further, the typical

way of addressing this question has been by way of discussing *arguments* for and against the existence of God. On the pro side, there are the traditional theistic proofs or arguments: the ontological, cosmological and teleological arguments, to use Kant's terms for them. On the other side, the anti-theistic side, the principal argument is the argument from evil, the argument that it is not possible or at least not probable that there be such a person as God, given all the pain, suffering and evil the world displays. This argument is flanked by subsidiary arguments, such as the claim that the very concept of God is incoherent (because, for example, it is impossible that there be a person without a body), and Freudian and Marxist claims that religious belief arises out of a sort of magnification and projection into the heavens of human attributes we think important.

But why has discussion centred on *justification* rather than warrant? And precisely what is justification? And why has the discussion of the justification of theistic belief focused so heavily on arguments for and against the existence of God?

As to the first question, we can see why once we see that the dominant epistemological tradition in modern Western philosophy has tended to *identify* warrant with justification. On this way of looking at the matter, warrant, that which distinguishes knowledge from mere true belief, just *is* justification. Indeed, until recently the "Justified True Belief" theory of knowledge – the theory according to which knowledge is justified true belief – has enjoyed the status of orthodoxy. According to this view, knowledge is justified true belief; therefore any of your beliefs has warrant for you if and only if you are justified in holding it.

But what *is* justification? What is it to be justified in holding a belief? To get a proper sense of the answer, we must turn to those twin towers of Western epistemology, René Descartes (*see* DESCARTES) and (especially) John Locke (*see* LOCKE). The first thing to see is that according to Descartes and Locke, there are *epistemic* or intellectual *duties*, or obligations, or requirements. Thus Locke:



Faith is nothing but a firm assent of the mind: which if it be regulated, as is our duty, cannot be afforded to anything, but upon good reason; and so cannot be opposite to it. He that believes, without having any reason for believing, may be in love with his own fancies; but neither seeks truth as he ought, nor pays the obedience due his maker, who would have him use those discerning faculties he has given him, to keep him out of mistake and error. He that does not this to the best of his power, however he sometimes lights on truth, is in the right but by chance; and I know not whether the luckiness of the accident will excuse the irregularity of his proceeding. This at least is certain, that he must be accountable for whatever mistakes he runs into; whereas he that makes use of the light and faculties God has given him, and seeks sincerely to discover truth, by those helps and abilities he has, may have this satisfaction in doing his duty as a rational creature, that though he should miss truth, he will not miss the reward of it. For he governs his assent right, and places it as he should, who in any case or matter whatsoever, believes or disbelieves, according as reason directs him. He that does otherwise, transgresses against his own light, and misuses those faculties, which were given him. . . .

(*Essay* 4.17.24)

Rational creatures, creatures with reason, creatures capable of *believing* propositions (and of disbelieving and being agnostic with respect to them), says Locke, have duties and obligations with respect to the regulation of their belief or assent. Now the central core of the notion of justification (as the etymology of the term indicates) is this: one is justified in doing something or in behaving a certain way, if in so doing one is innocent of wrongdoing and hence not properly subject to blame or censure. You are justified, therefore, if you have violated no duties or obligations, if you have conformed to the relevant requirements, if you are within your rights. To be justified in *believing* something, then, is to be within your rights in so believing, to be doing no wrong in believing in this way, to be flouting no duty, to be satisfying your epistemic duties and obligations. This way of thinking of justification has been the dominant way of thinking about justification; and this way of thinking has many important contemporary

representatives. Roderick Chisholm (see CHISHOLM), for example (as distinguished an epistemologist as the twentieth century can boast), in his earlier work explicitly explains justification in terms of epistemic duty (Chisholm, 1977, p. 14; 1982, p. 7).

The (or a) main epistemological question about religious belief, therefore, has been the question whether or not religious belief in general and theistic belief in particular is justified. And the traditional way to answer that question has been to inquire into the *arguments* for and against theism. Why this emphasis upon these arguments? An argument is a way of marshalling your *propositional evidence* – the evidence from other propositions you believe – for or against a given proposition. And the reason for the emphasis upon argument is the assumption that theistic belief is justified if and only if there is sufficient propositional evidence for it. If there isn't much by way of propositional evidence for theism, then you are not justified in accepting it. More exactly, if you accept theistic belief without having propositional evidence for it, then you are going contrary to epistemic duty and are therefore unjustified in accepting it. Thus W. K. Clifford (that “delicious *enfant terrible*”, as William James calls him) trumpets that “it is wrong, always, everywhere, and for anyone to believe anything upon insufficient evidence”; his is only the most strident in a vast chorus of voices insisting that there is an intellectual duty not to believe in God unless you have propositional evidence for that belief. (A few others in the choir: Sigmund Freud, Brand Blanshard, H. H. Price, Bertrand Russell and Michael Scriven.)

Now how is it that the *justification* of theistic belief gets identified with there being *propositional evidence* for it? Justification is a matter of being blameless, of having done one's duty (in this context, one's epistemic duty): what, precisely, has this to do with having propositional evidence?

The answer, again, is to be found in Descartes and especially Locke. Justification is the property your beliefs have when, in forming and holding them, you conform to your epistemic duties and obligations. But according to Locke, a central epistemic duty

is this: *to believe a proposition only to the degree that it is probable with respect to what is certain for you*. What propositions are certain for you? First (according to Descartes and Locke), propositions about your own immediate experience: that you have a mild headache, or that it seems to you that you see something red; and second, propositions that are self-evident for you: necessarily true propositions so obvious that you can't so much as entertain them without seeing that they must be true. (Examples would be simple arithmetical and logical propositions, together with such propositions as that the whole is at least as large as the part, that red is a colour, and that whatever exists has properties.) Propositions of these two sorts are certain for you; as for other propositions, you are justified in believing one of them only to the degree to which it is probable with respect to what is certain for you. According to Locke, therefore, and according to the whole modern foundationalist tradition (see FOUNDATIONALISM) initiated by Locke and Descartes (a tradition that until recently has dominated Western thinking about these topics) there is a duty not to accept a proposition unless it is certain or probable with respect to what is certain.

In the present context, therefore, the central Lockean assumption is that there is an epistemic duty not to accept theistic belief unless it is probable with respect to what is certain for you; as a consequence, theistic belief is justified only if the existence of God is probable with respect to what is certain. Locke doesn't *argue* for this proposition; he simply *announces* it; and epistemological discussion of theistic belief has for the most part followed him in making this assumption. This enables us to see why epistemological discussion of theistic belief has tended to focus on the arguments for and against theism: on the view in question, theistic belief is justified only if it is probable with respect to what is certain; and the way to show that it is probable with respect to what is certain is to give arguments for it from premisses that are certain (or are sufficiently probable with respect to what is certain).

There are at least three important problems with this approach to the epistemo-

logy of theistic belief. First, the standards for theistic arguments have traditionally been set absurdly high (and perhaps part of the responsibility for this must be laid at the door of some who have offered these arguments and claimed that they constitute wholly demonstrative proofs). The idea seems to be that a good theistic argument must start from what is self-evident and proceed majestically by way of self-evidently valid argument forms to its conclusion. It is no wonder that few if any theistic arguments meet *that* lofty standard – particularly in view of the fact that almost no philosophical arguments of *any* sort meet it. (Think of your favourite philosophical argument; does it really start from premisses that are self-evident and move by way of self-evident argument forms to its conclusion?)

Secondly, attention has been mostly confined to three theistic arguments: the traditional ontological, cosmological and teleological arguments, to use Kant's classification. But in fact there are many more good arguments: arguments from the nature of proper function, and from the nature of propositions, numbers and sets. There are arguments from intentionality, from counter-factuals, from the confluence of epistemic reliability with epistemic justification, from reference, simplicity, intuition and love. There are arguments from colours and flavours, from miracles, play and enjoyment, morality, from beauty, and from the meaning of life. There is even a theistic argument from the existence of evil.

But there is a third and deeper problem here. The basic assumption is that theistic belief is justified only if it is or can be shown to be probable with respect to some body of evidence or propositions – perhaps those that are self-evident or about one's own mental life, as Locke thought. But is this assumption true? The idea is that theistic belief is very much like a *scientific hypothesis*: it is acceptable if and only if there is an appropriate balance of propositional evidence in favour of it. But why believe a thing like that? Perhaps the theory of relativity or the theory of evolution is like that: such a theory has been devised to explain the phenomena and gets all its

warrant from its success in so doing. But other beliefs – e.g. memory beliefs (*see* MEMORY), belief in other minds (*see* OTHER MINDS) – are not like that; they aren't hypotheses at all, and are not accepted because of their explanatory powers. They are instead the propositions from which one starts in attempting to give evidence for a hypothesis. Now why assume that theistic belief, belief in God, is in this regard more like a scientific hypothesis than like, say, a memory belief? Why think that the justification of theistic belief depends upon the evidential relation of theistic belief to other things one believes? According to Locke and the beginning of this tradition, it is because there is a duty not to assent to a proposition unless it is probable with respect to what is certain to you; but is there really any such duty? No one has succeeded in showing that, say, belief in other minds or the belief that there has been a past, is probable with respect to what is certain for us. Suppose it isn't: does it follow that you are living in epistemic sin if you believe that there are other minds? Or a past?

There are urgent questions about any view according to which one has duties of the sort *don't believe p unless it is probable with respect to what is certain for you*. First, if this is a duty, is it one to which I can conform? My beliefs are for the most part not within my control: certainly they are not within my direct control. I believe that there has been a past and that there are other people; even if these beliefs are not probable with respect to what is certain for me (and even if I came to know this) I couldn't give them up. Whether or not I accept such beliefs isn't really up to me at all: I can no more refrain from believing these things than I can refrain from conforming to the law of gravity. Second, is there really any reason for thinking I *have* such a duty? Nearly everyone recognizes such duties as that of not engaging in gratuitous cruelty, taking care of one's children and one's aged parents, and the like; but do we also find ourselves recognizing that there is a duty not to believe what isn't probable (or what we can't see to be probable) with respect to what is certain for us? It hardly seems so. But if so, it is hard to see why being justified in

believing in God requires that the existence of God be probable with respect to some such body of evidence as the set of propositions certain for you. Perhaps theistic belief is *properly basic*, i.e. such that one is perfectly justified in accepting it without accepting it on the evidential basis of other propositions one believes.

Taking *justification* in that original etymological fashion, therefore, there is every reason to doubt that one is justified in holding theistic belief only if one has evidence for it. Of course, the term "justification" has undergone various analogical extensions in the work of various philosophers; it has been used to name various properties that are different from justification etymologically so-called, but analogically related to it. Thus the term is sometimes used just to mean propositional evidence (*see* EVIDENCE): to say that a belief is justified for someone is to say that he has propositional evidence (or sufficient propositional evidence) for it. So taken, however, the question whether theistic belief is justified loses some of its interest; for it isn't clear (given this use) that there is anything amiss with holding beliefs that are unjustified in *that* sense. Perhaps one also doesn't have propositional evidence for one's memory beliefs; if so, that would not be a mark against them and would not suggest that there is something wrong with holding them.

Another analogically connected way to think about justification (a way endorsed by the later Chisholm) is to think of it as simply a *relation of fittingness* between a given proposition and one's epistemic base – which includes the other things one believes, as well as one's experience. Perhaps that is the way justification is to be thought of; but then it is no longer at all obvious that theistic belief has this property of justification only if it is probable with respect to some body of evidence. Perhaps, again, it is like memory beliefs in this regard.

To recapitulate: the dominant Western tradition has been inclined to identify warrant with justification; it has been inclined to take the latter in terms of duty and the fulfillment of obligation, and hence to suppose that there is an epistemic duty not to believe in God

unless you have good propositional evidence for the existence of God. Epistemological discussion of theistic belief, as a consequence, has concentrated upon the propositional evidence for and against theistic belief – i.e. on arguments for and against theistic belief. But there is excellent reason to doubt that there are epistemic duties of the sort the tradition appeals to here.

And perhaps it was a mistake to identify warrant with justification in the first place. The beliefs of a madman who thinks he is Napoleon have little warrant for him; his problem, however, need not be dereliction of epistemic duty. He is in difficulty, but it is not or not necessarily that of failing to fulfill epistemic duty. He may be doing his epistemic best; he may be doing his epistemic duty *in excelsis*; but his madness prevents his beliefs from having much by way of warrant. His lack of warrant is not a matter of being unjustified, i.e., failing to fulfill epistemic duty. So warrant and *being epistemically justified* are by no means the same thing. Another example: suppose (to use the favourite twentieth-century variant of Descartes' evil demon example) I have been captured by Alpha-Centaurian super-scientists; running a cognitive experiment, they remove my brain, keep it alive in a vat of artificial nutrients, and by virtue of their advanced technology induce in me the beliefs I might otherwise have if I were going about my usual business. Then my beliefs would not have much by way of warrant; but would it be because I was failing to do my epistemic duty? Hardly.

As a result of these and other problems, another, *externalist* way of thinking about knowledge has appeared in recent epistemology (see EXTERNALISM/INTERNALISM). Or perhaps the thing to say is that it has reappeared; for the dominant strains in epistemology prior to the Enlightenment were really externalist. According to this externalist way of thinking, warrant does not depend upon satisfaction of duty, or upon anything else to which the knower has special cognitive access (as he does to what is about his own experience and to whether he is trying his best to do his epistemic duty); it depends instead upon factors "external" to the epistemic agent –

such factors as whether his beliefs are produced by a reliable cognitive mechanism, or whether they are produced by epistemic faculties functioning properly in an appropriate epistemic environment.

How shall we think about the epistemology of theistic belief in this more externalist way (which is at once both satisfyingly traditional and agreeably up to date)? The chief thing to see here, I think, is that the *ontological* question whether there is such a person as God is in a way prior to the *epistemological* question about the warrant of theistic belief. It is natural to think that if in fact we *have* been created by God, then the cognitive processes that issue in belief in God are indeed reliable belief-producing processes; if in fact God has created us, then no doubt the cognitive faculties that produce belief in God are functioning properly in an epistemically congenial environment. On the other hand, if there isn't any such person as God, if theistic belief is an illusion of some sort, then things are much less clear. Then belief in God in the basic way will no doubt be produced by wishful thinking or some other cognitive process not aimed at truth; thus it will have little or no warrant. And belief in God on the basis of argument would be like belief in false philosophical theories on the basis of argument: do such beliefs have warrant? More exactly, would they have warrant if the externalist views of warrant I mentioned above are true? It isn't easy to say.

In any event, the usual custom of discussing the epistemological questions about theistic belief as if they could be profitably discussed independently of the ontological issue as to whether or not theism is true, is misguided. These two issues are intimately intertwined.

*See also* BELIEF IN AND BELIEF THAT; EPISTEMIC VIRTUE; ETHICS AND EPISTEMOLOGY; VIRTUE EPISTEMOLOGY in Part I.

#### BIBLIOGRAPHY

- Alston, William P.: *Perceiving God* (Ithaca, NY: Cornell University Press, 1992).  
 Blanshard, Brand: *Reason and Belief* (London: Allen and Unwin, 1974).



- Chisholm, Roderick: *Theory of Knowledge* (New York: Prentice-Hall, 1966; 2nd edn 1977).
- Chisholm, Roderick: *The Foundations of Knowing* (Minneapolis, MN: University of Minnesota Press, 1982).
- Clifford, William K.: "The Ethics of Belief," in *Lectures and Essays* (London: Macmillan, 1879).
- Freud, Sigmund: *The Future of an Illusion* (New York, Norton, 1961; first German edn 1927).
- Mackie, John: *The Miracle of Theism* (Oxford: Oxford University Press, 1982).
- Mavrodes, George: *Belief in God; A Study in the Epistemology of Religion* (New York: Random House, 1970).
- Plantinga, A. and Wolterstorff, N., eds: *Faith and Rationality* (Notre Dame, IN: University of Notre Dame Press, 1983).
- Price, Henry H.: *Belief* (London: Allen and Unwin, 1969).
- Russell, Bertrand: *A History of Western Philosophy* (New York: Simon and Schuster, 1945).
- Scriven, Michael: *Primary Philosophy* (New York: McGraw-Hill, 1966).

ALVIN PLANTINGA

### religious belief, epistemology of – recent developments

There are at least three recent developments worth highlighting in the epistemology of religious belief. The first is the detailed presentation and defense of Reformed Epistemology – the view described above according to which theistic belief, along with other religious beliefs, can be justified or warranted for humans *non-inferentially*, i.e., not on the basis of inference or argument. (The view is called "Reformed Epistemology" because of its association with John Calvin, though one finds hints of this view in Aquinas, Augustine, and St Paul as well.) The chief defender of Reformed Epistemology is Alvin Plantinga. His *Warrant* trilogy began with the 1993 publication of two volumes discussing the epistemology of belief in general (including primarily non-religious beliefs) and culminated in 2000 with *Warranted Christian Belief*. In that book, Plantinga argues against a prominent objection to religious belief, one which says "whether or not religious belief is true, we can see that it is unreasonable and unwar-

ranted". He first notes that objections of this sort typically presuppose the falsity of a religious position according to which humans have non-inferential ways of knowing or rationally believing in God's existence. Then, after fleshing out an account of how religious belief can be non-inferentially warranted (via an innate sense of divinity or the testimony of the Holy Spirit), he argues that this account is likely to be true if Christianity is true, in which case objections presupposing that account's falsity are effective only if accompanied by arguments for the falsity of Christianity. One common complaint against Plantinga's *Warranted Christian Belief* is that he argues only that Christian belief is warranted *if it's true*, but he doesn't argue that it's true. However, the fact that Plantinga doesn't argue for the truth of Christian belief seems perfectly compatible with his position that Christian belief, like theistic belief, can be warranted independently of any arguments for it.

The second recent development worth noting – one which is in opposition to the first – is the continued clarification and defense of Theistic Evidentialism, the view that theistic or religious belief is often rational, warranted, and justified for people in contemporary Western societies but that it can be so *only* on the basis of good arguments. (Non-theistic Evidentialism also maintains this "good argument" requirement for the rationality of theism but it goes on to deny that there are any such arguments and concludes, accordingly, that theistic belief isn't rational.) In connection with the Theistic Evidentialist's reaction to Reformed Epistemology, there have been ongoing attempts to develop strong arguments for theism and for more specific religious doctrines as well. Pre-eminent among these are the works of the now retired Oxford philosopher Richard Swinburne, including most notably his trilogy (1977, 1979, 1981) arguing for theism and several other books (e.g. 1991, 2003) in support of more specific religious doctrines.

The third development is an increased appreciation and understanding of the similarities and differences between Reformed Epistemology and Theistic Evidentialism. For



example, it is now widely recognized that to insist that theistic beliefs are justified or warranted only on the basis of evidence (*see* EVIDENCE) is consistent with both positions (though Reformed Epistemologists aren't committed, as such, to insisting on this). The way the views differ on the topic of evidence is that Theistic Evidentialists say that the required evidence must be propositional, whereas Reformed Epistemologists think *non-propositional* (experiential) evidence can be sufficient for warranted theistic belief. Similarly, the view that there are good arguments for theism and other religious doctrines – arguments strong enough to justify firmly held beliefs in them – is compatible with both Reformed Epistemology and Theistic Evidentialism. The difference here is that Theistic Evidentialists think that, without such arguments, their theism is irrational whereas Reformed Epistemologists insist that, even if there are such arguments, they aren't *required* for rational or warranted theistic belief. It's true that most Reformed Epistemologists – while perhaps willing to admit that theistic arguments are often as impressive as arguments for their favorite philosophical views on non-religious topics – are inclined to think the available arguments aren't strong enough to warrant the level of conviction with which religious believers typically hold their religious positions. But this is not essential to the Reformed Epistemologist's position, even if it can play some role in motivating it.

So the difference between Reformed Epistemology and Theistic Evidentialism does not come down to the question of whether warranted theistic belief requires evidence or to the question of whether there are strong theistic arguments. Nor does it come down to the externalism/internalism dispute in epistemology (*see* EXTERNALISM/INTERNALISM). It's true that most Theistic Evidentialists are internalists and most Reformed Epistemologists are externalists. But Theistic Evidentialism is consistent with externalism: an externalist could think that, just as humans lack any cognitive faculty enabling them to see non-inferentially that there are electrons, so

also humans lack any cognitive faculty enabling them to see non-inferentially that God exists. Such an externalist could, therefore, endorse Theistic Evidentialism. Likewise, Reformed Epistemology is consistent with internalism: an internalist could think that we in fact have a sense of divinity enabling us to believe non-inferentially and with justification that God exists; and she could add that just as our perceptual beliefs are justified only if we are *aware* of the experiential grounds which adequately support them, so also our theistic beliefs are justified only if we are *aware* of the experiential grounds which adequately support them. By adding this awareness requirement, this sort of Reformed Epistemologist would qualify as an internalist (at least, the kind that requires awareness *of* adequate grounds, even if not the kind that requires awareness *that* the grounds are adequate).

The main difference, therefore, between Reformed Epistemologists and Theistic Evidentialists comes to this: the former think humans have some way – perhaps via some innate cognitive faculty – of *non-inferentially* knowing or rationally believing in theism with a degree of confidence that matches what we in fact find among religious believers. The latter think humans don't have such a source of non-inferential warrant or justification – at least, not one that provides sufficient warrant or justification to withstand the intellectual challenges to religious belief that are widespread in the contemporary academic climate of the West.

#### BIBLIOGRAPHY

- Plantinga, A.: *Warrant: The Current Debate* (New York: Oxford University Press, 1993).
- Plantinga, A.: *Warrant and Proper Function* (New York: Oxford University Press, 1993).
- Plantinga, A.: *Warranted Christian Belief* (New York: Oxford University Press, 2000).
- Swinburne, R.: *The Coherence of Theism* (Oxford: Clarendon Press, 1977).
- Swinburne, R.: *The Existence of God* (Oxford: Clarendon Press, 1979; 2nd edn 2004).

Swinburne, R.: *Faith and Reason* (Oxford: Clarendon Press, 1981; 2nd edn 2005).

Swinburne, R.: *Revelation* (Oxford: Clarendon Press, 1991).

Swinburne, R.: *The Resurrection of God Incarnate* (Oxford: Clarendon Press, 2003).

MICHAEL BERGMANN

## representation

### TERMINOLOGY

#### *Representation*

Contemporary philosophy of mind, following cognitive science, uses the term “representation” to mean just about anything that can be semantically evaluated. Thus, representations may be said to be true, to refer, to be true-of something, to be about something, to be accurate, etc. Representation thus conceived comes in many varieties. The most familiar are pictures, three-dimensional models (e.g. statues, scale models), linguistic text (including mathematical formulas) and various hybrids of these such as diagrams, maps, graphs and tables. It is an open question in cognitive science whether mental representation, which is our real topic here, falls within any of these familiar sorts.

#### *Content*

“Content” has become a technical term in philosophy for whatever it is a representation has that makes it semantically evaluable. Thus, a statement is sometimes said to have a proposition or truth condition as its content; a term is sometimes said to have a concept as its content. Much less is known about how to characterize the contents of non-linguistic representations than is known about characterizing linguistic representations. “Content” is a useful term precisely because it allows one to abstract away from questions about what semantic properties representations have; a representation’s content is just whatever it is that underwrites its semantic evaluation.

### REPRESENTATION AND THOUGHT

#### *The representational theory of cognition*

It is uncontroversial in contemporary cognitive science that cognitive processes are processes that manipulate representations. This idea seems nearly inevitable. What makes the difference between processes that are cognitive – solving a problem, say – and those that are not – a patellar reflex, for example – is just that cognitive processes are epistemically assessable. A solution procedure can be justified or correct; a reflex cannot. Since only things with content can be epistemically assessed, processes appear to count as cognitive only in so far as they implicate representations.

It is tempting to think that thoughts are the mind’s representations: Aren’t thoughts just those mental states that have (semantic) content? This is, no doubt, harmless enough provided we keep in mind that cognitive science may attribute to thoughts properties and contents that are foreign to common-sense. First, most of the representations hypothesized by cognitive science do not correspond to anything common sense would recognize as thoughts. Standard psycholinguistic theory, for instance, hypothesizes the construction of representations of the syntactic structures of the utterances one hears and understands. Yet we are not aware of, and non-specialists do not even understand, the structures represented. Thus, cognitive science may attribute thoughts where common sense would not. Second, cognitive science may find it useful to individuate thoughts in ways foreign to common sense. (See the discussion of “internalistic” theories of content below.)

#### *The representational theory of intentionality*

The representational theory of cognition gives rise to a natural theory of intentional states such as believing, desiring and intending. According to this theory, intentional states factor into two aspects: a *functional* aspect that distinguishes believing from desiring and so on, and a *content* aspect that

distinguishes beliefs from each other, desires from each other, and so on. A belief that  $p$  might be realized as a representation with the content that  $p$  and the function of serving as a premise in inference. A desire that  $p$  might be realized as a representation with the content that  $p$  and the function of initiating processing designed to bring it about that  $p$  and terminating such processing when a belief that  $p$  is formed.

#### REPRESENTATION AND THE THEORY OF CONTENT

A great deal of philosophical effort has been lavished on the attempt to *naturalize content*, i.e. to explain in non-semantic, non-intentional terms what it is for something to be a representation (have content), and what it is for something to have some particular content rather than some other. There appear to be only four types of theory that have been proposed: theories that ground representation in (1) similarity, (2) covariance, (3) functional role, (4) teleology.

*Similarity theories* hold that  $r$  represents  $x$  in virtue of being similar to  $x$ . This has seemed hopeless to most as a theory of mental representation because it appears to require that things in the brain must share properties with the things they represent: To represent a cat as furry appears to require something furry in the brain. Perhaps a notion of similarity that is naturalistic and does not involve property sharing can be worked out, but it is not obvious how.

*Covariance theories* hold that  $r$ 's representing  $x$  is grounded in the fact that  $r$ 's occurrence covaries with that of  $x$ . This is most compelling when one thinks about detection systems: the firing of a neural structure in the visual system is said to represent vertical orientations if its firing covaries with the occurrence of vertical lines in the visual field. Dretske (1981) and Fodor (1987) have, in different ways, attempted to promote this idea into a general theory of content.

*Functional role theories* hold that  $r$ 's representing  $x$  is grounded in the functional role  $r$  has in the representing system, i.e. on the

relations imposed by specified cognitive processes between  $r$  and other representations in the system's repertoire. Functional role theories take their cue from such common-sense ideas as that people cannot believe that cats are furry if they do not know that cats are animals or that fur is like hair. For a defence of the functional role approach, see Block (1986), Loar (1981), and Harman (1982).

*Teleological theories* hold that  $r$  represents  $x$  if it is  $r$ 's function to indicate (i.e. covary with)  $x$ . Teleological theories differ depending on the theory of functions they import. Perhaps the most important distinction is that between historical theories of functions (Millikan, 1984; Papineau, 1988), and a-historical theories (Stampe, 1977; Fodor, 1987, 1990; Cummins, 1987). Historical theories individuate functional states (hence contents) in a way that is sensitive to the historical development of the state, i.e. to factors such as the way the state was "learned", or the way it evolved. An historical theory might hold that the function of  $r$  is to indicate  $x$  only if the capacity to token  $r$  was developed (selected, learned) because it indicates  $x$ . Thus a state physically indistinguishable from  $r$  (physical states being a-historical) but lacking  $r$ 's historical origins would not represent  $x$  according to historical theories.

Theories of representational content may be classified according to whether they are atomistic or holistic and according to whether they are externalistic or internalistic. (see EXTERNALISM/INTERNALISM; HOLISM). Atomistic theories take a representation's content to be something that can be specified independently of that representation's relations to other representations. What Fodor (1987) calls the crude causal theory, for example, takes a representation to be a  $|cow|$  – a mental representation with the same content as the word "cow" – if its tokens are caused by instantiations of the property of being-a-cow, and this is a condition that places no explicit constraints on how  $|cow|$ 's must or might relate to other representations. Holistic theories contrast with atomistic theories in taking the relations a representation bears to others to be essential to its content. According to functional role theories, a representation is a

|cow| if it behaves like a |cow| should behave in inference.

Internalist theories take the content of a representation to be a matter determined by factors internal to the system that uses it. Thus, what Block (1986) calls “short-armed” functional role theories are internalist. Externalist theories take the content of a representation to be determined, in part at least, by factors external to the system that uses it. Covariance theories, as well as teleological theories that invoke an historical theory of functions, take content to be determined by “external” factors. Crossing the atomist–holist distinction with the internalist–externalist distinction gives us the table below.

Externalist theories (sometimes called non-individualistic theories, following Burge, 1979) have the consequence that molecule for molecule identical cognitive systems might yet harbor representations with different contents (Putnam, 1975; Burge, 1979; Millikan, 1984). This has given rise to a

controversy concerning “narrow” content. If we assume some form of externalist theory is correct, then content is, in the first instance “wide” content, i.e. determined in part by factors external to the representing system. On the other hand, it seems clear that, on plausible assumptions about how to individuate psychological capacities, internally equivalent systems must have the same psychological capacities. Hence, it would appear that wide content cannot be relevant to characterizing psychological equivalence. Since cognitive science generally assumes that content is relevant to characterizing psychological equivalence, philosophers attracted to externalist theories of content have sometimes attempted to introduce “narrow” content, i.e. an aspect or kind of content that is equivalent in internally equivalent systems. The simplest such theory is Fodor’s idea (1987) that narrow content is a function from contexts (i.e. from whatever the external factors are) to wide contents.

	<i>Atomist</i>	<i>Holist</i>
<i>Externalist</i>	Covariance	Historical-teleological
<i>Internalist</i>	Ahistorical-teleological	Similarity
		“Long-armed” functional role
		“Short-armed” functional role

#### MISREPRESENTATION

Perhaps the most serious single problem facing the theory of representation is to give a satisfactory account of misrepresentation. Consider covariance theories. A paradigm case of misrepresentation is a case in which a |cow| is tokened in response to a horse. But if *r* is tokened in response to a horse, then its tokenings don’t covary with the occurrence of cows, hence it cannot be a |cow|. Perhaps it represents something common to cows and horses, e.g. the property of being a cow-or-horse. (This is why Fodor (1987) has dubbed the problem the disjunction problem.) But then it is not a |cow|, and not a misrepresentation.

Functional role theories face an analogous problem. A case of error is a case in which

$\Sigma$  tokens a representation *m* that it uses as a |mouse|, in a situation in which it should have tokened something else, say *s*, something  $\Sigma$  uses as a |shrew|. So *m* was tokened by a process *C* that is designed to work only if the slot that  $\Sigma$  fills with *m* is something that represents shrews. But if *C* uses *m*, then *m* isn’t used (exclusively) as a |mouse| in  $\Sigma$ , contrary to hypothesis. It appears that, at a minimum,  $\Sigma$  uses *m* as a |mouse or shrew|. Hence, *m* represents being-a-mouse-or-shrew, and there was no error after all.

Teleological theories appear to have the right form to deal with misrepresentation, for they distinguish between what *r* indicates (what it covaries with) and what it is *r*’s function to indicate, thus making room for error as a mismatch between the two. A representation *r* misrepresents *x*, according to this

approach, when it is  $r$ 's function to indicate  $x$  but  $r$  does not indicate  $x$ .

A problem with this approach is that many representations do not have indication as their function – the antecedents of hypothetical thoughts, for example. Teleological theories are therefore forced to hold that representations that do not have indication as their function are complex, the idea being that non-indicating representations derive their content from their structure together with their indicating constituents. Teleological theories are therefore committed to a kind of verificationist reductionism (see VERIFICATIONISM): the content of every representation must reduce to the contents of those whose function is to indicate. Aside from general worries about semantic reductionism, this commits the teleological theory to the view that every serious scheme of mental representation must have a componential semantics.

Covariationist theories seem to have the same implication, since many non-erroneous representations do not covary with what they represent (almost every case of non-perceptual thought). While it might be true that every serious scheme of mental representation must have a componential semantics, (for an argument, see Fodor and Pylyshyn, 1988), it appears to be an empirical hypothesis about the mind, and therefore should not be an analytic consequence of the theory of content.

See also CONCEPTS; EXTERNALISM/INTERNALISM.

## BIBLIOGRAPHY

- Block, N.: "Advertisement for a Semantics for Psychology," *Midwest Studies in Philosophy* 10 (1986), 615–78.
- Burge, T.: "Individualism and the Mental," *Philosophical Review* 95 (1979), 3–45.
- Cummins, R.: *Meaning and Mental Representation* (Cambridge, MA: MIT Press/Bradford, 1987).
- Dretske, F.: *Knowledge and the Flow of Information* (Cambridge, MA: MIT Press/Bradford, 1981).
- Dretske, F.: "Misrepresentation," in *Belief*, ed. R. Bogdan (Oxford: Oxford University Press, 1986).

- Fodor, J. and Pylyshyn, Z.: "Connectionism and Cognitive Architecture," *Cognition* 28 (1988), 3–71.
- Fodor, J.: *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (Cambridge, MA: MIT Press/Bradford, 1987).
- Fodor, J.: *A Theory of Content and Other Essays* (Cambridge, MA: MIT Press/Bradford, 1990).
- Harman, G.: "Conceptual Role Semantics," *Notre Dame Journal of Formal Logic* 23 (1982), 242–56.
- Loar, B.: *Mind and Meaning* (Cambridge: Cambridge University Press, 1981).
- Millikan, R.: *Language, Thought and Other Biological Categories* (Cambridge, MA: MIT Press/Bradford, 1984).
- Papineau, D.: *Reality and Representation* (Oxford: Blackwell, 1988).
- Putnam, H.: "The Meaning of Meaning," in *Language, Mind and Knowledge*, ed. K. Gunderson, (Minneapolis, MN: University of Minnesota Press, 1975).
- Stampe, D.: "Towards a Causal Theory of Linguistic Representation," *Midwest Studies in Philosophy* 2 (1977), 42–63.

ROBERT CUMMINS

**representative realism** This theory holds that (1) there is a world whose existence and nature is independent of us and of our perceptual experience of it, (2) perceiving an object located in that external world necessarily involves causally interacting with that object, and (3) the information acquired in perceiving an object is indirect; it is information most immediately about the perceptual experience caused in us by the object, and only derivatively about the object itself. Clause (1) makes representative realism a species of realism, clause (2) makes it a species of causal theory of perception, and clause (3) makes it a species of representative as opposed to direct realism (see REALISM; CAUSAL THEORIES IN EPISTEMOLOGY; DIRECT REALISM).

Traditionally, representative realism has been allied with an act/object analysis (see ACT/OBJECT ANALYSIS) of sensory experience. (See, for instance, Locke, 1690, representative realism's most famous advocate.) Indeed, this act/object analysis is traditionally a major plank in arguments for representative realism.



One use of terms such as “looks”, “seems” and “feels” is to express opinion. “It looks as if the Labour Party will win the next election” expresses an opinion about the party’s chances, and does not describe a particular kind of perceptual experience. We can, however, use such terms to describe perceptual experience divorced from any opinion to which the experience may incline us. A straight stick half in water looks bent, and does so to people completely familiar with this illusion who have, therefore, no inclination to hold that the stick is in fact bent. Such uses of “looks”, “seems”, “tastes”, etc., are commonly called phenomenological.

The act/object theory holds that the sensory experiences recorded by sentences employing these terms in their phenomenological sense are a matter of being directly acquainted with something which actually bears the apparent property. When something looks red to me, I am acquainted with a red expanse (in my visual field); when something looks bent to me, I am acquainted with a bent shape (in my visual field); when something tastes bitter to me, I am directly acquainted with a sensation with the property of being bitter; and so on and so forth. (If you do not understand the term “directly acquainted”, stick a pin into your finger. The relation you will then bear to your pain, as opposed to the relation of concern you might bear to another’s pain when told about it, is an instance of direct acquaintance in the intended sense.)

The act/object account of sensory experience combines with various considerations traditionally grouped under the head of the argument from illusion to provide arguments for representative realism, or more precisely for the clause in it that contends that our sensorily derived information about the world comes indirectly, that what we are most directly acquainted with is not an aspect of the world but an aspect of our mental, sensory response to it. Consider, for instance, the familiar refractive illusion mentioned already, that of a straight stick in water looking bent. The act/object account holds that in this case we are directly acquainted with a bent shape. This shape, so the argument runs, cannot be the stick as it is straight, and thus must be a

mental item, commonly called a sense-datum. And, in general, sense-data – visual, tactual, etc. – are held to be the objects of direct acquaintance. Perhaps the most striking use of the act/object analysis to bolster representative realism turns on what modern science tells us about the fundamental nature of the physical world. Modern science tells us that the objects of the physical world around us are literally made up of enormously many, widely separated, tiny particles whose nature can be given in terms of a small number of properties like mass, charge, spin and so on. (These properties are commonly called the primary qualities; *see* PRIMARY AND SECONDARY QUALITIES.) But of course, that is not how the objects look to us, not how they present to our senses. They look continuous and coloured. What then can these coloured expanses with which we are directly acquainted be other than mental sense-data?

Two objections dominate the literature on representative realism: one goes back to Berkeley (*see* BERKELEY) and is that representative realism leads straight to scepticism about the external world; the other is that the act/object account of sensory awareness is to be rejected in favour of an adverbial account.

Traditional representative realism is a “veil of perception” doctrine, in Bennett’s (1971) phrase. Locke’s idea (*see* LOCKE) was that the physical world was revealed by science to be in essence colourless, odourless, tasteless and silent, and that we perceive it by, to put it metaphorically, throwing a veil over it by means of our senses. It is the veil we see, in the strictest sense of “see”. This does not mean that we do not really see the objects around us. It means that we see an object in virtue of seeing the veil, the sense-data, causally related in the right way to that object. An obvious question to ask, therefore, is what justifies us in believing that there is anything behind the veil; and if we are somehow justified in believing that there is something behind the veil, how can we be confident of what it is like? (*See* PROBLEM OF THE EXTERNAL WORLD.)

There seems, however, to be a good answer to this fair question. The hypothesis

of the external world is the best explanation of the course of our sensory experience. The tracks in a Wilson cloud chamber justify believing in electrons because electrons are the best explanation of those tracks. On a vastly more massive scale, the history of our sense-data justify believing in an external world because the external world is the best explanation of the sensory history (*see* INFERENCE TO THE BEST EXPLANATION).

It might well be observed that this reply to scepticism fares better as a justification for believing in the existence of external objects, than as a justification of the views we have about their nature. It is incredible that nothing independent of us is responsible for the manifest patterns displayed by our sense-data, but granting this leaves open many possibilities about the nature of the hypothesized external reality. Direct realists often make much of the apparent advantage that their view has on the question of the nature of the external world. The fact of the matter is, though, that it is much harder to arrive at tenable views about the nature of external reality than it is to defend the view that there is an external reality of some kind or other. The history of human thought about the nature of the external world is littered with what are now seen (with the benefit of hindsight) to be egregious errors – the four element theory, phlogiston, the crystal spheres, vitalism, and so on. It can hardly be an objection to a theory that it makes the question of the nature of external reality much harder than the question of its existence.

The way we talk about sensory experience certainly suggests an act/object view. When something looks thus and so in the phenomenological sense, we naturally describe the nature of our sensory experience by saying that we are acquainted with a thus and so “given”. But suppose that this is a misleading grammatical appearance, engendered by the linguistic propriety of forming complex, putatively referring expressions like “the bent shape in my visual field”, and that there is no more a bent shape in existence for the representative realist to contend to be a mental sense-datum, than there is a bad limp in existence when someone has, as we say, a bad

limp. When someone has a bad limp, they limp badly. Similarly, according to adverbial theorists, when, as we naturally put it, I am aware of a bent shape, we would better express the way things are by saying that I sense bent shape-ly. What the act/object theorist analyses as a feature of the object which gives the nature of the sensory experience, the adverbial theorist analyses as a mode of sensing which gives the nature of the sensory experience. (The decision between the act/object and adverbial theories is a hard one, addressed in detail in e.g. Jackson, 1977; Chisholm, 1957; and Cornman, 1975. Jackson defends the act/object view, Chisholm and Cornman defend the adverbial view. *See* ACT/OBJECT ANALYSIS; ADVERBIAL THEORY.)

As we noted above, traditionally representative realism is allied with the act/object theory. But we can approach the debate between representative realism and direct realism via the notion of information processing. (Indeed, Mackie (1976, ch. 2) argues that Locke can be read as approaching the debate in this way.) I am watching a football game on television. My senses, in particular my eyes and ears, “tell” me that Carlton is winning. What makes this possible is the existence of a long and complex causal chain of electro-magnetic radiation running from the game through the television cameras, various cables, my television set and a region of space between my eyes and the television screen. Each stage of this process carries information about preceding stages in the sense that the way things are at a given stage depends on the way things are at preceding stages. Otherwise the information would not be transferred from the game to my brain. There needs to be a systematic covariance between the state of my brain and the state of the match, and that will not obtain unless it obtains between intermediate members of the long causal chain. For instance, if the state of my retina did not systematically covary with the state of the television screen before me, my optic nerve would have, so to speak, nothing to go on to tell my brain about the screen, and so in turn would have nothing to go on to tell my brain about the game. There is no “information at a distance”.

A few of the stages in this transmission of information between game and brain are special in the sense that I am in some sense perceptually aware of them. Much of what happens between brain and match I am quite ignorant about, some of what happens I know about from books, but some of what happens I am *perceptually* aware of. For instance, I am perceptually aware of the images on the screen. I am also perceptually aware of the game. Otherwise I could not be said to *watch* the game on television. Now my perceptual awareness of the match depends on my perceptual awareness of the screen. The former goes via the latter. In saying this I am not saying that I go through some sort of internal monologue like "Such and such images on the screen are moving thus and so, therefore, Carlton is attacking the goal". Indeed, if you suddenly covered the screen with a cloth and asked me (1) to report on the images, and (2) to report on the game, I might well find it easier to report on the game than on the images. But that does not mean that my awareness of the game does not go via my awareness of the images on the screen. It shows that I am more interested in the game than in the screen, and so am storing beliefs about it in preference to beliefs about the screen.

We can now see how to elucidate representative realism independently of the debate between act/object and adverbial theorists about sensory experience. Our initial statement of representative realism talked of the information acquired in perceiving an object being most immediately about the perceptual experience caused in us by the object, and only derivatively about the object itself. In the act/object, sense-datum approach, what is held to make that true is that the fact that what we are immediately aware of is a mental sense-datum. But instead, representative realists can put their view this way: just as awareness of the match goes via awareness of the screen, so awareness of the screen goes via awareness of experience, and in general when subjects perceive objects, their perceptual awareness always goes via awareness of experience.

Why believe such a view? Because of the point we referred to earlier: the picture of the

world provided by our senses is so very different from the picture provided by modern science. It is so different in fact that it is hard to grasp what might be meant by insisting that we are in epistemologically *direct* contact with that world.

See also ACT/OBJECT ANALYSIS; ADVERBIAL THEORY; DIRECT REALISM; EXPERIENCE; the GIVEN; PERCEPTUAL KNOWLEDGE.

#### BIBLIOGRAPHY

- Bennett, J.: *Locke, Berkeley, Hume: Central Themes* (Oxford: Oxford University Press, 1971).  
 Berkeley, G.: *The Principles of Human Knowledge* (1710), in *Berkeley: Philosophical Works*, ed. M. R. Ayers (London: Dent, 1975).  
 Chisholm, R. M.: *Perceiving* (Ithaca, NY: Cornell University Press, 1957).  
 Cornman, J.: *Perception, Commonsense and Science* (New Haven, CT: Yale University Press, 1975).  
 Jackson, F.: *Perception* (Cambridge: Cambridge University Press, 1977).  
 Locke, J.: *Essay Concerning Human Understanding* (1690), ed. P. H. Nidditch (Oxford: Oxford University Press, 1975).  
 Mackie, J. L.: *Problems from Locke* (Oxford: Oxford University Press, 1976).

FRANK JACKSON

**Rescher, Nicholas (1928– )** In various publications Rescher offers a detailed and systematic view of human knowledge and its limits along with core implications for value theory and ethics broadly conceived. In epistemology and philosophy of science, Rescher is best described as an analytic pragmatist (see PRAGMATISM) placing epistemic priority on the methods of the natural sciences as a source of both understanding the empirical world and directing our actions within it. Rescher regards science as seeking the best fit between the data of experience and the conjectures we make in our attempts to resolve questions. He sees scientific methods as the product of an evolutionary process of *rational* selection which leaves us with only those methods that have proved to work.

With regard to *foundational beliefs* or *basic knowledge*, Rescher asserts that basic beliefs, like all factual beliefs, are fallible and hence subject to revision in the light of ongoing evidence. Such beliefs begin as working presumptions and are accepted as true until experience requires rejection; but until experience forces such rejection they qualify for acceptance as items of human knowledge.

On the question of *non-basic knowledge* or *scientific knowledge*, he has argued in *Methodological Pragmatism* that while particular scientific theses established by the inductive methods of science may be false (although we must presume them to be true) *rationality* requires us to use such methods because they generally tend to produce more effectively supplementable beliefs about the physical world than any other methods available to us.

*Truth* Rescher construes in terms of correspondence and argues that the criterion for it is fully warrantably assertible belief. The satisfaction of the criterion does not entail logically that the proposition is true as construed, but it would be irrational to ask for anything more in the pursuit of truth.

*Human knowledge* should not be construed in terms of an impossibly idealized requirement assuring logical certainty. Such a requirement can only guarantee a global scepticism having no plausible justification.

Though Rescher ascribes a certain primacy to induction and the methods of the natural sciences because they are the product of the evolutionary process, he has not argued that the *only* answerable questions are those that admit of answer by appeal to the methods of the natural sciences; he has argued against that thesis. On the question of *scientific progress*, he has argued that unto eternity science is progressive and revolutionary, meaning thereby that there will never be a time when we would be justified in believing that we had answered all answerable questions about the world; but owing to an inevitable exponential decay in our economic capacity to fund scientific technology, scientific progress will accordingly slow, without stopping, to increasingly infrequent theoretical and factual advances.

On the question of *scientific realism*, Rescher has argued for a particular form of instrumentalism

in science without endorsing instrumentalism as a whole on the issue of factual knowledge. For Rescher, common-sense beliefs (those beliefs so obviously true that we cannot even imagine factual conditions under which they would be false) do succeed in correctly describing the physical world because such beliefs are not in any way likely to suffer truth value revision. Scientific beliefs, however, have no such property and must, for that reason, be regarded as instrumentally reliable beliefs which we can plausibly presume to be true.

Finally, the characteristic feature of Rescher's epistemology is its systematic integration of matters of value (i.e. norms – be they cognitive or affective) and matters of experientially determined fact. For Rescher morality is basically a matter of safeguarding the real or best interests of people, and while the identification of such interests involves an irreducibly normative element, the processes of their effective cultivation are something we can only learn about empirically. Morality thus weaves issues of fact and value into a seamless whole.

See also IDEALISM; PRAGMATISM.

#### WRITINGS

*The Coherence Theory of Truth* (Oxford: Clarendon Press, 1973).

*Empirical Inquiry* (Totowa, NJ: Rowman and Littlefield, 1981).

*Methodological Pragmatism* (Oxford: Blackwell, 1977).

*Scepticism* (Oxford: Blackwell, 1980).

*Scientific Progress* (Oxford: Blackwell, 1978).

*Scientific Realism* (Dordrecht: Reidel, 1987).

ROBERT ALMEDER

**Rorty, Richard (1931–2007)** American philosopher who has taught at Wellesley, Princeton and the University of Virginia.

Much of Rorty's early writing was in the philosophy of mind, where he became an influential defender of eliminative materialism. But beginning with *Philosophy and the Mirror*

of *Nature* (1979), his work has taken an increasingly radical turn. Rorty is a critic of the very idea of a “theory of knowledge” and hence of philosophy itself, conceived as a distinct, professionalized research activity with epistemological questions at its core. It is Rorty’s work that has done most to fuel speculation about the DEATH OF EPISTEMOLOGY.

Rorty presents a complex historical-explanatory argument in support of the thesis that epistemology and its problems are not a perennial concern, arising as soon as one reflects, but rather the products of a distinctive, historically contingent constellation of ideas. Developments that led to the emergence of epistemology include Descartes’ (see DESCARTES) redefinition of the “mind” as that to which each of us has privileged access; Locke’s (see LOCKE) suggestion that by investigating the Cartesian mind we can determine the scope and limits of human knowledge; and Kant’s (see KANT) thought that since all empirically knowable objects, “outer” as well as “inner”, are subject to conditions inherent in our cognitive constitution, we can have a priori knowledge (see A PRIORI KNOWLEDGE in Part I) of features necessarily characteristic of the world as we know it. The culmination of these developments is the idea of a non-empirical discipline that determines the cognitive status of all other forms of discourse – science, morality, art or religion – according to how well they represent reality, if indeed they represent it at all.

What human beings have put together, they can take apart. This, Rorty argues, should and indeed has begun to happen with the discipline we think of as the theory of knowledge. One idea that is absolutely crucial to this essentially neo-Kantian project is that there are two cleanly distinguishable components in knowledge: the factual element “given” to consciousness and the constructive or interpretative element contributed by the mind, or latterly language. But, taken together, Sellars’s attack (see SELLARS) on the “myth of the given” and Quine’s (see QUINE) scepticism about the language-fact distinction constitute a decisive rejection of these

indispensable ideas. Another idea, equally indispensable, is that the aim of thought or language is correspondence to reality, accuracy of representation. But in the light of Wittgenstein’s later philosophy, which approaches language through the notion of “use” rather than that of “picturing”, this idea too has come to seem increasingly dubious. This rejection of the “representationalist” model of mind and language becomes increasingly prominent in Rorty’s later writings and reflects his growing identification with the tradition of American pragmatism (see PRAGMATISM). It should be noted, however, that Rorty does not endorse a pragmatist definition of truth, holding with Donald Davidson that, with a primitive concept like truth, the real mistake is not to offer this or that particular analysis but to try to provide any kind of informative analysis at all.

#### WRITINGS

- Philosophy and the Mirror of Nature* (Princeton, NJ: Princeton University Press, 1979).
- Consequences of Pragmatism* (Minneapolis, MN: University of Minnesota Press, 1982).
- Contingency, Irony and Solidarity* (Cambridge: Cambridge University Press, 1989).
- Essays on Reality and Representation* (Cambridge: Cambridge University Press, 1991).

MICHAEL WILLIAMS

**rule-following** see PROBLEM OF RULE-FOLLOWING.

**Russell, Bertrand Arthur William (1872–1970)** British philosopher, mathematician and social campaigner. Russell himself used the phrase “theory of knowledge” to include a wide variety of topics: realism (see REALISM) and idealism (see IDEALISM), neutral monism, the nature of experience, the epistemology of language, the nature of truth and falsehood, induction and causation, and the nature of judgement and thought. Part of the explanation for this broad taxonomy is that, like many of his empiricist predecessors, his theory of



knowledge constrained his metaphysics. The question of the nature of such-and-such would turn into a question which can reasonably be classified as epistemological: what would such-and-such have to be like in order for us to know it? Moreover, he included under theory of knowledge issues which are now often classified under philosophy of mind or of language, those which concern the nature of the vehicles and sources of knowledge: mental states with content and perceptual experiences. For Russell, an account of experience was, for most of his philosophical career, inseparable from the question whether there is a genuine subject of experience, which is why even neutral monism, an apparently purely metaphysical doctrine, gets classified under theory of knowledge.

These close connections are one of the exciting features of Russell's philosophy. However, I will concentrate on his epistemology in the narrower sense current today; in particular, I will describe some of his views concerning our knowledge of the external world. (It is noteworthy that, despite the importance he attached to logicism in his early work, he has almost nothing to say about the nature of a priori knowledge.)

We can roughly divide Russell's working life into three periods: that belonging to the nineteenth century, when he had still not found the voice for which he is now famous; the middle period, occupying the first two decades of this century, in which most of his best-known views were formulated; and the late period, in which, although his output was still considerable, he was rather detached from the mainstream of academic philosophy. In the middle period, the problem of "knowledge of the external world" presented itself to him in an entirely traditional way: "I think on the whole the sort of method adopted by Descartes is right: that you should set to work to doubt things and retain only what you cannot doubt because of its clearness and distinctness" (Russell, 1918, p. 182). Our own experiences or sense-data resist doubt, and the problem is to give an account of how we in addition have knowledge of mind-independent continuants. He adopts three kinds of strategy.

One aims to narrow the gap between experience and material continuants by construing the latter as close in nature to the former; the other proceeds in the reverse direction, by enlarging the cognitive capacities available to the mind; the third allows a measure of scepticism, but tries to identify a kind of knowledge of physical continuants (merely structural) which is available without taking too narrow a view of their nature, or too generous a view of our capacities.

As an example of the first strategy, Russell (1914) suggests that material continuants are logical constructions (*see* LOGICAL CONSTRUCTION) out of sense-data, so they are at least of a kind with things which we can unproblematically know. Although Ayer (*see* AYER) has found some merit in this approach, it seems to me not to touch the problem as stated. A material continuant is construed as a very large set of sense-data, actual and merely possible, and no one subject's experience contains them all. Thus no subject can know any proposition of the form "This (material continuant) is thus-and-so" merely by knowing what sense-data he has. New principles of knowledge are involved, and these principles are no more plausible when they involve extrapolation to the existence of sense-data with which one will never be acquainted than when they involve extrapolation to material continuants which will never themselves be directly accessible to experience.

In his (1948) Russell deploys the second strategy, claiming that unless we have a priori knowledge of certain substantive contingent facts, then "science is moonshine" (1948, p. 524). These facts he called "postulates of scientific inference". An example is: "Given any event A, it happens very frequently that, at any neighbouring time, there is at some neighbouring place an event very similar to A" (p. 506). Russell implies that we do indeed have a priori knowledge of such facts, the knowledge being of a kind which he explains in terms of "animal expectation". This kind of knowledge is available to non-language users and is arguably non-propositional. This is a cognitive faculty often ignored in attempts to show how scepticism can be avoided.

The third strategy is hinted at in his (1912) and developed in more detail in his (1927). Russell lays down as a basic postulate that experiences are caused by something other than experiences – call the causes material events. Implicitly assuming some principle of like cause, like effect, he says that one can infer that properties of or relations among experiences mirror properties of or relations among material events. Material continuants are constructed out of material events. The upshot is that we know the structure of matter, but not its intrinsic nature. The strategy leaves room for scepticism about the real nature of material continuants, but is supposed to capture enough for an interpretation of science upon which most scientific beliefs are true.

Russell's views, different as they were among themselves, no doubt all deserve to be called foundationalist (see FOUNDATIONALISM). However, his (1948) is very explicit that a datum is to be defined merely as a proposition having some intrinsic credibility. A datum does not have to be certain, and does not have to be incapable of having its credibility raised by other propositions. Some standard objections to foundationalism will not touch this position.

In this same work, there is also a quite radical streak, suggesting that Russell came close to breaking out of his formulation of the problem. He stresses the importance of “animal expectation” as a model for the kind of knowledge we possess of the postulates of scientific inference. This is non-propositional, and a matter of the animal's causal relations with its environment. Here we see Russell operating in a different mode: naturalizing epistemology (see NATURALIZED EPISTEMOLOGY), rather than keeping within the confines of the Cartesian problematic.

See also KNOWLEDGE BY ACQUAINTANCE/BY DESCRIPTION; PHENOMENALISM; PROBLEM OF THE EXTERNAL WORLD; SANTAYANA.

#### WRITINGS

*The Problems of Philosophy* (London: Williams and Norgate, 1912).

“The Relation of Sense-data to Physics,” *Scientia* 16 (1914), 1–27; reprinted in his *Mysticism and Logic* (New York: Longmans, 1918; London: Allen and Unwin, 1963).

*Theory of Knowledge: The 1913 Manuscript* (London: Allen and Unwin, 1984).

*The Analysis of Matter* (London: Allen and Unwin, 1927).

*Human Knowledge: Its Scope and Limits* (London: Allen and Unwin, 1948).

#### BIBLIOGRAPHY

Ayer, A. J.: *Russell* (London: Fontana, 1972).

Chisholm, R. M.: “On the Nature of Acquaintance: A Discussion of Russell's Theory of Knowledge,” in *Bertrand Russell's Philosophy*, ed. G. Nakhnikian (London: Duckworth, 1974), 47–56.

Eames, E. R.: *Bertrand Russell's Theory of Knowledge* (London: Allen and Unwin, 1969).

R. M. SAINSBURY

**Ryle, Gilbert (1900–76)** British philosopher, who spent most of his working life at Oxford. Ryle rejected many of the ideas and issues of traditional epistemology. Just as his work in the philosophy of mind rejected the whole traditional framework of Cartesian dualism, so his detailed accounts of knowledge, belief and perception sought to move beyond that tradition.

Like many analytic philosophers during the 1950s and 1960s Ryle took the central role of philosophy to be that of “conceptual geography”. In such early papers as “Systematically Misleading Expressions” and “Categories” he constructed a framework for philosophy which was later the foundation for his pioneering work *The Concept of Mind*. He claimed that philosophical clarification has to do with the categorization of concepts, and is carried out for the sake of philosophers only and not for the benefit of the ordinary users of those concepts. In “Categories” he expresses a preference for the idea that philosophy has to do with the formulation of type liaisons, or type disciplines, rather than with the usual analysis associated with formal logic. Though he recognized a link between his category disciplines and Russell's theory of

types. Ryle insisted in the final chapter of *Dilemmas* that his own methods belonged to “informal” rather than to formal logic.

In those early papers the central thought is that the grammatical form of some expression may be “improper to the states of affairs it records” (1931–2, p. 142). When grammatical and logical form diverge in this way the expressions will be “systematically misleading” and will produce paralogisms and antinomies. These will then typically form the basis for traditional rivalries where the opposed theories are both wrong, since their disagreements rest on underlying confusions which an appropriate category discipline will resolve.

Ryle gives many examples of such simple category mistakes. The statement “Jones hates the thought of going to hospital” is not of the same logical type as “Jones hates Smith”. “Jones is popular” has a tacitly relational, rather than subject-predicate, form. The phrase “alleged murderer” does not ascribe a further property to one already classified as a murderer, just as the phrase “fictional country” does not identify a real country and characterize it with a further property. Though these may not be philosophically significant, or even contentious cases, Ryle believed that comparable errors were committed in traditional philosophy. In many such cases the error is that of treating a noun phrase as if it were a straightforward name for some identifiable object. He claims, for example, that the phrase “the meaning of . . .” is not referential and does not pick out any object as a meaning.

The most celebrated use of this apparatus came in Ryle’s attempt to demolish Cartesian dualism in *The Concept of Mind*. There Ryle offers the opposition between dualism and materialism as the erroneous consequence of a basic category mistake about mind. In *On Thinking* it is the opposition between dualism and behaviourism which is cited as the mistaken consequence. In both cases the category mistake is that of supposing that the phrase “the mind” and similar expressions for specific mental features name some specific object. With that mistaken assumption it seems inevitable that we should identify

minds either with physical objects (materialism) or with occult spiritual objects (dualism). Both such opposed theories, on Ryle’s view, failed to grasp the logical complexities and status of our mental language. This view is famously illustrated with Ryle’s account of the difference in logical status between expressions such as “the Oxford colleges” and “the university”. He thought that while specific colleges could be identified there was no comparable object for which the expression “the university” stood.

Though Ryle argued that there was one central category mistake at the heart of these traditional theories he examines in detail a wide variety of mental features, ranging from knowledge and belief to moods, emotions, sensations, memory and imagination. In that exploration he further deploys an ingenious array of informal devices, contrasting dispositions and occurrences, mongrel categoricals and hypotheticals, single-track and multi-track dispositions, task words and achievement words, in order to throw light on the logical complexities of our mental language. If there is one central point which his discussion seeks to make it is that while we do not believe that physical movements alone constitute mentality, or even evidence of mentality, nevertheless any additional feature required cannot be located within the occult framework of a Cartesian conception of mental substance.

At some points Ryle’s discussion impinges directly on issues of traditional epistemology. In *Dilemmas*, for example, he rejects a set of arguments for scepticism by insisting that descriptions such as “illusory”, or “fake”, make sense only where their contrasts, such as “veridical”, or “genuine”, make sense. In that argument he also sometimes seems to canvass the stronger, and fallacious, consequence that this establishes the existence and certainty of some veridical perception. In *The Concept of Mind*, too, his discussion of sensation and observation, about the success of which he expresses some doubt, criticises traditional theories of perception such as phenomenalism.

But perhaps the closest link with traditional epistemology is his discussion of “knowing

how” and “knowing that”. There he not only extends the traditional interest in knowledge, but also argues strongly against the traditional priority accorded to “knowing that”. He thought that this put too great a weight on intellectual theorizing, and wished instead to acknowledge the existence of other kinds of knowledge, particularly the practical skills of “knowing how”. He argued further that the traditional priority ought to be reversed, since even intellectual theorising presupposed some practical skill. His argument for this deployed characteristically a device of “informal logic” in claiming that to give priority to “knowing that” led to an infinite regress. For if intellectual knowledge required us to exercise practical mastery over the elements of some theory, and if that practical skill required another underlying theory to support it, then an endless regress would be bound to arise.

Ryle’s rejection of an extreme Cartesian dualism was successful, but critics objected that it had led him into an unacceptable behaviourism (see BEHAVIOURISM). His own text seemed to some ambiguous between two distinct theses: (A) the rejection of Cartesian dualism, and (B) the behaviourist rejection of *any* coherent notion of inner mental experience. Although Ryle’s official view treated behaviourism as an error on a par with its

Cartesian rival, nevertheless his own discussion seemed to some to canvass thesis (B).

See also BELIEF; DISPOSITION; INTROSPECTION; OTHER MINDS.

#### WRITINGS

“Systematically Misleading Expressions,” *Proceedings of the Aristotelian Society* 22 (1931–2), 139–70.

“Categories,” *Proceedings of the Aristotelian Society* 38 (1937–8), 189–206.

“Knowing How and Knowing That,” *Proceedings of the Aristotelian Society* 46 (1945–6), 1–16.

*The Concept of Mind* (London: Hutchinson, 1949). *Dilemmas* (Cambridge: Cambridge University Press, 1954).

*On Thinking*, ed. K. Kolenda (Oxford: Blackwell, 1979).

“Ordinary Language,” *Philosophical Review* 62 (1963), 167–86.

#### BIBLIOGRAPHY

Lyons, W.: *Gilbert Ryle: An Introduction to His Philosophy* (Hassocks: Harvester, 1980).

Wood, O. P. and Pitcher, G., eds: *Ryle* (London: Macmillan, 1970).

GRAHAM BIRD

# S

---

**safety** *see* SENSITIVITY AND SAFETY.

**Santayana, George (1863–1952)** American philosopher and writer, born in Spain. Santayana may well be the most resolute and most extreme realist in the history of thought (*see* REALISM). He maintained that both the immediate objects of consciousness (which he thought were universals) and the objects of knowledge (which he believed were entities in the material world) enjoy a reality independent of the mind. Universals have for him a timeless, self-identical being free from reliance on their embodiment in matter and envisagement by mind. They are infinite in number and constitute a realm of essence which includes every quality, every relation, and every form of definiteness.

In *Scepticism and Animal Faith* (1923), Santayana undertook to complete Descartes' project of universal doubt (*see* DESCARTES). He found that the resolute sceptic can doubt the existence of everything and is reduced to the unmeaning and non-existent vista of universals (essences) in present consciousness. Only this is certain but, since it is all transparent immediacy, it yields no knowledge of anything. Scepticism is, in this way, undefeatable, but it stands revealed as a self-stultifying mental exercise unfit to ground philosophy.

Santayana insists on the importance of honesty in philosophy, arguing that we should incorporate nothing into theory that we cannot act on in practice. Universal doubt fails to meet this criterion: the sceptic happily eats the food in whose existence he refuses to believe. Philosophy should consist of those beliefs that, in our actions, we constantly affirm.

This leads Santayana to a new, and hitherto inadequately appreciated, philosophical

method. The procedure he adopts consists of the critical reconstruction of the tenets of animal faith. Our actions reveal, and in a sense justify, certain commitments concerning the world; properly clarified and systematized, these constitute the philosophy of animal faith or the enduring commonsense of mankind.

To the animal in action, knowledge is belief or confidence mediated by the essences before the mind. In perception, we harness the immediate objects of consciousness and use them as symbols of surrounding material realities. The sign-significate relation between the contents of consciousness and the objects of knowledge enables Santayana to sidestep many of the problems of the correspondence theory of truth. The identity of the universal in mind with the one embodied in nature remains, for him, the ideal, but it is neither necessary for knowledge nor, given the discrepancy between subatomic reality and the scale of our senses, likely to be achieved.

Consciousness, for Santayana, is an irreducible by-product of material processes. It consists of a sequence of intentional acts directed upon essences. It lacks causal power and has as its sole function the grasp or intuition of non-existent forms. Although tethered to the organism, it provides escape from the cares of daily life by aesthetic immediacy or the disinterested contemplation of pure forms that Santayana calls "the spiritual life".

Santayana can be credited with three significant achievements. He pushed the sceptical enterprise farther than any other philosopher. He developed the promising new method of the philosophy of animal faith. Remarkably, he managed to combine in a single system a naturalistic view of human beings, a platonic theory of universals and an account of spirituality without theological trappings.

*See also* RUSSELL.



## WRITINGS

*The Life of Reason; or, The Phases of Human Progress*, I, Introduction and Reason in Common Sense, and V, Reason in Science (New York: Charles Scribner's Sons, 1905, 1906).

*Obiter Scripta* (New York: Charles Scribner's Sons, 1936).

*Physical Order and Moral Liberty*, ed. J. and S. Lachs (Nashville, TN: Vanderbilt University Press, 1969).

*Realms of Being* (New York: Charles Scribner's Sons, 1942).

*Scepticism and Animal Faith* (New York: Dover, 1955).

*Winds of Doctrine* (New York: Charles Scribner's Sons, 1926).

JOHN LACHS

**Sartre, Jean-Paul (1905–80)** Sartre lived and worked for the most part in Paris, and wrote novels and plays, as well as philosophy, on the main themes of existentialism, which borrow from and criticized elements of Husserl's phenomenology (see EXISTENCE; HUSSERL; PHENOMENOLOGY). After the Second World War he became increasingly involved in political issues and in the debates about Marxism.

In his main philosophical work, *Being and Nothingness*, Sartre examines the relationships between Being For-itself (consciousness), and Being In-itself (the non-conscious world). He rejects central tenets of the rationalist and empiricist traditions (see RATIONALISM; EMPIRICISM), calling the view that the mind or self is a thing or substance "Descartes's substantialist illusion" (1958, p. 84), and claiming also that consciousness does not contain ideas or representations of the material world ("representations . . . are idols invented by the psychologists"; 1958, p. 125) (see IDEA). Sartre also attacks idealism in the forms associated with Berkeley and Kant (see IDEALISM; BERKELEY; KANT), and concludes that his account of the relationship between consciousness and the world is neither realist (see REALISM) nor idealist (1958, Introduction).

Sartre also discusses Being For-others, which comprises the aspects of experience pertaining to interactions with other minds (see

OTHER MINDS). His views here are subtle: roughly, he holds that one's awareness of others is constituted by feelings of shame, pride, and so on.

Sartre's rejection of ideas, and the denial of idealism, appear to commit him to direct realism in the theory of perception (see DIRECT REALISM). This is not inconsistent with his claim to be neither realist nor idealist, since by "realist" he means views which allow for the mutual independence or in-principle separability of mind and world. Against this Sartre emphasizes, after Heidegger (see HEIDEGGER), that perceptual experience has an active dimension, in that it is a way of interacting and dealing with the world, rather than a way of merely contemplating it ("activity, as spontaneous, unreflecting consciousness, constitutes a certain existential stratum in the world"; 1962, p. 61). Consequently, he holds that experience is richer, and open to more aspects of the world, than empiricist writers customarily claim:

When I run after a streetcar . . . there is consciousness of-the-streetcar-having-to-be-over-taken, etc. . . . I am then plunged into the world of objects; it is they which constitute the unity of my consciousness; it is they which present themselves with values, with attractive and repellent qualities. . . .

(1957, p. 49)

Relatedly, he insists that I experience material things as having certain potentialities-for-me ("nothingnesses"). I see doors and bottles as openable, bicycles as rideable (these matters he links ultimately to the doctrine of extreme existentialist freedom). Similarly, if my friend is not where I expect to meet her, then I experience her absence "as a real event" (1958, p. 10).

These phenomenological claims are striking and compelling. But Sartre pays insufficient attention to such things as illusions and hallucinations, which are normally cited as problems for direct realists. In his discussion of mental imagery (1972), however, he describes the act of imaging as a "transformation" of "psychic material". This connects with his view that even a physical image such as a photograph of a tree does not figure

as an object of consciousness when it is experienced as a tree-representation (rather than as a piece of coloured card). But even so, the fact remains that the photograph continues to contribute to the character of the experience. Given this, it is hard to see how Sartre avoids positing a mental analogue of a photograph for episodes of mental imaging, and harder still to reconcile this with his rejection of visual representations. It may be that the regards imaging as debased and derivative perceptual knowledge (1972, ch. 2), but this merely raises once more the issue of perceptual illusion and hallucination, and the problem of reconciling them with direct realism.

## WRITINGS

*La Transcendance de l'Ego* (Paris: Vrin, 1936); trans. F. Williams and R. Kirkpatrick as *The Transcendence of the Ego* (New York: Hill and Wang, 1957).

*Esquisse d'une théorie des émotions* (Paris: Hermann, 1939); trans. P. Mairet as *Sketch for a Theory of the Emotions* (London: Methuen, 1962).

*L'Imaginaire* (Paris: Gallimard, 1940); trans. anon. as *The Psychology of Imagination* (London: Methuen, 1972).

*L'Être et le Néant* (Paris: Gallimard, 1943); trans. H. Barnes as *Being and Nothingness* (London: Methuen, 1958).

GREGORY MCCULLOCH

**scepticism** Scepticism is the view that we lack knowledge. It can be rather "local" (Pappas, 1978). For example, the view could be that we lack all knowledge of the future because we do not know that the future will resemble the past; or we could be sceptical about the existence of "other minds". But there is another view – the absolute global view that we do not have any knowledge whatsoever.

It is doubtful that any philosopher seriously entertained absolute global scepticism. Even the Pyrrhonist sceptics who held that we should refrain from assenting to any non-evident proposition had no such hesitancy about assenting to "the evident" (see

PYRRHONISM). The non-evident is any belief that requires evidence in order to be epistemically acceptable, i.e. acceptable because it is warranted. Descartes, in his sceptical guise, never doubted the contents of his own ideas (see DESCARTES). The issue for him was whether they "corresponded" to anything beyond ideas.

But Pyrrhonist and Cartesian forms of *virtual* global scepticism have been held and defended. Assuming that knowledge is some form of true, sufficiently warranted belief, it is the warrant condition, as opposed to the truth or belief condition, that provides the grist for the sceptic's mill. The Pyrrhonists will suggest that no non-evident, empirical proposition is sufficiently warranted because its denial will be equally warranted. A Cartesian sceptic will argue that no empirical proposition about anything other than one's own mind and its contents is sufficiently warranted because there are always legitimate grounds for doubting it. Thus, an essential difference between the two views concerns the stringency of the requirements for a belief's being sufficiently warranted to count as knowledge. A Cartesian requires certainty. A Pyrrhonist merely requires that the proposition be more warranted than its negation.

The Pyrrhonists do not *assert* that no non-evident proposition can be known, because that assertion itself is such a knowledge claim. Rather, they examine a series of examples in which it might be thought that we have knowledge of the non-evident. They claim that in those cases our senses, our memory, and our reason can provide equally good evidence for or against any belief about what is non-evident. Better, they would say, to withhold belief than to assent. They can be considered the sceptical "agnostics".

Cartesian scepticism, more impressed with Descartes' argument for scepticism than his own reply, holds that we do not have any knowledge of any empirical proposition about anything beyond the contents of our own minds. The reason, roughly put, is that there is a legitimate doubt about all such propositions because there is no way to justifiably deny that our senses are being stimulated by some cause (an evil spirit, for example) which

is radically different from the objects which we normally think affect our senses. Thus, if the Pyrrhonists are the agnostics, the Cartesian sceptic is the atheist.

Because the Pyrrhonist requires much less of a belief in order for it to be certified as knowledge than does the Cartesian, the arguments for Pyrrhonism are much more difficult to construct. A Pyrrhonist must show that there is no better set of reasons for believing any proposition than for denying it. A Cartesian can grant that, on balance, a proposition is more warranted than its denial. The Cartesian need only show that there remains some legitimate doubt about the truth of the proposition.

Thus, in assessing scepticism, the issues to consider are these: Are there ever better reasons for believing a non-evident proposition than there are for believing its negation? Does knowledge, at least in some of its forms, require certainty? And, if so, is any non-evident proposition certain?

See also CERTAINTY; DESCARTES; PYRRHONISM; SCEPTICISM, CONTEMPORARY.

#### BIBLIOGRAPHY

- Burnyeat, M., ed.: *The Sceptical Tradition* (Berkeley, CA: University of California Press, 1983).
- Descartes, R.: *Meditations on First Philosophy* (1641), in *The Philosophical Writings of Descartes*, ed. J. Cottingham, R. Stoothoff and D. Murdoch (Cambridge: Cambridge University Press, 1975), esp. meditation I.
- Klein, P.: *Certainty: A Refutation of Scepticism* (Minneapolis, MN: University of Minnesota Press, 1981).
- Moore, G. E.: "Four Forms of Scepticism," in his *Philosophical Papers* (New York: Macmillan, 1959).
- Pappas, G. S.: "Some Forms of Epistemological Scepticism," in *Essays on Knowledge and Justification*, ed. G. S. Pappas and M. Swain (Ithaca, NY: Cornell University Press, 1978), 309–16.
- Popkin, R.: *The History of Skepticism* (Berkeley, CA: University of California Press, 1979).
- Sextus Empiricus: *Outlines of Pyrrhonism*, trans. R.G. Bury (Cambridge, MA: Harvard University Press, 1976).

Unger, P.: *Ignorance: A Case for Scepticism* (Oxford: Oxford University Press, 1975).

Wittgenstein, L.: *On Certainty* (Oxford: Blackwell, 1969).

PETER D. KLEIN

**scepticism, contemporary** As with many things in contemporary philosophy, the current discussions about scepticism originate with Descartes' discussion of the issue, in particular, with his discussion of the so-called "evil spirit hypothesis" (see DESCARTES). Roughly put, that hypothesis is that instead of there being a world filled with familiar objects, there is just me and my beliefs and an evil genius who causes me to have just those beliefs that I would have were there to be the world which one normally supposes to exist. The sceptical hypothesis can be "up-dated" by replacing me and my beliefs with a brain-in-a-vat and brain-states and replacing the evil genius with a computer connected to my brain stimulating it to be in just those states it would be in were its states caused by objects in the world.

The hypothesis is designed to impugn our knowledge of empirical propositions by showing that our experience is not a reliable source of beliefs. Thus, one form of traditional scepticism developed by the Pyrrhonists, namely that reason is incapable of producing knowledge, is ignored by contemporary scepticism (see SCEPTICISM). Apparently, the sceptical hypothesis can be employed in two distinct ways. It can be used to show that our beliefs fall short of being certain and it can be used to show that they are not even justified. In fact, as we will see, the first use depends upon the second.

Letting "*p*" stand for any ordinary belief (e.g. there is a table before me) the first type of argument employing the sceptical hypothesis can be stated as follows:

1. If *S* knows that *p*, then *p* is certain.
2. The sceptical hypothesis shows that *p* is not certain.

Therefore, *S* does not know that *p*.

No argument for the first premiss is needed because this first form of the argument employing the sceptical hypothesis is only concerned with cases in which certainty is thought to be a necessary condition of knowledge (see CERTAINTY). Nevertheless, it could be pointed out that we often do say that we know something, although we would not claim that it is certain. In fact, Wittgenstein (1969) claims, roughly, that propositions which are known are always subject to challenge, whereas, when we say that *p* is certain, we are foreclosing challenges to *p*. As he puts it, “‘Knowledge’ and ‘certainty’ belong to different categories” (Wittgenstein, 1969, §308).

Both of these replies miss the basic point at issue – namely, whether ordinary empirical propositions are certain. A Cartesian sceptic could grant that there is a use of “know” – perhaps a paradigmatic use – such that we can legitimately claim to know something and yet not be certain of it. But it is precisely whether such a claim as “here’s one hand” is certain that is at issue. For if such propositions are not certain, then so much the worse for those propositions that we claim to know in virtue of being of being certain of our observations. The sceptical challenge is that, in spite of what is ordinarily believed, no empirical proposition is immune to doubt.

Implicit in the second premiss of the argument is a Cartesian notion of doubt which is roughly that a proposition, *p*, is doubtful for *S* if there is a proposition that (1) *S* is not justified in denying and (2) if added to *S*’s beliefs, would lower the warrant of *p* (see CERTAINTY). The sceptical hypothesis would lower the warrant of *p* if added to *S*’s beliefs. So it becomes clear that this first argument for scepticism will succeed just in case there is a good argument for the claim that *S* is not justified in denying the sceptical hypothesis.

That leads directly to a consideration of the second, more common, way in which the sceptical hypothesis has played a role in the contemporary debate over scepticism:

1. If *S* is justified in believing that *p*, then, since *p* entails the denial of the sceptical hypothesis, *S* is justified in believing the denial of the sceptical hypothesis.

2. *S* is not justified in denying the sceptical hypothesis.

Therefore, *S* is not justified in believing that *p*.

There are several things to note about this argument. First, if justification is a necessary condition of knowledge, this argument would succeed in showing that *S* does not know that *p*. Second, it explicitly employs the premiss needed by the first argument, discussed above, namely that *S* is not justified in denying the sceptical hypothesis. Third, the first premiss employs a version of the so-called transmissibility principle which probably first occurred in Edmund Gettier’s famous article (1963) (see GETTIER PROBLEM). Fourth, it is clear that *p* does, in fact, entail the denial of the most natural construction of the sceptical hypothesis since that hypothesis includes the statement that *p* is false. Fifth, the first premiss can be reformulated using some epistemic notion other than justification; in particular, with the appropriate revisions, “knows” could be substituted for “is justified in believing”. As such, the principle will fail for *uninteresting* reasons. For example, if belief is a necessary condition of knowledge, since we can believe a proposition without believing all of the propositions entailed by it, it is clear that the principle is false. Similarly, the principle fails for other *uninteresting* reasons. For example, if the entailment is a very complex one, *S* may not be justified in believing what is entailed because *S* does not recognize the entailment. In addition, *S* may recognize the entailment but believe the entailed proposition for silly reasons. But the *interesting* question is this: If *S* is justified in believing (or knows) that *p*, and *p* obviously (to *S*) entails *q*, and *S* believes *q* on the basis of believing *p*, then is *S* justified in believing (or in a position to know) that *q*?

The contemporary literature contains two general responses to the argument for scepticism employing an interesting version of the transmissibility principle. The most common is to challenge the principle. The second claims that the argument will, of necessity, beg the question against the anti-sceptic.

Nozick (1981), Goldman (1986), Thalberg (1974), Dretske (1970) and Audi (1988) have objected to various forms of the transmissibility

principle. Some of the arguments are designed to show that the first premiss with "knowledge" substituted for "justification" is interestingly false. But it is crucial to note that even if the principle, so understood, were false, as long as knowledge requires justification, the argument given above could still be used to show that  $p$  is beyond our ken because the belief that  $p$  would not be justified. Equally important, even if there is some legitimate conception of knowledge in which it did not entail justification, the sceptical challenge could simply be reformulated in terms of justification. For not being justified in believing that there is a table before me seems as disturbing as not knowing it.

The arguments against the first premiss take two forms: Some are based on supposedly clear-cut counterexamples to the principle; some are based on general theories of justification or knowledge which entail that the transmissibility principle is false.

Audi (1988) presents a plausible counterexample: Consider a person who adds a column of numbers enough times to be justified in believing that the sum is 1066. Now, *if the transmissibility principle is correct*,  $S$  is justified in believing that an expert arithmetician is wrong who has added the column and believes that column sums to 1067. But, so the objection continues,  $S$  is not justified in believing that the expert arithmetician is wrong. Dretske (1970) and Thalberg (1974) present similar counterexamples.

Nozick (1981) and Goldman (1986) develop general theories of knowledge which have the consequence that the transmissibility principle fails in an interesting way when "knowledge" is substituted for "is justified in believing". There are many important differences in their accounts, but very roughly put, the claim is that in order for a belief to count as knowledge, the mechanisms which produce our beliefs must be reliable. That is, those mechanisms must produce true beliefs in the actual as well as in a defined subset of the possible circumstances. Thus, we know that  $p$  but we are not in a position to know the denial of the sceptical hypothesis because there are reliable processes which produces the belief that  $p$

but none which reliably produce the belief that the sceptical hypothesis is false.

Both the counterexamples to the principle and the general theory of knowledge which has the consequence that the principle fails for interesting reasons seem to me to ignore what makes the principle intuitively plausible. It is plausible precisely because deduction is the very best way of expanding the corpus of what is justified or known. One is justified in believing some consequence of  $p$  or one is in a position to know some consequence of  $p$  because  $p$ , itself, is justified or known. Thus, if  $S$  is justified in believing that  $p$ , and  $p$  entails the denial of the sceptical hypothesis, then  $S$  is justified in denying that hypothesis because  $S$  has an adequate reason available, namely,  $p$ .

To return to the purported counterexample, if  $S$  is justified in believing that the sum is 1066, then  $S$  is justified in believing that anyone who says that it is 1067 is wrong because  $S$  has an adequate reason for that belief – namely that the sum is 1066. Of course, upon recognizing the entailment,  $S$  might re-evaluate the epistemic status of the original proposition that the sum is 1066. But that is a vindication of the principle. For  $S$ 's re-evaluation is based upon recognizing that if the original proposition is justified, so is the entailed proposition. But, as  $S$  now reasons, the latter is not justified. So, there is no time when  $S$  is both justified in believing that the sum is 1066 and not justified in believing that the expert is wrong.

Thus there is a plausible defence of the transmissibility principle that depends upon an internalist account of justification (see EXTERNALISM/INTERNALISM). But that very defence shows that the argument employing the transmissibility principle will beg the question. For if the transmissibility principle is true, then some form of the internalist account of justification must be correct and what justifies  $S$  in denying the sceptical hypothesis is  $p$  itself. But, then, an argument for the second premiss, i.e. that  $S$  is not justified in denying the sceptical hypothesis, must be sufficient to show that  $S$  does not have adequate "internal" reasons to deny the sceptical hypothesis. But since  $p$  would be such an



adequate reason, the argument for the second premiss would, of necessity, have to show that *S* is not justified in believing *p*. That, of course, is just the conclusion of the argument. Thus, there is an adequate response to the argument for scepticism which employs the transmissibility principle (see Klein, 1981).

A discussion of contemporary scepticism would not be complete without mentioning the views of Putnam (1981) (see PUTNAM) and Davidson (1986) (see DAVIDSON) concerning the sceptical hypothesis. They do not focus on the epistemic notions of certainty or justification. Rather, they have argued that the hypothesis, itself, must be false.

Their arguments differ in many important respects, but there is one central feature common to them both. Subtleties aside, it is this: the beliefs which we have depend upon the objects in our surroundings. Beliefs have a causal history such that it is not possible that most of them are false. Putnam argues that it is not possible, for example, for us to have the belief that water is  $H_2O$  if, in fact, no one has ever experienced water. Roughly, if humans had always been on a planet which contained no water, our beliefs could not be about *water*. Again, subtleties aside, Davidson's view is that our beliefs must be true in the main because "what stands in the way of global scepticism of the senses . . . is the fact that we must, in the plainest and methodologically most basic cases, take the objects of belief to be the causes of those beliefs" (Davidson, 1986, pp. 317–18).

There are at least four plausible responses to this way of regarding the sceptical hypothesis. First, even if some subset of my beliefs could not be false at some point in time because of the way in which they arise, a very strong form of scepticism remains possible. I could have acquired my beliefs just as the causal accounts dictate and then, later, I could be placed in the scenario described in the sceptical hypothesis and, from then on, my beliefs about my immediate environment would be in the main false. Second, one could argue that Davidson and Putnam are not entitled to employ claims about the causal order of things in presenting an argument against scepticism, because it is exactly

knowledge about the causal order which the sceptic is questioning. Third, if beliefs (as opposed to wishes, for example) are just those (mental) states of humans which have their intentional objects as their typical causes, then we cannot presuppose the existence of beliefs in an argument against scepticism. The Cartesian can do so because he/she thinks that types of mental states can be differentiated through "introspection". Thus, the sphere of the doubtful has increased because, now, one may have doubts that there are beliefs. Finally, one could hold that if an account of meaning is such that it makes global scepticism a priori impossible, then that view of meaning is *ipso facto* false. I have argued for the first three objections to this approach to scepticism (Klein, 1986). The final objection is due to McGinn (1986).

See also GETTIER PROBLEM; KNOWLEDGE AND BELIEF; PROPOSITIONAL KNOWLEDGE; RELIABILISM; SCEPTICISM; SEXTUS EMPIRICUS; TRANSCENDENTAL ARGUMENTS; and SCEPTICISM AND CLOSURE in Part I.

## BIBLIOGRAPHY

- Audi, R.: *Belief, Justification, and Knowledge* (Belmont, CA: Wadsworth, 1988).
- Davidson, D.: "A Coherence Theory of Truth and Knowledge," in *Truth and Interpretation*, ed. E. LePore (New York: Blackwell, 1986), 307–19.
- Dretske, F.: "Epistemic Operators," *Journal of Philosophy* 67 (1970), 1003–13.
- Gettier, E.: "Is Justified True Belief Knowledge?," *Analysis* 23 (1963), 121–3.
- Goldman, A. I.: *Epistemology and Cognition* (Cambridge, MA: Harvard University Press, 1986).
- Klein, P.: *Certainty: A Refutation of Skepticism* (Minneapolis, MN: University of Minnesota Press, 1981).
- Klein, P.: "Radical Interpretation and Global Scepticism," in *Truth and Interpretation*, ed. E. LePore (New York: Blackwell, 1986), 369–86.
- Lehrer, K.: *Knowledge* (Oxford: Oxford University Press, 1974).
- McGinn, C.: "Radical Interpretation and Epistemology," in *Truth and Interpretation*, ed. E. LePore (New York: Blackwell, 1986), 356–68.
- Nozick, R.: *Philosophical Explanations* (Cambridge, MA: Harvard University Press, 1981).

- Putnam, H.: *Reason, Truth and History* (Cambridge: Cambridge University Press, 1981).
- Thalberg, I.: "Is Justification Transmissible through Deduction?," *Philosophical Studies* 25 (1974), 357–64.
- Unger, P.: *Ignorance: A Case for Scepticism* (Oxford: Oxford University Press, 1975).
- Wittgenstein, L.: *On Certainty* (Oxford: Blackwell, 1969).

PETER D. KLEIN

**scepticism, modern** Modern scepticism began in the sixteenth century with the revival of knowledge of and interest in ancient Greek Pyrrhonian scepticism, as posed in the writings of Sextus Empiricus (see *SEXTUS EMPIRICUS*), and Academic scepticism as presented in Cicero's *De Academica*. The term "sceptic" was not used in the Middle Ages, and was first just transliterated from the Greek. Sextus' writings were published in Latin in 1562 and 1569, and in Greek in 1621. Editions of Cicero's text appeared during the sixteenth century. The revival of these writings took place at a time when a fundamental issue concerning religious knowledge had been raised by the Reformation and Counter-Reformation – how does one distinguish true religious knowledge from false or dubious views? Erasmus denied that this could be done, and advised following the sceptics in suspending judgement, *and* accepting the views of the Catholic Church on questions in dispute. The translator of Sextus, Gentian Hervet, a Catholic priest, said that the views of the Pyrrhonists constituted the perfect and complete answer to Calvinism. If nothing can be known, then Calvinism cannot be known. Counter-Reformers used the sceptical arguments to constitute a "machine of war" against their Protestant opponents, and Protestants sought to show that the Catholics would undermine their own views because of the same sceptical challenges.

The most important presentation of scepticism at the time was that of Montaigne, as presented in his "Apology for Raimond Sebond" (see *MONTAIGNE*). Montaigne had studied and been overwhelmed by the arguments in Sextus and Cicero. He put them

together in his long, rambling essay, and modernized them to fit sixteenth-century intellectual concerns. He also stated them in a vernacular language (French), which provided the vocabulary for modern discussions of the problem of knowledge.

Montaigne's questioning of the evidence for any knowledge claims, the adequacy of any purported criterion of knowledge, and the possibility of any universal ethical standards posed challenges to all views then being presented. Montaigne's work became a best-seller in France, and in English translation. With rising doubts about the prevailing intellectual tradition, Montaigne's effort set forth a general scepticism, not just against scholasticism or Renaissance naturalism, but against the possibility of there being any system of ideas that could not be cast in doubt. Montaigne's disciple, Father Pierre Charron, presented scepticism in a didactic form that was very widely read.

Early seventeenth-century philosophers tried to formulate answers to the new scepticism as a way of grounding modern philosophical theories that could justify the new science. Bacon (see *BACON*), Mersenne, Gassendi, Descartes (see *DESCARTES*), Pascal, among others, tried to deal with the sceptical menace that was overwhelming the intellectual world.

Mersenne and Gassendi in different ways formulated a mitigated or constructive scepticism, conceding a major part of the sceptic's case, while still contending that some kind of limited knowledge was possible and useful. Mersenne, in a dialogue with a sceptic, who set forth arguments from Sextus, said that although one could not answer the basic challenges of the sceptics, it did not matter because one did in fact have ways of dealing with the questions. One could predict from one experiential situation what would follow although one did not know the *real* causes of events. One could doubt whether any metaphysical knowledge is possible, while still developing a science relating appearances to appearances.

Gassendi carried this further in what he called a *via media* between scepticism and dogmatism. He developed a hypothetical Epicurean atomic theory relating appearances

together. This would provide a shadow of truth, rather than Truth itself.

Descartes (*see* DESCARTES) was unwilling to settle for this limited certitude. He sought truths that no sceptics could challenge. To find these, he first adopted a sceptical method of doubt, rejecting any beliefs that might, under any conceivable conditions, be false or dubious. He quickly rejected sensory beliefs, because our senses sometimes deceive us. He rejected beliefs about physical reality, because what we take to be such a reality may only be part of dream. He rejected beliefs based on reasoning, because we may be systematically deceived by some demonic force.

At this point, Descartes appeared to have created a greater scepticism than that of Montaigne. But Descartes went on to ask whether one can doubt or reject one's belief in one's own existence. Here we find that any attempt to do so is immediately overridden by our awareness that we, ourselves, are doing the doubting. So, the first truth Descartes claimed that could not be doubted was "I think, Therefore I am" – the *cogito* (*see* COGITIO). From this truth one could elicit the criterion that whatever we clearly and distinctly conceive is true. Using the criterion, we establish that God exists, that He is all-powerful, the creator of everything that exists, and that because He is perfect, He cannot deceive us. Therefore, whatever God makes us believe clearly and distinctly must be true. In this way Descartes' new philosophy is intended to rebut the new scepticism.

Descartes' system became the main target of the modern sceptics. It was criticized by Gassendi, Hobbes (*see* HOBBS) and Mersenne, as based on unjustified and unjustifiable dogmas. Why could an all-powerful God not deceive? How do we know that there is not a truth for God or angels that is different from what we are forced to accept as true? Why must what we clearly and distinctly conceive be true in reality, and be not just in our minds? How do we know that our entire subjective picture of the world, no matter how certain, is not just our illusion? Descartes retorted that to take such questions seriously was to shut the door on reason.

But this argument from catastrophe did not really answer the sceptical challenges.

In the next generation, highly detailed sceptical analyses of the questionable portions of Descartes' philosophy appeared. Pierre-Daniel Huet sought to show that all of Descartes' views, including the *cogito*, were open to doubt. Simon Foucher directed a similar attack against Malebranche as soon as the latter's philosophy was published. Foucher also fought with Leibniz's attempt to found a dogmatic system. The culmination of seventeenth-century scepticism appeared in the writings of Pierre Bayle, especially in his *Historical and Critical Dictionary* (1697–1702). Bayle put together all kinds of doubts to undermine both ancient and modern philosophy. He raised devastating sceptical challenges against Cartesianism, against Leibniz's new rationalism, against any and all contenders. The arguments in Bayle's *Dictionary*, especially in his articles on the Greek sceptic Pyrrho of Elis (*see* PYRRHONISM), and on Zeno of Elea, posed central problems for the next generation of philosophers.

Locke (*see* LOCKE) offered a way of evading scepticism by admitting that one could have no *real* knowledge beyond intuition and demonstration, but that nobody was so mad as to doubt that fire is hot, that rocks are solid, etc. Experience would defeat scepticism. Locke's critic, Bishop Stillingfleet, tried to show that his empiricism would end up in scepticism. Berkeley (*see* BERKELEY), raised on Bayle's arguments, saw that they could be turned against Locke's philosophy. Bayle had already shown that the distinction between primary and secondary qualities was indefensible. If the latter were subjective, and just in the mind, so were the former. Berkeley pressed this point to drive Locke's view into total scepticism. Berkeley claimed to have found an answer to scepticism in insisting that appearance is reality, whatever is perceived is real.

Hume (*see* HUME), a devoted Bayle reader, developed a more encompassing scepticism. We could know nothing beyond impressions and ideas. Our causal knowledge, all that takes us beyond our immediate experience, is based on no rational or defensible principle, but just

on a natural and unalterable psychological tendency to expect future experiences to resemble those we have had in the past. Any attempt to defend our unavoidable beliefs in causes, in the external world, or in a constitutive real self within us, end in absurdity and contradiction. So we are driven by any investigation of our beliefs in to total scepticism, but nature will not let us rest in this; we cannot help believing. So, Hume concludes, it is by animal faith that we get on in life, and calm our irresistible sceptical doubts.

Hume's scepticism was met by two sorts of answers that have played important roles in contemporary theories of knowledge, the Commonsense realist response of Thomas Reid (see REID) and the critical theory of Immanuel Kant (see KANT). Reid, a contemporary of Hume, insisted that though one could not answer the sceptical problems raised, no one really had doubts about the existence of causes, the external or internal world. Our common-sense leads us to positive views on these scores, and when commonsense conflicts with philosophy, one has to reject the philosophical conclusions. Hume indicated he agreed with Reid, but did not find this an answer to scepticism. In Hume's view this forced belief is a psychological fact of life, but not an anti-sceptical argument.

Kant said that Hume awoke him from his dogmatic slumbers, and made him see how parlous our knowledge claims are. But he insisted Hume had asked the wrong question. We do have unquestionable knowledge that tells us something about all possible experience, such as that all experience will be temporal and spatial. How is such knowledge possible, if we cannot get beyond our world of experience? Kant insisted that experience is the combination of how we project it and its content. There are forms of all possible perceptions, and there are categories through which we make judgements about all possible experiences. Whether these correspond to a world beyond experience, we can not know, but we can analyse what we can be sure of about possible experience. Hence, we can have some kind of knowledge, but not knowledge of things-in-themselves.

Kant proposed his critical philosophy as a way of resolving the sceptical problems imbedded in modern philosophy. He was quickly attacked as being himself just a very high-brow sceptic, since he too ended up denying our ability to have necessary knowledge about the world. German philosophy of the next half century consisted of attempts to avoid or overcome the scepticism implicit in Kant's analysis of the knowledge situation.

See also COMMONSENSISM AND CRITICAL COGNITIVISM; IN ITSELF/FOR ITSELF; NOUMENAL/PHENOMENAL; SEXTUS EMPIRICUS.

# BIBLIOGRAPHY

- Bayle, P.: *Historical and Critical Dictionary, Selections*, trans. R. H. Popkin (Indianapolis, IN: Hackett, 1991).
- Charron, P.: *Toutes les Oeuvres de Pierre Charron*, dernière édition (Paris: J. Villery, 1635).
- Descartes, R.: *The Philosophical Writings of Descartes*, ed. J. Cottingham, R. Stoothoff and D. Murdoch, 2 vols (Cambridge: Cambridge University Press, 1975).
- Gassendi, P.: *The Selected Works of Pierre Gassendi*, trans. and ed. C. Brush (New York: Johnson Reprint, 1972).
- Huet, P.-D.: *Traité Philosophique de la Faiblesse de l'Esprit Humain* (Amsterdam, 1723).
- Hume, D.: *A Treatise of Human Nature* (1739–40), ed. L. A. Selby-Bigge, revised by P. H. Nidditch (Oxford: Oxford University Press, 1978).
- Hume, D.: *Dialogues Concerning Natural Religion* (1779), ed. R. H. Popkin (Indianapolis, IN: Hackett, 1988).
- Mersenne, M.: *La Verité des Sciences Contre les Sceptiques ou Pyrrhoniens* (Paris: T. Du Bray, 1625).
- Montaigne, M. de: "Apology for Raimond Sebond," in *The Complete Works of Montaigne*, trans. D. Frame (Stanford, CA: Stanford University Press, 1958).
- Pascal, B.: *Selections*, trans. R. H. Popkin (New York: Macmillan, 1989).
- Sextus Empiricus: *Adversus Mathematicos . . . graece nunquam latine . . .*, ed. G. Hervet (Paris: Martin Juvenem, 1569).
- Sextus Empiricus: *Outlines of Pyrrhonism and Adversus Mathematicos*, 4 vols (Cambridge, MA: Harvard University Press, 1917–55).

RICHARD POPKIN



**Schlick, Moritz (1882–1936)** Schlick was the founder of what became known as the Vienna Circle (see VIENNA CIRCLE) and, with Rudolf Carnap and later Otto Neurath, its leading spirit. He had trained under Max Planck as a theoretical physicist, and was one of the first philosophers to understand and write on Einstein's theories of relativity. He was thus ideally qualified to lead the logical positivists (see LOGICAL POSITIVISM).

Schlick's most substantial contribution to epistemology is his early work, *General Theory of Knowledge*. At this time Schlick was not a strict positivist at all, though not a Kantian either. (In 1919, after the volume appeared, Einstein observed that Schlick would find it hard to obtain a professorship as "he does not belong to the philosophical established church of the Kantians".) *The General Theory* is empiricist (see EMPIRICISM) in that it rejects the possibility of Kantian synthetic a priori knowledge (see A PRIORI KNOWLEDGE in Part I; KANT), but defends a version of what might be called "structural realism", which Schlick termed "physicalism". Science deals with real, unobservable entities, which Schlick even calls "things-in-themselves", and which are not mere logical constructions from sense-data. His account of knowledge and judgement is explicitly "holistic", and knowledge or cognition ("Erkennen") is sharply separated from acquaintance ("Kennen") or experience ("Erleben") of the "given"; so Russell's notion of knowledge by acquaintance becomes incoherent (see KNOWLEDGE BY ACQUAINTANCE/BY DESCRIPTION). Knowledge must be propositional in nature, requiring concepts and judgements; mere acquaintance with the given is irredeemably subjective (see GIVEN). On self-knowledge, Schlick finds an additional fact, the "unity of consciousness", to be "directly given"; but though expressly anti-Humean, the account is also non-Kantian in that this "fact" seems to be empirical. Machian positivism – which tends to equate what is real with what is given in experience – receives a sympathetic diagnosis, but is firmly rejected.

Under the influence of Hilbert, the *General Theory* argued for an account of necessary propositions in geometry (see GEOMETRY),

arithmetic, logic and metaphysics as "conventions of symbolism" rather than synthetic a priori truths. However, it was Wittgenstein's *Tractatus* (see WITTGENSTEIN) rather than the work of Schlick which to the Vienna Circle provided the basis for a "consistent empiricism" (one which could give a plausible account of logical necessity). (See LOGICAL POSITIVISM; MATHEMATICAL KNOWLEDGE.)

The standpoint of the *General Theory* illustrates (as Friedman (1983) has argued) how the doctrines of the Vienna Circle were not simply a linear development from classical empiricism, but incorporated elements of the Kantian tradition. However, under the influence of his understanding of Relativity Theory, and later of Wittgenstein, Schlick's writings after the *General Theory* take on a more verificationist and positive tone (see VERIFICATIONISM). In the 1930s, he advocated a linguistic version of Mach's equation of the given and the real with his Verification Principle, which received varied definitions in the Vienna Circle; Schlick's tended to be liberal. In "Positivism and Realism" (1932), "On the Foundation of Knowledge" (1934) and "Facts and Propositions" (1935) he argued, against the coherentist objections of Neurath, for the existence of an incorrigible foundational class of *Konstatierungen* or "affirmations" (see COHERENTISM; NEURATH; FOUNDATIONALISM). What Schlick wanted to claim was that sentences *could* be compared with something distinct from other sentences; that statements may be warranted by experiences. Confusingly, though, the "affirmations" occupied an ambiguous position mid-way in character between judgements and sensations. The change in Schlick's epistemology should not be overstated; there was an enduring and unresolved tension in his philosophy between a naturalist realism and an empiricist, positivist foundationalism (see Gower, 1992).

Under the influence of Wittgenstein and Carnap (see CARNAP), Schlick came to regard metaphysics as consisting of "pseudo-problems", and, for instance, treated the dichotomy of idealism and realism as a pseudo-inconsistency between sentences devoid of cognitive significance. But he never accepted



Carnap's conception of philosophy as merely the "logic of science".

#### WRITINGS

*General Theory of Knowledge* (1918), trans. A. Blumberg (LaSalle, IL: Open Court, 1974).  
*Collected Philosophical Papers*, vol. I (1909–22), ed. H. L. Mulder and B. F. van de Velde-Schlick (Dordrecht: Reidel, 1978).  
 "Positivism and Realism" (1932), "On the Foundation of Knowledge" (1934) and "Facts and Propositions" (1935), in *Collected Philosophical Papers*, vol. I (1925–36), ed. H. L. Mulder and B. F. van de Velde-Schlick (Dordrecht: Reidel, 1979), and in *Essential Readings in Logical Positivism*, ed. O. Hanfling (Oxford: Blackwell, 1981).

#### BIBLIOGRAPHY

Friedman, M.: "Critical Notice: Moritz Schlick, *Philosophical Papers*," *Philosophy of Science* 50 (1983), 498–514.  
 Gower, B.: "Realism and Empiricism in Schlick's *Philosophy*," in *Science and Subjectivity*, ed. D. Bell and W. Vossenkuhl (Berlin: Akademie Verlag, 1992).

ANDY HAMILTON

**scientia media** *see* SUAREZ.

**self-consciousness** The expression "self-consciousness" can mean different things. In the sense (I) "consciousness of self" it refers to the awareness a subject (of experience) has of itself i.e. of the typical referent of the pronoun *I*. It is not merely a grasp of the entity that happens to be myself, but an awareness of myself known as myself. The philosophical issues here revolve around how such awareness is generated and what its logical structure is. Alternatively, self-consciousness can be (2) "experience of the items in one's consciousness or the contents of consciousness" like sensations, thoughts, feelings etc. This leaves open the possibility of such awareness being a result of the special faculty of introspection. However, there is a use of self-consciousness that refers to the "self-intimation" of every

conscious state and in this sense it means (3) the "ability of a conscious state to become an object to itself". The philosophical problem here is to cash out in epistemic and metaphysical terms the metaphor of "phosphorescence" that is generally used to capture this reflexivity of consciousness.

*See also* INTROSPECTION; SELF-KNOWLEDGE AND SELF-IDENTITY.

#### BIBLIOGRAPHY

Anscombe, G. E. M.: "The First-person," in *Mind and Language*, ed. S. Guttenplan (Oxford: Clarendon Press, 1975).  
 Kant, I.: *Critique of Pure Reason*, trans. N. Kemp Smith (London: Macmillan, 1964).  
 Locke, D.: *Myself and Others* (Oxford: Clarendon Press, 1968), esp. ch. 2.  
 Ryle, G.: *The Concept of Mind* (London: Hutchinson, 1949).

VRINDA DALMIYA

**self-evidence** Self-evident propositions are those evident in themselves or known independently of all other propositions and evidence. To comprehend such a proposition is to be fully justified in believing it, or to know it. Such propositions might include first, some necessary truths of logic, for example laws of noncontradiction and identity; second, analytically true propositions, such as "bachelors are unmarried"; and third, some contingent propositions, such as "I exist" or "I am appeared to redly". The latter two classes are more controversial: the second because the analytic/synthetic distinction may not be clear-cut, and the third because such propositions may be justified in relation to experience, if not to other propositions.

*See also* ANALYTICITY; the GIVEN; KNOWLEDGE BY ACQUAINTANCE/KNOWLEDGE BY DESCRIPTION.

#### BIBLIOGRAPHY

Audi, R.: *Belief, Justification, and Knowledge* (Belmont, CA: Wadsworth, 1988), ch. 4.

Butchvarov, P.: *The Concept of Knowledge* (Evanston, IL: Northwestern University Press, 1970), 61–75.

Chisholm, R. M.: *Theory of Knowledge*, 3rd edn (Englewood Cliffs, NJ: Prentice-Hall, 1989), chs 3, 4.

ALAN H. GOLDMAN

**self-knowledge and self-identity** Normally the way one knows something about oneself is significantly different from the way one knows the same sort of thing about someone else. Knowledge of one's own current mental states is ordinarily not grounded on information about behaviour and physical circumstances. Knowledge of one's actions, and of such facts as that one is sitting or standing, is usually "without observation" or, at any rate, not based on the sorts of observations that ground one's knowledge of the actions and posture of others. One's perceptual knowledge of one's situation in the world, e.g. that one is facing a tree, differs markedly from the perceptual knowledge others have of the same facts, since it usually doesn't involve perceiving *oneself*. And one's memory knowledge of one's own past is normally very different from one's memory knowledge of the pasts of others; one remembers one's thoughts, feelings, perceptions and actions "from the inside", in a way that does not depend on the use of any criterion of personal identity to identify a remembered self as oneself.

Although in all these cases one could speak of a "special" first-person access, it is the access people have to their own mental states that has attracted the most attention. Some philosophers, e.g. Ryle (1949) (*see* RYLE), have denied that there is a fundamental difference between first-person and third-person knowledge of mental states. Others, most notably Wittgenstein (1953) (*see* WITTGENSTEIN), have maintained that where the difference seems most pronounced, e.g. in the case of pain ascriptions, the first-person "avowals" are not really expressions of *knowledge* at all (*see* AVOWALS). Such views are manifestations of the twentieth-century reaction against Cartesian views about self-knowledge

that are often associated with the claim that there are radical first-person/third-person asymmetries (*see* CARTESIANISM). These include the views that the mind is transparent to itself, that mental states are "self-intimating", that first-person ascriptions of mental states are infallible, and that self-knowledge of mental states serves as the foundation for the rest of our empirical knowledge (*see* FOUNDATIONALISM). Such views have been undermined by the work of Freud (*see* FREUD), with its postulation of a realm of unconscious wishes, intentions, etc., by work in cognitive psychology which shows most of the "information processing" in the mind to be unconscious and which shows many sorts of introspective reports to be unreliable (*see*, e.g., Nisbett and Wilson, 1977), and by philosophical criticisms of foundationalist accounts of knowledge. But most recent theorists who reject these Cartesian claims would agree that the reasons for their rejection are not reasons for denying that there is first-person knowledge of mental states that differs importantly from third-person knowledge of the same phenomena.

One question about such knowledge is whether it is appropriately thought of as observational, i.e. as grounded in a kind of perception that could be called "inner sense". Modern defenders of the view that such knowledge is observational (e.g. D. M. Armstrong, 1968) take perceiving something to be a matter of being so related to it that its having certain properties is apt to give rise to the non-inferential belief that there is something that has them. On this conception, it seems plausible to say that one perceives mental states and events occurring in one's own mind, in virtue of an internal mechanism by which mental states give rise to true beliefs about themselves, but cannot perceive those occurring in the minds of others, and that it is in this that one's "special access" to one's mind consists.

Some who agree with such a "reliable internal mechanism" view of introspective awareness would object to describing such knowledge as perceptual. In paradigm cases of perception, e.g. vision, the causal connection between the object perceived and the

perceiver's beliefs about it is mediated by a state of the perceiver, a "sense-experience", which in some sense represents the object, and which the subject can be aware of (in being aware of the look or feel of a thing).

There seem to be no such intermediaries between our sensations, thoughts, beliefs, etc. and our beliefs about them, and this seems a reason for denying that our awareness of them is perceptual.

A different objection questions the idea, implicit in the perceptual model, that there is only a contingent connection between having mental states and being aware of them just as there is only a contingent connection between there being trees and mountains and there being perceptual awareness of them). It makes doubtful sense to suppose that there are creatures that have pain without having any capacity whatever to be aware of their pains. And a consideration of the explanatory role of self-knowledge suggests that for many kinds of mental states the very capacity to have and conceive of such states involves immediate "first person access" to the existence of these states in oneself. To mention just one instance, if being a subject of beliefs and desires involves being at least minimally rational, and if rational revision of one's belief-desire system in the light of new experience requires some knowledge of what one's current beliefs and desires are, then being a subject of such states requires the capacity to be aware of them. While we should reject any self-intimation thesis strong enough to rule out the possibility of self-deception, or to deny mental states to animals and infants, it is far from obvious that the nature of mental states is distinct from their introspective accessibility in the way the observational model implies (see Shoemaker, 1988).

Lichtenberg denied that Descartes had a right to say "I think", claiming that he was only entitled to "It thinks". And Hume (1739) (see HUME) famously denied that when one introspects one finds any item, over and above one's individual perceptions, that could be the self or subject that "has" them. Such denials have led some (including Hume) to deny that there is any such self or subject, and

have led others to wonder how we can have knowledge of such a thing or refer to it with "I". Arguably, such denials lose their force if we abandon the observational model of self-knowledge; what is disturbing is the idea that we do perceive "by inner sense" perceptions, thoughts, etc., but do not perceive anything that could be their subject. Of course, if perceiving something is construed merely as being so related to it as to acquire, in a reliable way, true beliefs about it, then our capacity for self-knowledge involves our being able to perceive both individual mental events or states and the self (person) who has them (see Shoemaker, 1986).

The peculiarities of self-knowledge are, in any case, closely tied to the peculiarities of self-reference. If the amnesiac Joe Jones discovers that Joe Jones is the culprit, without realizing that he himself is Joe Jones, this will not be a case of self-knowledge in the sense that concerns us, even though it is a case in which the person known about is the knower himself. We are concerned with cases in which someone knows that *he himself*, or she *herself*, is so and so, where this is knowledge the knower would express by saying "I am so and so" (see Castaneda, 1968). One feature of first-person reference is that it in no way depends on the availability of individuating descriptions; one can refer to oneself with "I" without knowing of any descriptions that could be used to fix its reference. A related feature of "I"-judgements is their "immunity to error through misidentification" (see Shoemaker, 1968; Evans, 1982). In central cases of the use of "I" (what Wittgenstein (1958) called its use "as subject"), making the judgement "I am F" does not involve identifying as oneself the thing that is judged to be F; such a judgement may be mistaken (if the judger is not F), but one sort of mistake is ruled out: it cannot be that one is right in thinking someone to be F, but wrong in thinking that person to be oneself. This immunity to error through misidentification differs from that which characterizes judgements having demonstratives such as "this" as subject; where both "I am F" and "This is F" are immune to such error, the memory judgement "I was F" preserves the immunity

while the memory judgement "This was F" does not (for a qualification of this, see Shoemaker, 1986). This is related to the fact, already mentioned, that first-person memory judgements typically do not need to be grounded on any criterion of identity.

It is precisely where "I"-judgements are known in distinctively first-personal ways that they have this immunity to error through misidentification. And "I"-judgements that do not have this immunity (e.g. "I am bleeding", if inferred from the blood on the floor) always have among their grounds some that do (e.g. "I see blood" or "There is blood near me"). It is arguable that part of what gives first-person content to beliefs and other mental states is their relation to distinctively first-person ways of knowing, and that without such "special access" there could be no first-person reference at all (see Evans, 1982). (But another important feature of "I"-judgements is their intimate relation to action: the amnesiac Joe Jones will not be moved to action by learning that Joe Jones is in danger, but will be if he learns in addition that he is Joe Jones and so that *he himself* is in danger; see Perry, 1989.)

A stronger and more controversial claim is that the special access persons have to themselves enters into the very identity conditions for the sorts of things persons are. Many have argued, following Locke, that memory access is part of what determines the temporal boundaries of persons. A major determinant of the spatial boundaries of persons, i.e. of what counts as part of a person's body, is the extent of direct voluntary control, and this is intimately tied to the special epistemic access persons have to their own voluntary actions. And a familiar Kantian idea is that unity of consciousness – different states belonging to the same conscious subject – in some way involves consciousness, or the possibility of consciousness, of this unity.

See also APPERCEPTION; INTROSPECTION; KANT.

#### BIBLIOGRAPHY

Armstrong, D. A.: *A Materialist Theory of the Mind* (London: Routledge and Kegan Paul, 1968).

Evans, G.: "Self-identification," ch. 7 of his *The Varieties of Reference* (Oxford: Clarendon Press, 1982).

Hume, D.: *Treatise of Human Nature* (1739–40), ed. L. A. Selby-Bigge, revised by P. H. Nidditch (Oxford: Oxford University Press, 1978).

Nisbett, R. and Wilson, T. De C.: "Telling More than We Know: Verbal Reports on Mental processes," *Psychological Review* 84 (1977), 231–59.

Perry, J.: "The Essential Indexical," *Nous* 13 (1989), 3–21.

Ryle, G.: *The Concept of Mind* (London: Hutchinson, 1949).

Shoemaker, S.: "Self-reference and Self-awareness," *Journal of Philosophy* 65 (1968), 555–67.

Shoemaker, S.: "Introspection and the Self," *Midwest Studies in Philosophy* 11 (1986).

Shoemaker, S.: "On Knowing One's Own Mind," *Philosophical Perspectives* 2 (1988), 183–209.

Wittgenstein, L.: *Philosophical Investigations*, trans. G. E. M. Anscombe (Oxford: Blackwell, 1953).

Wittgenstein, L.: *The Blue and Brown Books* (Oxford: Blackwell, 1958).

SYDNEY SHOEMAKER

**self-presenting** This term refers to an alleged property of certain first-person mental states. One is supposed to be justified in believing that one is in such a state simply in virtue of being in such a state, independently of all other evidence. Alternatively, to be in such a state is to be aware that one is in such a state. These states include intentional states, such as thinking, believing and hoping, as well as perceptual states, such as being appeared to redly. Critics of the notion argue that one can be justified in ascribing properties to one's own mental states only if one is justified in believing oneself reliable in ascribing such properties generally.

See also EXPERIENCE; the GIVEN; SENSATION/COGNITION.

#### BIBLIOGRAPHY

Brentano, F.: *Psychology from an Empirical Standpoint*, trans. A. C. Rancurello, D. B. Terrell and L. L. McAlister (New York: Humanities Press, 1973), book 2, ch. 2.

Chisholm, R. M.: *Theory of Knowledge*, 3rd edn (Englewood Cliffs, NJ: Prentice-Hall, 1989), ch. 3.  
 Sellars, W.: "Empiricism and the Philosophy of Mind," in his *Science, Perception and Reality* (London: Routledge and Kegan Paul, 1963), 127–96.

ALAN H. GOLDMAN

**Sellars, Wilfrid (1912–89)** American philosopher. The epistemological view most closely associated with Wilfrid Sellars is surely his thoroughgoing critique of what he called "the Myth of the Given". The philosophical framework of givenness has historically taken on many guises, of which classical sense-datum theory is but one. But Sellars considered the very idea that empirical knowledge rests on a foundation at all, of *whatever* kind, to be a manifestation of the Myth of the Given, as was the assumption that one's "privileged access" to one's own mental states is a primitive feature of experience, logically and epistemologically prior to all intersubjective concepts pertaining to inner episodes. At the heart of Sellars' critique of "the entire framework of givenness" is his articulate recognition of the irreducibly normative character of epistemic discourse.

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.

(1956, p. 169)

The mere occurrence of "experiences", however *causally* systematic, consequently does not yet constitute perception in the epistemic sense that allows for a meaningful contrast of "veridical perception" and "misperception". Perception in this sense is always of something *as* something, and so requires more than mere exercise of differential response propensities. In so far as an instance of perceiving something *as* such-and-so is a candidate for epistemic appraisal, it necessarily encompasses the judgement that something *is* such-and-so, and *a fortiori* a classification of its content *under concepts*. It follows that the

senses *per se* grasp no facts, and so it becomes clear that

instead of coming to have a concept of something because we have noticed that sort of thing, to have the ability to notice a sort of thing is already to have the concept of that sort of thing, and cannot account for it.

(1956, p. 176)

Consonant with his rejection of the Myth of the Given, Sellars proceeded to interpret an individual's first-person epistemic authority with respect to aspects of his or her own experience as built on and presupposing an intersubjective status for sensory concepts. Correlatively, sensory consciousness cannot provide a form of knowledge of empirical facts that is "foundational" in being immediate (non-inferential), presupposing no knowledge of other matters of fact (particular or general), and constituting the ultimate court of appeals for all other factual claims. Although a person can indeed *directly know* an empirical fact in a sense which implies that he has not inferred what he justifiably believes from other propositions, Sellars insisted that the belief constituting such direct knowledge was not on that account somehow *self-justifying*, *self-warranting*, or *self-authenticating*. Rather

to say that someone directly knows that-*p* is to say that his right to the conviction that-*p* essentially involves the fact that the idea [belief] that-*p* occurred to the knower in a specific way. I shall call this kind of credibility "trans-level credibility", and [speak of] the inference scheme . . . to which it refers, as trans-level inference.

(Sellars, 1963a; p. 88)

So, for example, when someone sees there to be a red apple in front of him – a perceptual taking that Sellars models by a candid, spontaneous thinking-out-loud of the form: "Lo! Here is a red apple" – then,

given that he 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.

(1975a, pp. 341–2)

This reasoning does not have the original perceptual judgement as its conclusion, but is rather an inference from the character and context of the original non-inferential experience to the existence of a good reason for accepting it as veridical. The “trans-level” character of such a justificatory argument derives from the fact that its main premise asserts the occurrence (in a certain manner and context) of precisely the belief that is warranted by the reasoning as a whole, and it is this fact, too, that creates the impression that the belief in question is somehow *self-warranting*.

Sellars’ thoroughlygoingly holistic view of cognition and justification (*see* HOLISM) implies that the reasonableness of accepting even *first* principles is a matter of the availability of good arguments warranting their acceptance (Sellars, 1988). What is definitive of *first* principles, FP, is the unavailability of sound reasonings in which they are derived from still more basic premises:

(A1) .....  
 .....  
 Therefore, FP

The unavailability of sound reasonings of the form (A1), however, is entirely compatible with the existence of good “trans-level” arguments of the form:

(A2) .....  
 .....  
 Therefore, *it is reasonable to accept* FP,

in whose conclusion the principle FP is, in essence, mentioned.

Since accepting principles is something persons *do*, Sellars interpreted the conclusion of (A2) as the claim that a particular course of epistemic conduct could be supported by adequate reasons, i.e. that there is a sound *practical* argument whose conclusion expresses the intention to engage in such conduct:

(A3) I shall achieve desirable epistemic end E

Achieving E implies accepting principles of kind K.

The principle FP is of kind K

Therefore, I shall accept FP.

On Sellars’ view the forms of justificatory reasoning governing the acceptance of individual lawlike generalizations (both universal and statistical) and whole theoretical systems mobilize just such patterns of practical inference. Adopting a systematic theoretical framework is ultimately justified by an appeal to the epistemic end of “being able to give non-trivial explanatory accounts of established laws” (1964, p. 384); and adopting statistical nomologicals which project the observed frequency of a property in a class (including the case in which this frequency = 1) to unobserved finite samples of the class, by the epistemic end of

being able to draw inferences concerning the composition with respect to a given property Y of unexamined finite samples . . . of a kind, X, in a way which also provides an explanatory account of the composition with respect to Y of the total examined sample, K, of X.

(1964, p. 392)

Crucially, these epistemic ends are concerned with “the realizing of a logically necessary condition of being in the framework of explanation and prediction” at all (1964, p. 397). Since, on Sellars’ view, inductive reasoning does not need to be *vindicated*, i.e. shown to be truth-preserving, but rather is itself a form of *vindication*, i.e. (deductive) practical justificatory reasoning, the ends-in-view to which it appeals must be capable of being known to obtain or be realized. The end of being in possession of laws and principles that enable one to draw predictive inferences and produce explanatory accounts satisfies this practical constraint: such Reichenbachian ends as being in possession of (approximately) true limit-frequency statements, where such limits exist, do not.

*See also* the GIVEN.

## WRITINGS

- "Empiricism and the Philosophy of Mind," in Sellars 1963, pp. 127–96.
- Science, Perception and Reality* (London: Routledge and Kegan Paul, 1963).
- "Phenomenalism," in Sellars 1963, pp. 60–105; cited as (1963a).
- "Induction as Vindication," in Sellars 1975, pp. 367–416.
- Philosophical Perspectives* (Springfield, IL: Thomas, 1967); reprinted in two volumes, *Philosophical Perspectives: History of Philosophy and Philosophical Perspectives: Metaphysics and Epistemology* (Reseda, CA: Ridgeview, 1977).
- Essays in Philosophy and Its History* (Dordrecht: Reidel, 1975a).
- "The Structure of Knowledge," in *Action, Knowledge, and Reality*, ed. H.-N. Castaneda (Indianapolis, IN: Bobbs-Merrill, 1975b).
- The Carus Lectures for 1977–8, *The Monist* 64 (1981).
- "On Accepting First Principles," *Philosophical Perspectives* 2 (1989). Contains a complete bibliography of Sellars' published work through 1989.

JAY ROSENBERG

**sensation/cognition** There are various ways of classifying mental activities and states. One useful distinction is that between the propositional attitudes and everything else. A propositional attitude is one whose description takes a sentence as complement of the verb. Belief is a propositional attitude: one believes (truly or falsely as the case may be) that there are cookies in the jar. That there are cookies in the jar is the proposition expressed by the sentence following the verb. Knowing, judging, inferring, concluding and doubting are also propositional attitudes: one knows, judges, infers, concludes or doubts that a certain proposition (the one expressed by the sentential complement) is true.

Though the propositions are not always explicit, hope, fear, expectation, intention and a great many other terms are also (usually) taken to describe propositional attitudes: one hopes that (is afraid that, etc.) there are cookies in the jar. Wanting a cookie is, or can be construed as, a propositional attitude: wanting that one have (or eat or whatever) a cookie.

Intending to eat a cookie is intending that one will eat a cookie.

Propositional attitudes involve the possession and use of *concepts* and are, in this sense, representational. One must have some knowledge or understanding of what Xs are in order to *think, believe* or *hope* that something is X. In order to *want* a cookie, *intend* to eat one, or be *disappointed* that they are all gone, one must, in some way, know or understand what a cookie is. One must have this concept. (There is a sense in which one can want to eat a cookie without knowing what a cookie is – if, for example, one mistakenly thinks there are muffins in the jar and, as a result, wants to eat what is in the jar (= cookies). But this sense is hardly relevant; for in this sense one can want to eat the cookies in the jar without wanting to eat any cookies.) For this reason (and in this sense) the propositional attitudes are *cognitive*: they require (or presuppose) a level of understanding and knowledge, the kind of understanding and knowledge required to *possess* the concepts involved in occupying the propositional state.

Though there is sometimes disagreement about their proper analysis, non-propositional mental states, on the other hand, do not (at least not on the surface) take propositions as their object. Being in pain, being thirsty, smelling the flowers and feeling sad are introspectively prominent mental states that do not, like the propositional attitudes, require the application or use of concepts. One doesn't have to understand what pain or thirst is to experience pain or thirst. Assuming that pain and thirst are *conscious* phenomena, one must, of course, be conscious or aware of the pain or thirst to experience them, but awareness *of* must be carefully distinguished from awareness *that*. One can be aware of X – thirst or a toothache – without being aware that that (i.e. thirst or a toothache) is what, in fact, one is aware of. Awareness *that* (like belief that and knowledge that) is a propositional attitude; awareness *of* is not.

As the examples (pain, thirst, tickles, itches, hunger) are meant to suggest, the non-propositional states have a felt or experienced ("phenomenal") quality to them that is absent in the case of the propositional attitudes.

Aside from who it is we believe to be playing the tuba, believing that John is playing the tuba is much the same as believing that Joan is playing the tuba. These are different propositional states, different beliefs, yes, but they are distinguished entirely in terms of their propositional content – in terms of what they are beliefs about. Contrast this with the difference between *hearing* John play the tuba and *seeing* him play the tuba. Hearing John play the tuba and *seeing* John play the tuba differ, not just (as do beliefs) in what they are *of* or *about* (for these experiences are, in fact, of the same thing: John playing the tuba), but in their qualitative character: the one involves a visual, the other an auditory, experience. The difference between seeing John play the tuba and hearing John play the tuba, then, is a *sensory*, not a cognitive, difference.

Some mental states are a combination of sensory and cognitive elements. Fear and terror, sadness and anger, joy and depression, are ordinarily thought of in the way sensations are: not in terms of what proposition (if any) they are directed at, not in what (if anything) they represent, but (like visual and auditory experiences) in their intrinsic character, in how they feel to the person experiencing them. But when we describe a person as being afraid that, sad that, upset that (as opposed to merely thinking or knowing that) so-and-so happened, we typically mean to be describing the kind of sensory (feeling or emotional) quality accompanying the cognitive state. Being afraid that the dog is going to bite me is both to think (that he might bite me) – a cognitive state – and feel fear or apprehension (sensory) at the prospect.

The perceptual verbs exhibit this kind of mixture, this duality between the sensory and the cognitive. Verbs like “to hear”, “to see”, and “to feel” are (often) used to describe pro-positional (cognitive) states, but they describe these states in terms of the *way* (sensory) one comes to be in them. Seeing that there are two cookies left is coming to know that there are two cookies left *by seeing*. Feeling that there are two cookies left is coming to know this in a different way, by having tactile experiences (sensations).

On this model of the sensory-cognitive distinction (at least as it is realized in perceptual phenomena), sensations are a pre-conceptual, a pre-cognitive, vehicle of sensory information. The terms “sensation” and “sense-data” (or simply “experience”) were (and, in some circles, still are) used to describe this early phase of perceptual processing. It is currently more fashionable to speak of this sensory component in perception as the percept, the sensory information store, the icon (for these, see Neisser, 1967) or (from computer science) the 2 1/2-D sketch (Marr, 1982). But the result is generally the same: an acknowledgement of a stage in perceptual processing in which the incoming information is embodied in “raw” sensory (pre-categorical, pre-recognitional) form (Crowder and Morton, 1969). This early phase of the process is comparatively modular – relatively immune to, and insulated from, cognitive influences (Fodor, 1983). The emergence of a propositional (cognitive) state – seeing *that* an object is red – depends, then, on the earlier occurrence of a conscious, but none the less non-propositional condition: seeing (under the right conditions, of course) the red object. The sensory phase of this process constitutes the delivery of information (about the red object) in a particular form (visual); cognitive mechanisms are then responsible for extracting and using this information – for generating the belief (knowledge) that the object is red. (The existence of Blindsight suggests that this information can be delivered, perhaps in degraded form, at a non-conscious level.)

To speak of sensations *of* red objects, tubas, and so forth is to say that these sensations carry information about objects, information (in the case of vision) about an object’s colour, its shape, orientation and position; and (in the case of audition) information about acoustic qualities (pitch, timbre, volume). It is not to say that the sensations share the properties of the objects they are sensations of or that they *have* the properties they carry information about. Auditory sensations are not loud and visual sensations are not coloured. Sensations are bearers of (unconceptualized) information, and the bearer of the information that something is red need not itself *be* red. It need

not even be the sort of thing that could be red; it might be a certain pattern of neuronal events in the brain. None the less, the sensation, though not itself red, will (being the normal bearer of this information) typically produce in the subject who undergoes the experience a belief, or tendency to believe, that something red is being experienced. Hence the existence of hallucinations.

Just as there are theories of the mind (e.g. certain forms of functionalism) that would deny the existence of any state of mind whose essence was purely qualitative (i.e. did not consist of the state's extrinsic, causal, properties), there are theories of perception and knowledge – *cognitivist theories* – that deny a sensory component to ordinary sense perception (e.g. Gibson, 1950; Armstrong, 1961; Pitcher, 1971). The sensory dimension (the look, feel, smell, taste of things) is (if it isn't altogether denied) identified with some cognitive condition (knowledge or belief) of the experiencer. All seeing (not to mention hearing, smelling, and feeling) becomes a form of believing or knowing. As a result, organisms that cannot know (believe, etc.) cannot have experiences. Often, to avoid these strikingly counterintuitive results, cognitive theorists resort to unconscious, implicit, or otherwise unobtrusive (and, typically, undetectable) forms of believing (or knowing).

Aside, though, from introspective evidence (closing and opening one's eyes, if it changes beliefs at all, doesn't *just* change beliefs; it eliminates and restores a distinctive kind of conscious experience), there is a variety of empirical evidence for the existence of a stage in perceptual processing that is conscious without being cognitive (in any recognizable sense). For example, experiments with brief visual displays (Sperling, 1960; Averbach and Coriell, 1961) reveal that when subjects are exposed for very brief (e.g. 50 msec) intervals to information-rich stimuli (e.g. a  $3 \times 3$  array of numbers), there is persistence (at the conscious level) of what is called (by psychologists) an image or visual icon that embodies *more* information about the (now absent) stimulus than the subject can cognitively process or report on. Subjects can (depending on later prompts) exploit the

information in this persisting icon by reporting on *any* part of the absent array of numbers (they can, for instance, report the top three numbers; the middle three; or the bottom three). They cannot, however, identify *all* nine numbers. They report *seeing* all nine, and they can identify any one of the nine (actually, any three), but they cannot identify all nine. Knowledge and belief, recognition and identification – these cognitive states, though present (on any given trial) for any two or three numbers in the array, are absent for all nine numbers in the array. Yet, the image carries information (sufficient unto recognition and identification) about *all nine* numbers (how else account for subjects' ability to identify *any* number in the absent array?). Obviously, then, information is there, in the experience itself, *whether or not it is (or even can be) used*. As psychologists (e.g., Lindsay and Norman, 1972, p. 329) conclude, there is a limit on the information processing capabilities of the later (cognitive) mechanisms that is not shared by the sensory stages themselves.

See also CONCEPTS; EXPERIENCE; the GIVEN; LINGUISTIC UNDERSTANDING; SENSE-DATA.

#### BIBLIOGRAPHY

- Armstrong, D.: *Perception and the Physical World* (Harmondsworth: Penguin, 1961).
- Averbach, E. and Coriell, A. S.: "Short-term Storage of Information in Vision," *Information Theory: Proceedings of the Fourth London Symposium*, ed. C. Cherry (London: 1961).
- Dretske, F.: *Knowledge and the Flow of Information* (Cambridge, MA: MIT Press, 1981).
- Fodor, J.: *The Modularity of Mind* (Cambridge, MA: MIT Press, 1983).
- Gibson, J.: *The Perception of the Visual World* (Boston, MA: Houghton Mifflin, 1950)
- Lindsay, P. H. and Norman, D. A.: *Human Information Processing* (New York: Academic Press, 1972).
- Marr, D.: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (San Francisco, CA: Freeman, 1982).
- Neisser, U.: *Cognitive Psychology* (New York: Appleton-Century-Crofts, 1967).

Pitcher, G.: *A Theory of Perception* (Princeton, NJ: Princeton University Press, 1971).  
 Sperling, G.: "The Information Available in Brief Visual Presentations," *Psychological Monographs* 74: 11 (1960).

FRED DRETSKE

Moore, G. E.: "Sense-data," in his *Some Main Problems of Philosophy* (London: Allen and Unwin, 1953), 28–51.  
 Price, H. H.: *Perception* (New York: McBride, 1933).  
 Russell, B.: "On Our Knowledge of the External World," in his *Our Knowledge of the External World* (New York: Norton, 1929), 67–103.

ALAN H. GOLDMAN

**sense-data** Sense-data are the twentieth-century successors to the empiricists' concepts of ideas of sense (see EMPIRICISM; IDEA). The term "sense-data", introduced by Moore and Russell, refers to the immediate objects of perceptual awareness, such as colour patches and shapes, usually supposed distinct from surfaces of physical objects. Qualities of sense-data are supposed to be distinct from physical qualities because their perception is more relative to conditions, more certain, and more immediate, and because sense data are private and cannot appear other than they are. They are objects that change in our perceptual fields when conditions of perception change and physical objects remain constant.

Critics of the notion question whether, just because physical objects can appear other than they are, there must be private, mental objects that have all the qualities the physical objects appear to have. There are also problems regarding the individuation and duration of sense-data and their relations to physical surfaces of objects we perceive. Contemporary proponents counter that speaking only of how things appear cannot capture the full structure within perceptual experience captured by talk of apparent objects and their qualities.

See also ACT/OBJECT ANALYSIS; ARGUMENT FROM ILLUSION; CERTAINTY; EXPERIENCE; FOUNDATIONALISM; the GIVEN; INFALLIBILITY; REPRESENTATIVE REALISM.

#### BIBLIOGRAPHY

Barnes, W. H. F.: "The Myth of Sense Data," in *Perceiving, Sensing, and Knowing*, ed. R. Swartz (Garden City, NY: Anchor, 1965), 138–67.  
 Jackson, F.: *Perception* (Cambridge: Cambridge University Press, 1977).

**sensitivity and safety** Conditions subjects must arguably meet if their true beliefs are to count as instances of not epistemic luck but knowledge. Epistemologists often note that knowledge is, at root, non-lucky or non-accidental true belief. What we seek when we try to analyse knowledge is that condition (or set of conditions) which ensures that when our true belief is an instance of knowledge it is not subject in any substantive way to luck. On the face of it, a justification condition of some sort would seem to achieve this end. A *justified* true belief is preferable to a *mere* true belief because, it seems, genuine justification eliminates the possibility that one's true belief is true only by accident. Famously, however, the cases cited in Gettier's (1963) seminal article showed that justification, at least as it is usually understood, cannot achieve that end, and thereby opened up the discussion of a whole range of potential new ways of defining knowledge.

One of the most influential of these new proposals was the idea – prefigured in Dretske, 1970, but first clearly outlined in Nozick, 1981, ch. 3 – that we should understand knowledge primarily in terms of true belief that meets some sort of *sensitivity* condition. Nozick's insight was to recognise that the anti-luck intuition underlying much of our thinking about knowledge was in effect a *modal* intuition regarding how our beliefs in genuine cases of knowledge should be responsive to the facts not just in the *actual* world (i.e. be true), but also in a relevant range of *possible* worlds. Here, in essence, is how sensitivity is usually understood:

#### *Sensitivity*

An agent, *S*, has a *sensitive* belief in a true contingent proposition, *p*, if, and only if, in the



nearest possible worlds in which  $p$  is not true,  $S$  does not believe  $p$ .

Consider my belief right now that I am sitting at my desk. Provided that circumstances are normal, this belief is not only true but also sensitive. In the nearest possible worlds in which this belief is *not* true – those worlds in which I am standing next to my desk, for example – I do not believe that I am seated at my desk (but believe instead that I am standing).

According to Nozick, in addition to safeguarding true belief against luck, the sensitivity principle secures a second benefit: it enables us to resolve the problem of scepticism. The traditional challenge articulated by the sceptics presupposes what is known as the *closure principle*. A sensitivity-based theory of knowledge rejects scepticism on the grounds that the closure principle is false. This principle states, roughly, that if you know one proposition, and know that it entails a second proposition, then you know that second proposition. The sceptical import of this principle arises from the fact that we are, it seems, unable to know the denials of sceptical hypotheses, such as the hypothesis that one is presently a brain in a vat (BIV) who is being “fed” one’s experiences by supercomputers. However, ordinary propositions we all believe (such as “I have hands” or “I am presently seated at my desk”) entail these denials. (For example, if I have hands, I’m not a handless BIV.) Consequently, sceptics can reason that, since we don’t know that the sceptical hypotheses are false, and since they are entailed by the ordinary propositions we believe in ordinary life, we don’t know these ordinary propositions.

Nozick agrees with the sceptics on this much: We cannot know the denials of sceptical hypotheses. On Nozick’s theory, we cannot know them for the reason that the relevant beliefs do not meet the sensitivity condition. Consider my belief that I am not a BIV. In the nearest possible worlds in which that belief is false – worlds in which I *am* a BIV – I am deceived and thus believe that I’m not a BIV. In contrast, our beliefs in everyday propositions tend to be sensitive. Consider the belief that I have hands. In the closest worlds in which this belief is false (worlds in which

my hands have been amputated, for example), I do not believe that I have hands. Now, the conclusion that we lack knowledge of everyday propositions follows only if we accept closure. Consequently, Nozick argues that the principle of closure has to go, claiming that, unsurprisingly, a modal notion like knowledge should not support a principle of this sort.

The main problem for Nozick’s response to scepticism lies in the fact that the closure principle is highly intuitive, so intuitive in fact that many feel rejecting closure is too costly a price to pay for a resolution of the sceptical challenge. It may be, however, that one could retain closure *and* endorse a sensitivity-based theory of knowledge. In order to see how this might be so, we need to consider an essential modification – relativization to a method of belief formation – that Nozick makes to the principle of sensitivity.

Consider the following case, involving a grandmother who has a highly reliable ability to tell that her grandson is well just by looking at him. However, in the nearest possible worlds in which her grandson is unwell, the boy’s parents wouldn’t want grandmother to worry. They would keep him away from Grandmother and tell her that he is well regardless. So the grandmother’s belief, “He is well”, is not sensitive, for she would believe that her grandson is well even if he were not. On a sensitivity theory of knowledge, it follows that the grandmother’s belief is not a case of knowledge. But that does not look like a plausible outcome. After all, the grandmother’s belief is formed in a highly reliable fashion, for when the grandmother does get to look at her grandson she will not believe he is well when in fact he is not.

Relativizing sensitivity to a method solves this problem, since what is crucial to this example is that the method by which the belief is formed – i.e. having a good look at her grandson in good conditions of observation – is not the same method by which she forms her belief in the worlds in which her belief is false. In those worlds, the method by which she forms the belief is trusting the testimony of the boy’s parents. If, however, we keep the initial

method of *looking* at her grandson fixed, the grandmother's belief turns out to be sensitive, for in the nearest possible worlds in which she gets a good look at her grandson in good environmental conditions and he is unwell, she will recognise this and *not* believe that he is well. Hence, holding the initial method of belief formation fixed, we are able to retain the intuition that the grandmother has knowledge.

With the sensitivity principle modified to deal with this problem, it is no longer clear that, as Nozick thought, sensitivity entails closure failure because of our seeming inability to know the denials of sceptical hypotheses. After all, the method by which one forms, in the actual (non-sceptical) world, the belief that one is not a BIV is surely, at least in substantial part, the method of using and relying on one's normal perceptual faculties. However, in the nearest worlds in which one *is* a BIV, one cannot, using that method, be forming the belief that one is not a BIV because, *ex hypothesi*, those perceptual faculties are not available to one in these worlds (see Williams, 1991, ch. 8; Black, 2002). So arguably there cannot be a possible world in which one believes falsely that one is not a BIV and arrives at that belief via the same method as one does in the actual world. If it is true that there can't be such a world, then one's belief that one is not a BIV can, it seems, be sensitive and thus an instance of knowledge.

Nozick does not exercise this option of retaining closure because he understands the notion of a method "internalistically". He insists that any two methods which are "experientially the same, the same 'from the inside', will count as the same method" (1981, pp. 184–5). So construed, it could well be that one's method of forming the belief "I am not a BIV" is the same in both the actual world (in which one is not a BIV) and the nearest worlds in which one is a BIV, for arguably one's experiences in the two cases are the same. This is an oddly restrictive way of understanding methods, however, especially if, like Nozick, one is in other respects an epistemic externalist (see EXTERNALISM/INTERNALISM). Why should the nature of one's

experiences be a deciding factor here? Couldn't there be two different cognitive processes, properly classified as two methods, which have the same experiential upshot?

A different problem for the Nozickean sensitivity-based account of knowledge arises from counter-examples to the effect that the sensitivity condition prevents us from having lots of everyday knowledge. Sosa (1999), for instance, offers the following example. Suppose that I drop my rubbish down the rubbish chute in my high-rise condo. Do I know, moments later, that my rubbish is in the basement? According to the sensitivity-based account of knowledge, I do not because my belief "The rubbish is in the basement" is not sensitive. After all, were the bag to have somehow snagged on the way down the chute, I would still believe that the rubbish is now in the basement regardless. However, the belief that one's rubbish is now in the basement is, according to Sosa, a fairly paradigm instance of everyday knowledge. If sensitivity-based theories cannot account for cases like this, then the view is in serious trouble.

Sosa offers a different modal principle to replace sensitivity, which he refers to as the "principle of safety". Here, in essence, is how safety is usually understood:

### *Safety*

An agent, *S*, has a *safe* belief in a true contingent proposition *p* if, and only if, in most nearby possible worlds in which *S* believes *p*, *p* is true.

Like sensitivity, safety also aims to capture our intuition that knowledge is non-lucky true belief, the guiding interpretation of that intuition being that one's true beliefs should, if they are to count as knowledge, be safe in the sense that they could not easily have been false (see Pritchard, 2005, ch. 6).

Safety can account for our putative knowledge in the rubbish-chute case. There are of course possible worlds in which one's rubbish (having snagged on the way down) is *not* in the basement, yet one believes it is. That's why the belief in question is not sensitive. However, in most nearby possible worlds in which one believes one's rubbish is

in the basement, one's belief is true. After all, one's rubbish does not easily snag on its way down the chute. The belief in question is, therefore, safe.

Safety can also be put into service to explain why we can know the denials of sceptical hypotheses. Provided the worlds in which sceptical hypotheses are true are indeed far-off, there are no nearby possible worlds in which one believes "I'm not a BIV" when in fact one is a BIV. Consequently, one's belief that one is not a BIV is safe. As pointed out above, sensitivity-based theories of knowledge arguably yield the same result. Such theories, however, were not designed with the intention to secure this result. In contrast, proponents of safety-based theories of knowledge, such as Sosa, 1999, and Pritchard, 2005, have made the fact that safety can account for our knowledge of the denials of sceptical hypotheses an explicit part of the motivation for the principle.

As with sensitivity, some sort of relativisation to methods will be required to enable safety to deal with potential counter-examples. The main attack on safety, however, is not based on counter-examples but rather arises from the complaint that there is no stable construal of the principle which can accommodate both everyday cases of knowledge and the kind of knowledge that one has in the *lottery* case (see Greco, 2002). Suppose you own a single lottery ticket for a fair lottery with long odds, where the draw for the lottery has taken place but you have yet to hear the result. Reflecting on the long odds involved, you form the belief that yours is a losing lottery ticket. Intuitively, this belief is not an instance of knowledge, for one cannot acquire such knowledge on a purely statistical basis. Two points about this case are noteworthy: (i) You fail to have knowledge although you have impeccable evidence in favour of your belief. (ii) Upon consulting a reliable newspaper, you can come to know that your ticket is a loser even though the chance of error is hardly less in this second case.

The sensitivity theorist can easily explain the data with which the lottery case presents us. Distinguish between two methods: *reflecting*

*on the odds* (M1) versus *looking up the numbers in a reliable newspaper* (M2). Your belief that your ticket is a loser is sensitive if you use M2 but is not sensitive if you use M1. Consider the nearest possible worlds in which your ticket bears the winning number. Upon using M1, you will in these worlds arrive at the false belief "My ticket is a loser". So your belief is not sensitive relative to M1. However, upon using M2, you will in these worlds arrive at the true belief that you won the lottery. Hence your belief is sensitive relative to M2.

The safety theorist, however, has trouble explaining our lack of knowledge in the lottery case. Given the probabilities involved in the envisaged lottery, in most nearby possible worlds in which one believes that one has lost (using the same method as in the actual world) one's belief will be true. The belief that one's ticket is a loser is therefore safe, and hence an instance of knowledge. Many epistemologists would say that this outcome is contrary to intuition.

To avoid this objection, the safety theorist could propose to construe safety so that it requires that one's belief matches the truth not just in most nearby possible worlds, but in all (or at least nearly all) of them (see e.g. Pritchard, 2005, ch. 6). This would ensure that one's belief that one has lost the lottery fails the safety test and thus is no longer an instance of knowledge. This move is not unproblematic, however, for it makes it hard to account for one's knowledge in the rubbish-chute case. After all, surely there are *some* nearby possible worlds in which the rubbish chute snags on the way down and yet one believes (on the same basis as in the actual world) that the rubbish is in the basement. The challenge for the safety theorist is thus to articulate a notion of safety that meets two desiderata: it should be robust enough to account for the lack of knowledge in the lottery case, while at the same time liberal enough to allow for having knowledge in cases like the rubbish chute example.

See also EPISTEMIC LUCK; EXTERNALISM/INTERNALISM; RELIABILISM; SCEPTICISM.

## BIBLIOGRAPHY

- Black, T.: "A Moorean Response to Brain-in-a-vat Scepticism," *Australasian Journal of Philosophy* 80 (2002), 148–63.
- Dretske, F.: "Epistemic Operators," *The Journal of Philosophy* 67 (1970), 1007–23.
- Gettier, E.: "Is Justified True Belief Knowledge?," *Analysis* 23 (1963), 121–3.
- Greco, J.: "Virtue and Luck, Epistemic and Otherwise," *Metaphilosophy* 34 (2002), 353–66.
- Nozick, R.: *Philosophical Explanations* (Oxford: Oxford University Press, 1981).
- Pritchard, D. H.: *Epistemic Luck* (Oxford: Oxford University Press, 2005).
- Sosa, E.: "How to Defeat Opposition to Moore," *Philosophical Perspectives* 13 (1999), 141–54.
- Williams, M.: *Unnatural Doubts: Epistemological Realism and the Basis of Scepticism* (Oxford: Blackwell, 1991).

DUNCAN PRITCHARD

**Sextus Empiricus** A Greek sceptic of the Pyrrhonian school, and a practicing physician, who flourished probably during the late second century AD. His exact dates are controversial and the details of his life virtually unknown to us, yet he is the most important source of our knowledge of ancient Greek sceptical philosophies. The works that have come down to us are the *Outlines of Pyrrhonism* in three books, which provides us with Sextus' own positive account of Pyrrhonian scepticism, and a larger work in eleven books, commonly referred to collectively as *Against the Mathematicians*. The latter contains much of the same material as found in the *Outlines*, but it also provides additional sceptical arguments against dogmatic philosophers as well as valuable information about the major philosophical schools of the Hellenistic period. It appears that little of the philosophical material in Sextus' writings is original with him. We know, for example, that he drew freely on the thought of earlier Pyrrhonian sceptics, especially Aenesidemus in the first century BC (see PYRRHONISM).

Sextus describes scepticism as a "philosophy" and a "way of life", identifying both a theoretical and an anti-theoretical practical component in Pyrrhonian scepticism. The appearance of

paradox is genuine, for the sceptic engages in theoretical reasoning in order finally to do away with it. The sceptic is an "inquirer" into truth, but unlike other philosophers whom Sextus classifies as either dogmatists or Academics, he neither claims to have discovered the truth nor does he say that it cannot be discovered (*Outlines*, I 3–4). The Pyrrhonian sceptic, *qua* philosopher, just continues to inquire. Sextus himself goes beyond the role of inquirer in presenting a positive and highly sophisticated theoretical account of scepticism as a way of life in which the practical dimension of Pyrrhonian scepticism remains one of its most important and distinctive features (Stough, 1984). Sceptical argumentation always has a practical aim. Sextus likens the sceptic's arguments to a drug designed to cure a peculiarly philosophical ailment characteristic of dogmatists, who boldly and uncritically put forth theories about how things really are (*Outlines*, III 280–1). The sceptic's strategy is intended to cure this pathological inclination and indeed, when properly understood, his own arguments refute themselves along with all the rest. Sextus' writings contain elaborate and extensive arguments which were deployed against all dogmatic philosophers, but most notably Stoics who strenuously defended a criterion of truth as the foundation of their philosophical system (see STOIC EPISTEMOLOGY).

The sceptic himself, Sextus says, initially set out to determine the truth or falsity of his impressions of things in an effort to attain ataraxia, the tranquil and untroubled state of mind put forth as the end (*telos*) of scepticism (see ATARAXIA). But instead he found himself confronted with contradictory appearances and arguments of equal weight and credibility (see ISOSTHENEIA). Unable to decide between them, he adopted a neutral attitude, suspending judgement about their truth or falsity (*epoche*), and found "as if by chance" that *ataraxia* followed "as a shadow follows its object" (*Outlines*, I 26). Sextus' narrative provides us with a model for the sceptical method. Scepticism is defined as the ability to produce oppositions between appearances and judgements "in any manner whatsoever" as a means of facilitating the noncommittal

attitude of neither affirming nor denying anything (*Outlines* I, 8). Sceptics make a practice of balancing the opposing claims of the dogmatists against each other. In doing so, they find “no more” reason to prefer one position than another. The arguments in support of competing theories are equally strong, hence equally persuasive. Pyrrhonists therefore suspend judgement, taking no position at all about what is true or false.

Characteristic of the sceptical method are the “modes” (*tropoi*) of Pyrrhonian scepticism preserved in Sextus’ writings. The sceptical modes are patterns of argument designed to induce suspense of judgement. The best-known of several different groups of modes are the Ten Modes of *Epoche* attributed to Aenesidemus, which are elaborated at length in the first book of Sextus’ *Outlines of Pyrrhonism*. Each mode makes distinctive use of the fact that things “appear” different to us in different situations. How something looks (tastes, smells, feels, and the like) is determined by ten factors delineated by the modes, such as conditions affecting subject and object and the circumstances in which the object appears. These variations are invoked to produce “oppositions”, which are expressed in propositions typically assigning incompatible appearances (properties) to an object. The sceptic then claims *isostheneia* of appearances and ends by suspending judgement about how things really are. Schematically the argument looks like this (modification of Annas and Barnes, 1985):

1.  $x$  appears F in situation S1.
2.  $x$  appears F' in situation S2.
3. We have no criterion (or proof) independent of S1 and S2 to judge between F and F'.
4. We can neither affirm nor deny that  $x$  is really F or F'.

F and F' represent oppositions of appearances, predicates which in the sceptic's view cannot be jointly true of a subject. In one mode S1 and S2 vary in the positions occupied by the perceiver. The same boat looks small and stationary when viewed from a distance, but large and in motion from close at hand. Another mode notes the effects of the various

circumstances or conditions a subject may be in when an object is perceived. Air that feels cold to an older person feels mild to someone in his prime. Further modes appeal to differences among the senses, differences in the quantity and composition of an object, the effect of admixtures, the effect of relativity, and numerous other factors influencing the way things appear to us. Since the oppositions are equally balanced, the appearances are equally credible. The sceptic therefore withholds assent and neither affirms nor denies that  $x$  is really F or F'.

As a consequence of suspending judgement about how things really are (all such matters being “non-evident” (*adela*) according to Sextus), the sceptic holds no true or false beliefs. Sextus insists, however, in response to criticism, that the sceptic is not reduced to inactivity, nor is he forced into inconsistency, by his neutrality in regard to belief (Frede, 1979; Burnyeat, 1980). The Pyrrhonian sceptic assents to his impressions of things and follows appearances as a criterion for acting in everyday life. Although he holds no beliefs, and makes no assertions, about what is or is not so, he never questions the fact that things appear this way or that. His speech, which merely records how things appear to him, has expressive and regulative functions pertaining solely to action (Stough, 1984). Sextus maintains that the sceptic can follow appearances as a practical criterion for everyday affairs without committing himself to any beliefs or statements about what really is the case.

See also MONTAIGNE; PYRRHONISM; SCEPTICISM, MODERN; STOIC EPISTEMOLOGY.

#### WRITINGS

*Opera*, 3 vols (Leipzig: Teubner, 1912–54). Vols 1, 2, ed. H. Mutschmann (1912–14); vol. 3, ed. J. Mau (indices, K. Janacek) (1954).

*Works*, 4 vols, trans. R. G. Bury, Loeb Classical Library (London: William Heinemann, 1933–49): vol. 1, *Outlines of Pyrrhonism* (1933); vol. 2, *Against the Mathematicians* (1935); vol. 3, *Against the Physicists; Against the Ethicists* (1936); vol. 4, *Against the Professors* (1949).



## BIBLIOGRAPHY

- Annas, J. and Barnes, J.: *The Modes of Scepticism* (Cambridge: Cambridge University Press, 1985).
- Barnes, J.: "The Beliefs of a Pyrrhonist," *Proceedings of the Cambridge Philological Society* 29 (1982), 1–29, and *Elenchos* 4 (1983), 5–43.
- Brochard, V.: *Les Sceptiques grecs* (Paris: F. Alcan, 1887); 2nd edn reprinted (Paris: Librairie Philosophique J. Vrin, 1959).
- Burnyeat, M.: "Can the Sceptic Live His Scepticism?," in *Doubt and Dogmatism*, ed. M. Schofield, M. Burnyeat and J. Barnes (Oxford: Clarendon Press, 1980), 20–53.
- Frede, M.: "Des sceptikers Meinungen," *Neue Hefte für Philosophie* 15/16 (1979), 102–29.
- Long, A.: "Sextus Empiricus on the Criterion of Truth," *Bulletin of the Institute of Classical Studies* 25 (1978), 35–49.
- Stough, C.: *Greek Scepticism* (Berkeley, CA: University of California Press, 1969).
- Stough, C.: "Sextus Empiricus on Non-assertion," *Phronesis* 29 (1984), 137–64.
- Stough, C.: "Knowledge and Belief," *Oxford Studies in Ancient Philosophy* 5 (1987), 217–34.
- Striker, G.: "The Ten Tropes of Aenesidemus," in *The Sceptical Tradition*, ed. M. Burnyeat (Berkeley, CA: University of California Press, 1983), 95–115.

CHARLOTTE STOUGH

**simplicity** Philosophers and scientists have often held that the simplicity or parsimony of a theory is one reason, all else being equal, to view it as true. This goes beyond the unproblematic idea that simpler theories are easier to work with and have greater aesthetic appeal.

One theory is more parsimonious than another when it postulates fewer entities, processes, changes or explanatory principles; the simplicity of a theory depends on more or less the same considerations, though it is not obvious that parsimony and simplicity come to the same thing. It is plausible to demand clarification of what makes one theory simpler or more parsimonious than another before the justification of these methodological maxims can be addressed.

If we set this descriptive problem to one side, the major normative problem is as follows: What reason is there to think that simplicity is a sign of truth? Why should we accept a

simpler theory instead of its more complex rivals? Newton and Leibniz thought that the answer was to be found in a substantive fact about nature. In *Principia*, Newton laid down as his first Rule of Reasoning in Philosophy that "nature does nothing in vain . . . for Nature is pleased with simplicity and affects not the pomp of superfluous causes". Leibniz hypothesized that the actual world obeys simple laws because God's taste for simplicity influenced his decision about which world to actualize.

Epistemology since Hume and Kant has drawn back from this theological underpinning. Indeed, the very idea that nature is simple (or uniform) has come in for critique. The view has taken hold that a preference for simple and parsimonious hypotheses is *purely methodological*; it is constitutive of the attitude we call "scientific" and makes no substantive assumption about the way the world is.

A variety of otherwise diverse twentieth-century philosophers of science have attempted, in different ways, to flesh out this position. Two examples must suffice here; see Hesse (1969) for summaries of other proposals. Popper (1959) (*see* POPPER) holds that scientists should prefer highly falsifiable (improbable) theories; he tries to show that simpler theories are more falsifiable. Quine (1966) (*see* QUINE), in contrast, sees a virtue in theories that are highly probable; he argues for a general connection between simplicity and high probability.

Both these proposals are *global*. They attempt to explain why simplicity should be part of the scientific method in a way that spans all scientific subject matters. No assumption about the details of any particular scientific problem serves as a premiss in Popper's or Quine's arguments.

Newton and Leibniz thought that the justification of parsimony and simplicity flows from the hand of God; Popper and Quine try to justify these methodological maxims without assuming anything substantive about the way the world is. In spite of these differences in approach, they have something in common. They assume that all uses of parsimony and simplicity in the separate sciences can be encompassed in a single justifying argument.

Recent developments in confirmation theory suggest that this assumption should be scrutinized. Good (1983) and Rosenkrantz (1977) have emphasized the role of auxiliary assumptions in mediating the connection between hypotheses and observations. Whether an hypothesis is well supported by some observations, or whether one hypothesis is better supported than another by those observations, crucially depends on empirical background assumptions about the inference problem at hand. The same view applies to the idea of prior probability (or prior plausibility) (see PROBABILITY, THEORIES OF). If one hypothesis is preferred over another even though they are equally supported by current observations, this must be due to an empirical background assumption.

Principles of parsimony and simplicity mediate the epistemic connection between hypotheses and observations. Perhaps these principles are able to do this because they are surrogates for an empirical background theory. It isn't that there is one background theory presupposed by every appeal to parsimony; this has the quantifier order backwards. Rather, the suggestion is that each parsimony argument is justified only to the degree that it reflects an empirical background theory about the subject matter at hand. Once this theory is brought out into the open, the principle of parsimony is entirely dispensable (Sober, 1988).

This *local* approach to the principles of parsimony and simplicity resurrects the idea that they make sense only if the world is one way rather than another. It rejects the idea that these maxims are purely methodological. How defensible this point of view is will depend on detailed case studies of scientific hypothesis evaluation and on further developments in the theory of scientific inference.

See also NATURAL SCIENCE.

#### BIBLIOGRAPHY

Good, I.: *Good Thinking: The Foundations of Probability and Its Applications* (Minneapolis, MN: University of Minnesota Press, 1983).

Hesse, M.: "Simplicity," in *The Encyclopedia of Philosophy*, ed. P. Edwards (New York: Macmillan, 1969), 445–8.

Popper, K. R.: *The Logic of Scientific Discovery* (London: Hutchinson, 1959).

Quine, W. V.: "Simple Theories of a Complex World," in his *The Ways of Paradox and Other Essays* (New York: Random House, 1966), 242–6.

Rosenkrantz, R.: *Inference Method, and Decision* (Dordrecht: Reidel, 1977).

Sober, E.: *Reconstructing the Past: Parsimony, Evolution and Inference* (Cambridge, MA: MIT Press, 1988).

ELLIOTT SOBER

**social sciences, epistemology of** Do the social sciences require any special treatment or can the results of general epistemology, which have in modern times usually been formulated with an eye to the natural sciences, be applied directly to them? This question has exercised philosophers and social scientists more or less from the beginnings of self-consciously "scientific" investigations of social phenomena. Among the founders of modern social science, Mill (see MILL), Weber and Durkheim, to name a few, all commented on it. In the contemporary context, this issue is often addressed under the heading *naturalism*, i.e. the doctrine that the methods of the natural sciences can be used in the study of society. Self-proclaimed anti-naturalists nowadays abound, but, ironically, frequently situate themselves with respect to characterisations of the natural scientific enterprise that are themselves now dated. For, after Kuhn (1970), it will certainly no longer suffice to place oneself in opposition to naturalism, to rail against the restrictive tenets of verificationism (see VERIFICATIONISM), a position which no longer commands great respect even in relation to epistemology of the natural sciences.

There are at least three *prima facie* reasons for wondering whether naturalism can be sustained or whether it is necessary, instead, to articulate for the social sciences an epistemology which doesn't simply ape the preoccupations and conclusions of "mainstream" natural-scientifically oriented epistemology.

First, there is the *reflexivity* of the social sciences in relation to the objects of their scrutiny. By this I refer, in particular, to the fact that human beings, the objects of study, come to take up points of view towards their own activities which are influenced, however indirectly, by the results of social scientific investigations, and which therefore in some sense retrospectively invalidate these investigations (which made no provision for the “appropriation” of their results by their subjects). (Alfred Schutz (1971, p. 495) was already aware of a cognate phenomenon – the “pre-interpreted” character of “social reality”, which implies, he thought, that social scientists are, at least in the first instance, dependent for their understanding of social reality on the understandings of their subjects.) Although this phenomenon was perhaps most poignantly identified in relation to so-called “reflexive predictions” (self-fulfilling or self-defeating “prophecies”), it is of much more general significance. When social actors acquire familiarity with social scientific ideas, their behaviour and attitudes are no longer “naïve” with respect to them and the ideas themselves no longer apply as straightforwardly. And this, it is sometimes said, is a distinguishing mark of the social sciences; there is nothing in the domain of natural scientific concern that corresponds to the mutual interplay that seems unavoidable between social scientists and the subjects of their research. (The influence of the observer on what is observed of quantum mechanical phenomena is, surely, only vaguely analogous.)

Second, there is the *complexity* of social phenomena, or perhaps, more correctly, their relative imperviousness to controlled (and thought-) experimental manipulation in the “Galilean style” which involves abstraction, idealization and the identification of underlying mechanisms – *prima facie* complexity then being interpreted as a product of the interaction of these, themselves simple mechanisms. Mill had already identified the impediments, for the development of a genuinely predictive social science, thrown up by the complexity of social phenomena, and the difficulties seem to remain. To some, such as F. A. Hayek (1973–8, vol. 2, ch. 10), this

feature of social scientific research suggests that students of social phenomena confine their interest to the elucidation of patterns and other structures of order, abandoning as futile the aspiration to develop a genuinely predictive science of human behaviour and social interaction. Since it is widely believed that such complexity as is manifested in relation to “purely natural” phenomena is susceptible to manipulation using broadly “Galilean” techniques and that the products of such manipulation *are* genuinely predictive theories, this feature is also held to mark a difference between natural and social scientific domains.

Third, there is the (essential) *contestability* of many of the theoretical concepts of the social sciences (see Gallie, 1955–6). By this I refer, of course, to the fact that we are required to make what could reasonably be called “value judgements” in order to apply or refuse the application of many of the most important social scientific concepts. This is so, according to Gallie and his followers, because the concepts themselves are at least partially, and ineradicably, *evaluative* rather than strictly “descriptive” (see FACT/VALUE), but, perhaps more importantly, because the criteria for their application are many and because judgements, which themselves have an evaluative dimension, need to be made in order to determine whether these criteria have been adequately satisfied in any particular case. (The evaluative character of many social scientific concepts may be a legacy of the dependence, stressed by Schutz, of social scientific concepts on the concepts of ordinary social agents, which are certainly themselves evaluative. It seems to mark a limit of neo-Weberian aspirations for a “value-free” social science and to explain the strenuous, but apparently pointless attempts to “operationalize” the concepts in terms of which theories are framed – or otherwise purge them of connotations unacceptable to the positivist tradition (see LOGICAL POSITIVISM).) Since such fundamental social theoretical concepts as power and rationality (see RATIONALITY) are, *prima facie*, essentially contestable in this sense, while all the key concepts of the natural sciences appear not to be, this feature too

is held to mark a distinction between natural and social scientific investigations, and therefore, along with reflexivity and complexity, to herald the need for a distinctive epistemology for the social sciences.

Many are those who have been persuaded by these and similar considerations to accept that there is a difference in kind between the techniques appropriate to the social sciences and those appropriate to the natural sciences and, what is more, between the epistemological orientations of these two great branches of learning. If the fundamental orientation of the natural sciences is a *technical* one, aiming at prediction and control of natural phenomena, then, according to Habermas (see HABERMAS) among many others, that of the social sciences is or ought to be either *practical*, aiming at understanding or *critical*, aiming at liberation. In either case, we encounter the idea that the social sciences do not or ought not to aim at the development of abstract general theories which provide a basis for fine-grained predictions of concrete social phenomena, but that they ought, instead, to aim at understanding or to interpret.

One way of capturing what's at issue on this account is by reference to a distinction between "thin" and "thick" descriptions of an agent's action or of some social-structure feature (cf. Geertz, in Gibbons, 1987). We give a thin description of some action, for instance, when we describe it in a more or less "behaviouristic" vocabulary of bodily movements. We give a thick description of this same action when we describe it in the vocabulary of native members of the community in which it occurs, i.e. according to their understanding of its significance and character. So an action is thinly described as sticking pieces of flesh in the fire and thickly described as cooking or sacrificing to the gods, as the case may be. Clearly, thick descriptions are interpretations, "from the native's point of view" according to Geertz, and the provision of them represents an important contribution to our understanding of human affairs. Furthermore, the provision of such descriptions is unthreatened by the phenomena of reflexivity, complexity, and contestability. Of course, we may have to provide different

interpretations when interpretees become aware of the interpretations we have already provided – reflexivity is catered for, not eliminated. But we are not incapacitated, in the provision of interpretations, by either complexity or contestability: in the latter case, the contestability of native concepts is simply "reflected" in the contestability of our thick descriptions or interpretations of native behaviour and structures.

(For earlier generations, the idea of *verstehen* embodied this understanding of the distinctive interpretive orientation of the social sciences. For some of our contemporaries this understanding is embodied in the project of hermeneutics especially associated, perhaps, with the work of Hans-Georg Gadamer (see HERMENEUTICS; GADAMER); for others in the project of genealogy especially associated perhaps with the work of Michel Foucault (see FOUCAULT); see Rabinow, 1984.)

The notion of *interpretation*, then, is central to anti-naturalist conceptions of social science. But, as has often been noted, it contains an ambiguity of crucial political importance. There is a hermeneutics of *recovery*, on the one hand, and a hermeneutics of *suspicion*, on the other hand. In the one case, we aim to recover the socially constitutive self-understandings of ordinary social agents; in the other (Gibbons, 1987, p. 4) we aim to penetrate ideology (see IDEOLOGY), self-delusion and other mystificatory aspects of agents' self-understandings to "uncover the real or true meaning of social and political practice". In the one case, we aim to reconstitute selfconsciously and explicitly what agents have themselves already constituted as a basis for their social lives. This aim is especially important in relation to ways of life initially quite "alien" to us and so is commonly, and rightly, associated with the anthropological enterprise. In the other case, our aim is informed by the fundamental insights of Nietzsche, Marx and Freud (see NIETZSCHE; MARXISM; FREUD). They have taught us that social actors do not always adequately comprehend the character of their motives or of the institutional framework in which they perform, and that any genuine understanding of these matters will require treating with suspicion agents' self-understandings, which

are, on this account, themselves to be understood rather than, as on the more “naïve” account, treated as sources of understanding. (Also of some relevance, ironically, are the insights of “mainstream” politico-economic theorists from the time of Adam Smith and including Popper (*see* POPPER) who have identified “unintended consequences” as an additional source of self-misunderstanding, or, at least, of the inadequacy, for an understanding of social life, of an understanding of agents’ self-understanding of their actions (*see* e.g. Hayek, 1973–8, vol. 1).)

How firm is this contrast between the hermeneutics of recovery and the hermeneutics of suspicion? For the hermeneutics of recovery we need to assume a broad basis of agreement in beliefs, desires and other propositional attitudes between social scientists and the subjects of their research. This assumption is embodied in what Donald Davidson (1984) calls the principle of charity (*see* DAVIDSON; PRINCIPLE OF CHARITY); and, unless we make and can legitimately make this assumption about the subjects of our research, their activities must remain mysterious to us. Of course, it is not implied, simply because we assume a broad basis of agreement, that we will never be able to discover disagreement, should it exist. Although we may be reluctant to attribute “odd” attitudes to the subjects of our interpretive investigations (*see* Quine, 1960) we may sometimes be right to do so – e.g. where we have a good explanation of the “oddity” of these attitudes.

There are nevertheless many obscurities and much controversy about criteria of interpretive adequacy precisely in relation to this issue. Many interpreters have been notably reluctant to “take literally” the religious or “philosophical” attitudes of interpretees (at least where these diverge too much from “mainstream” attitudes) and have claimed, in defence of this reluctance, that good reasons can be offered to treating the utterances apparently expressive of such attitudes as metaphorical, or more broadly, symbolic in character (*see* Skorupski, 1976). But these kinds of cases, as difficult as they are to resolve, point, more deeply, to the fundamentally holistic character (*see* HOLISM) of the

interpretive project and to the potential for an apparently irreducible multiplicity of interpretations – a manifestation, in this domain, of the indeterminacy of translation (*see* INDETERMINACY OF TRANSLATION).

As Quine has emphasized (*see* QUINE), how a particular object of interpretation is to be understood is determined, where it can be, only against a very broad background of considerations. It is possible, by making adjustments at relevant loci in a complex system of attitudes, to interpret any given object in a number of different ways, each of which is “correct” relative to the (suitably adjusted) system in which it is incorporated (*see* ONTOLOGICAL RELATIVITY).

Other important contributions to our understanding of the hermeneutics of recovery are John Rawls’ idea (1973) of “reflective equilibrium”, which seems to capture Quinean ideas about adjusting some parts of a larger system in order to accommodate other potential elements, and Ronald Dworkin’s idea (1986) of the “interpretive attitude”, which draws special attention to the requirement that we strive, in interpretation, “to make an object the best it can be, as an instance of some assumed enterprise”.

All these ideas seem to suggest that a kind of relativism (*see* RELATIVISM) is not easily to be avoided in the realm of interpretation. They also suggest that the distinction between the hermeneutics of recovery and the hermeneutics of suspicion is somewhat overdrawn. If some attitudes of interpretees may have to be treated “symbolically” interpretation sometimes involves *correction* of agents’ naïve self-understandings, with the possibility, at least, of concomitant liberation from them. And if there are always going to be several different and incompatible ways of making “best sense” of one’s interpretees, this undermines the thought that in the hermeneutics of suspicion we look down upon our interpreters from a position whose nature gives our results a claim to objective validity.

Still, what holds the distinction in place is the idea that while a hermeneutics of recovery rests on the principle of charity and the assumption that many attitudes are shared



by interpreter and interpretee, a hermeneutics of suspicion seems to presuppose that the interpreter occupies a vantage-point which is privileged in relation to that of the interpretee (in Marxism, this is the point of view of the vanguard party, or perhaps the proletariat; in Freudianism, it is the point of view of the analyst) – or perhaps that both take up a point of view which provides a basis for some kind of superior insight. In the work of Habermas (1981) (see HABERMAS) this appears as the notion of communicative action which may involve the testing, in discourse, of the validity claims which are implicit in the various acts of interpretation. Claims which can pass such tests provide points of leverage against our pre-reflective beliefs. (There is resonance here with the idea, in moral-political theory, of an Archimedean point – see Rawls, 1973, p. 41.)

The idea that the interpreter is in a privileged position, and the idea that there is a privileged position which lies at some remove from those of interpreter and interpretee, can be contrasted with the concept of genealogy associated with the work of Michel Foucault (see Rabinow, 1984). This seems, on the one hand, to deny that there is *any* privileged perspective from which hidden realities can be perceived, while, on the other hand, claiming to provide an account of those institutional, ideological and historical contingencies (masquerading as necessities) which form our ideas and practices.

If the idea of interpretation is central to the anti-naturalistic conception of social scientific activity, it is by no means clear, as many anti-naturalists have tended to assume, that this sets their activities apart from those of natural scientist engaged in the study of non-human phenomena. Certainly, many of the themes of anti-naturalists and interpretivists are to be found in (and, perhaps ironically, are derived from) the work of Thomas Kuhn and his predecessors and successors (see Bernstein, 1983). Just as social scientists seek to interpret the attitudes, actions and works of human beings, so too, it might be said, do natural scientists search for an interpretation of natural phenomena, a way of making sense of them to ourselves. And they too are

likely to find, as Quine has emphasized, that their efforts are only holistically constrained, rather than tied to some incorrigible fundament, as previously sought in both rationalism and empiricism alike (see RATIONALISM; EMPIRICISM). And so, too, are they likely to find, as Richard Rorty (1979) has recently much emphasized (see RORTY), that there is no absolutely secure vantage-point which, could we take it up, would provide us with what Hilary Putnam has called a “God’s-eye view” of the natural world (see PUTNAM). There is, then, a kind of double irony about the epistemology of the social sciences. In articulating a distinctive account of the activities and aspirations of social scientists, philosophers have contributed to a re-orientation of thinking about general epistemology which somewhat effaces the distinction with which they started – between a naturalistically oriented epistemology of the “hard” sciences and an interpretivist epistemology of the “soft” sciences.

See also HISTORICAL KNOWLEDGE; IDEOLOGY; NATURAL SCIENCE.

#### BIBLIOGRAPHY

- Bernstein, R. J.: *Beyond Objectivism and Relativism* (Oxford: Blackwell, 1983).  
 Davidson, D.: *Inquiries into Truth and Interpretation* (Oxford: Oxford University Press, 1984).  
 Dworkin, R.: *Law's Empire* (London: Fontana, 1986).  
 Gallie, W. B.: “Essentially Contested Concepts,” *Proceedings of the Aristotelian Society* 56 (1955–6), 167–98.  
 Gibbons, M. T., ed.: *Interpreting Politics* (Oxford: Blackwell, 1987).  
 Habermas, J.: *The Theory of Communicative Action*, trans. T. McCarthy (Boston, MA: Beacon Press, 1981).  
 Hayek, F. A.: *Law, Legislation and Liberty*, 3 vols (Chicago, IL: University of Chicago Press, 1973–8).  
 Kuhn, T. S.: *The Structure of Scientific Revolutions*, rev. edn (Chicago, IL: University of Chicago Press, 1970).  
 Quine, W. V.: *Word and Object* (Cambridge, MA: MIT Press, 1960).  
 Rabinow, P., ed.: *The Foucault Reader* (Harmondsworth: Penguin, 1984).

- Rawls, J.: *A Theory of Justice* (Oxford: Oxford University Press, 1973).
- Rorty, R.: *Philosophy and the Mirror of Nature* (Princeton, NJ: Princeton University Press, 1979).
- Schutz, A.: "Concept and Theory Formation in the Social Sciences," reprinted in *Sociological Perspectives*, ed. K. Thompson and J. Tunstall (Harmondsworth: Penguin, 1971).
- Skorupski, J.: *Symbol and Theory* (Cambridge: Cambridge University Press, 1976).

FRED D'AGOSTINO

**sociology of knowledge** The aim of the sociology of knowledge is to locate whatever body of belief a group accepts as a true account of reality, and then try to illuminate it by reference to social variables. In the first instance the questions are:

1. What exactly is believed?
2. How is that belief distributed, e.g. who believes and who doesn't?
3. What are the sources of its credibility?
4. How is it defended against doubt and anomaly?

The explanatory resources available to sociologists that can shed light on these issues are:

- (a) a concern for the processes of socialization by which a group tries to transmit its culture to new members;
- (b) identifying the goals and interests of members of the group;
- (c) studying the processes of negotiation by which a degree of consensus may be achieved;
- (d) attending to the rhetorical devices that are used in the course of disputes; and
- (e) revealing the way the classifications, meanings and judgements that structure a body of knowledge are held as conventions.

Thus, studying the precise use to which a belief is put, whether implicitly or explicitly, can prove a valuable exercise. This can show a body of knowledge in an unexpected light, revealing the operation of interests that more traditional, "internalist" accounts may have

passed over. An example would be the studies which show us the social use given to the Corpuscular and Newtonian philosophies (cf. Jacob, 1976; Shapin and Schaffer, 1985).

The typical conception of the sociology of knowledge presents it as a limited exercise which is confined to studying the general conditions that might encourage or inhibit the growth of knowledge. Genuine knowledge, especially scientific knowledge, is assumed to be the result solely of our rational commerce with nature. The most that sociologists could hope to explain would be how we come to be in a position to conduct such enquiries and why, on occasion, our perception of their results is clouded or distorted; e.g. "ideology" might suppress "free enquiry", and bias the handling of evidence. Sociologists might therefore explain why, say, the Edinburgh middle classes in the mid-nineteenth century were widely attracted to phrenology, but they are unlikely to help us understand why the university anatomists and philosophers of the city dismissed it as error. Error and falsehood might need to be propped up by tradition, convention, authority and interest, but the real core of knowledge, when properly identified, will stand on its own two feet (i.e. observation and reason).

This limited conception, in which we have primarily a sociology of error, might be called the "weak programme" in the sociology of knowledge. Over the past two decades cogent reasons have emerged for throwing aside these restrictions and pursuing a stronger programme, i.e. one which sees all knowledge claims as social phenomena that are equally problematic. On the strong programme (*see STRONG PROGRAMME*) the aim is to avoid limiting our curiosity by treating what we take for granted as if it were unproblematic or self-explanatory. Consequently all bodies of institutionalized belief, whether we evaluate them as true or false, rational or irrational, are to be analysed in the same general way. For example, the practices of members of a culture where witchcraft is accepted should be no more problematic than those of a culture where it is rejected. A deviant member of the former, e.g. an individual who did not believe in witchcraft, should be no less a source of

explanatory interest than deviant members of ours, i.e. individuals who *do* believe in it. Of course, the sociologist must make various assumptions about the social agents under study, e.g. about their experience and material circumstances, but this can be done in ways that highlight the problematic character of their beliefs, whatever they are. How the world is doesn't fix how it will be described or understood by the people whose beliefs are to be explained. It is to be noted, though, that all bodies of institutionalized belief, however bizarre they may seem to us, must be compatible with ordinary human reasoning propensities, i.e. with "natural rationality".

Confidence in the viability of the stronger rather than the weaker programme in the sociology of knowledge has two sources: one empirical, the other theoretical. First, historians have given us impressively detailed studies of episodes from the history of science that have implicitly supported the stronger claim. Although historians do not necessarily see themselves as supporting any particular vision of knowledge, their enquiries have eroded the false distinction between socially sustained error and rationally sustained truth. This is because in practice they have uncovered ever more of the social contingencies in the development of knowledge, such as the operation of particular interests. They have done so regardless of the evaluative distinction between true or false theories, and correct or incorrect conclusions. Significantly, most of this work has not dealt with the genesis of theories (that remains locked in the enigmatic thought processes of individuals), but with their evaluation, i.e. with the context of justification, *not* the context of discovery. The scope of the historical work is very wide, ranging from studies of the way broad political concerns impinged on early conceptions of force, to the impact of narrowly professional interests on specific techniques and practices in botanical classification. Some of the work deals with episodes in the history of modern science, such as the disputes over Boyle's air pump experiments, while other work deals with early evolutionary debates, conflicts in modern statistics, and even the structure of elementary particle physics. It is far too extensive

and detailed to survey here, but a bibliographical guide, including a most valuable survey, is given at the end. This work provides the real substance of current thinking in the sociology of knowledge. It is detailed and empirical, not broad or programmatic; it exemplifies the concrete virtues of Anglo-Saxon historiography rather than the more speculative standpoint of continental philosophy. The persuasive power of the discipline cannot be appreciated without it.

Such work will not, however, convince a determined sceptic. It is always possible to gloss it in ways that minimize its significance. Just as it is always possible to read, say, the record of moral diversity as a catalogue of error and evil (and hence miss its full relativizing potential), we can do the same with the social history of science. It can be read as a record of mistaken ideas plus a study of the more or less accidental circumstances surrounding certain discoveries. At most it will then reveal the contingencies that happened to place people in a position to receive the indications of reason and observation. To offset this minimizing tendency it is necessary to furnish and defend the theoretical perspective that brings out the full significance of the historical work. It is the increasing awareness of this perspective that provides the second of the reasons for appreciating the true scope of the sociology of knowledge.

Philosophers themselves, especially philosophers of science, have done most of the groundwork. In particular they have exhibited the systematic underdetermination of theory by evidence in science (*see UNDERDETERMINATION OF THEORY*). Philosophers have then gone on to ask what is the *best* way to close the evidential gap. Sociologists and historians, by contrast, have enquired into the *actual* way it is closed. Clearly, underdetermination calls for the involvement of further factors in the construction of a rounded conception of the world. These will be partly psychological – our natural tendencies to draw certain kinds of conclusion – and partly sociological, e.g. our inherited assumptions, our local purposes and interests. To study this in more detail, assume we are scientists trying to test a theory. This is the point at which we interact

most directly with reality. We begin with a body of background assumptions (A) and an hypothesis (H). From these we draw a prediction (O) about an observation to be expected under certain precise conditions. Thus we have  $(A \& H) \rightarrow O$ . We then compare the predicted O with the report that sums up the results of our investigations, i.e. what we actually observed (call this O'). The effect of the comparison will be some rational state of belief regarding H. Precisely what that belief should be is described differently by different philosophers depending on whether they are realists, instrumentalists, or whatever. It may be probable belief of a certain degree, pragmatic acceptance or rejection, or some other appropriate rational attitude. For our purposes these different accounts of the correct stance that ought to be taken are not important, because they are normative rather than descriptive. What does matter, however, is the underlying logical structure that we have just described that these different accounts have in common. It can serve as a general description of the structure of our thinking. It emerges – and this is the vital point – that at each stage and facet of the process we are dealing with a social phenomenon. The claim is not that the processes are “purely” social, or “merely” social; we must never forget the psychological component of “natural rationality” running through them, but they each have a social aspect which is central to them and which cannot be removed without altering their character. Let us therefore go step by step through the standard structure, at each point locating its social character.

First, the background assumptions (A). These will mostly be inherited elements of an accepted culture, though they will be a special purpose selection from them, a selection sustained by the experience, traditions and purposes of some group of practitioners. Although many of these assumptions will be articulated and framed as propositions, some of them are best thought of as built into the habits and behaviour patterns of the group. They will be shared “practices”; and even the explicitly formulated assumptions will only have a determinate meaning because of

their being integrated into these practices. This is the sociological significance of the work of Wittgenstein (*see* WITTGENSTEIN) and its application to the philosophy of science in Kuhn's idea of a “paradigm” – a concrete scientific achievement that acts as a model for subsequent work.

Second, take the hypothesis under test (H). This will have no unique, privileged formulation. Any given articulation of it might be modified or improved or altered in order to bring out what it had “all along” been meant to express. A theory, even if it is identified with a specific set of sentences, is a fluid thing whose substance is, in principle, constantly negotiable, and which at any one time will have the status of a convention. The extent to which it is modified in the face of inevitable anomaly and trouble depends on a trade-off between the theory and the background assumptions. As Duhem made clear, we have a choice here. Nature, observation and the indications of reason do not point to a unique response. Once again the judgement involved, if it becomes accepted into the scientific culture, will have the form of a convention.

Third, the process by which the implications of the theory, the prediction, are extracted from it is not simply a matter of logically drawing out meaning which is unproblematically contained within the theory. There is no such unproblematic meaning. Every avenue by which a theory may be related to a particular and concrete episode of observation is fraught with difficulty, and can be rationally contested. At every point assumptions have to be made, e.g. about the purity of chemicals, the integrity and closure of the apparatus, etc. These assumptions might be otherwise, and their acceptability and determination again depend on the conventional practices of the group.

Fourth, the process of comparison between a prediction and the results of investigation is likewise underdetermined by reason and observation alone. As Kuhn pointed out, the tables of figures in text books which set predictions and observations side by side actually serve to define what counts, in the circumstances, as a “good fit”, i.e. what counts as good enough for the practitioners. Naïvely,

it may seem simple to decide if the predicted O and the observed O' are, or are not the "same", but of course criteria of "sameness" pose us with a classificatory problem, and the conventional and interested character of classification is widely acknowledged. The excellent fit between Gay-Lussac's work on gases and Dalton's atomic theory didn't impress Dalton. What may pass as a negligible difference from one point of view may look like a significant divergence from another. What is mere anomaly for one scientist could be a decisive refutation for another.

Little wonder, then, that a careful scrutiny of the details of a scientific episode reveals that it is shot through with contingencies whose matter-of-fact resolution depends on the ways that practitioners organise their interactions with one another. The sociology of knowledge isn't a subject whose attractions and plausibility depend on taking a distant and vaguely focused look at science in the attempt to trace broad patterns and trends. Quite the contrary; it makes better sense the more closely we look, because the higher the resolution of our microscope, the more obvious are the holes in the threadbare fabric of more traditional philosophical accounts.

See also GENETIC EPISTEMOLOGY; NATURAL SCIENCE; RELATIVISM.

#### BIBLIOGRAPHY

- A most valuable survey of the relevant historical literature up to 1982 is S. Shapin, "History of Science and Its Sociological Reconstructions," *History of Science* 20 (1982), 157–210. Other important works, which mostly appeared after the above, include:
- Collins, H.: *Changing Order* (Beverly Hills, CA: Sage, 1985).
- Desmond, A.: *The Politics of Evolution: Morphology, Medicine and Reform in Radical London* (Chicago, IL: University of Chicago Press, 1989).
- Jacob, M. C.: *The Newtonians and the English Revolution, 1689–1720* (Hassocks: Harvester, 1976).
- MacKenzie, D.: *Statistics in Britain, 1865–1930: The Social Construction of Scientific Knowledge* (Edinburgh: Edinburgh University Press, 1981).
- Pickering, A.: *Constructing Quarks: A Sociological History of Particle Physics* (Edinburgh: Edinburgh University Press, 1984).
- Pinch, T.: *Confronting Nature: The Sociology of Solar-Neutrino Detection* (Dordrecht: Reidel, 1986).
- Rudwick, M. J. S.: *The Great Devonian Controversy: The Shaping of Scientific Knowledge among Gentlemanly Specialists* (Chicago, IL: University of Chicago Press, 1985).
- Shapin, S. and Schaffer, S.: *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life* (Princeton, NJ: Princeton University Press, 1985).

DAVID BLOOR

**Socrates** see PLATO.

**solipsism** Doctrines and threats of solipsism are much older than the introduction of the term "solipsism" to mark them. The term derives from the Latin "solus ipse". This means literally "self alone", and less literally either "I alone exist" or else "I alone am conscious", yielding in the first case a more idealist form of solipsism querying the existence of an independent material world, and in the second case a more materialist form allowing for the (possible) existence of a material world but again not countenancing the existence of other minds or centres of consciousness.

It is common to distinguish between *metaphysical* (or ontological) *solipsism*, *epistemological solipsism* and *methodological solipsism*. The metaphysical solipsist contends that he alone is (exists, or is real); the epistemological solipsist contends that only he *certainly* is (exists, or is real); the methodological solipsist typically holds to solipsism as some sort of methodological strategy (either a method of inquiry or else a constructional programme). Descartes (see DESCARTES), in seeking to reconstruct his knowledge and base science on "a foundation wholly my own", can be seen as following a policy of Methodological Solipsism: this involved passing through a brief phase of Epistemological Solipsism from which he escaped by (supposedly) proving the existence of a non-deceiving God, while still retaining the first-personal basis.

Russell (see RUSSELL), who described solipsism as "hard to refute, but still harder to believe",



drew a distinction between what he called *dogmatic solipsism* and what he called *sceptical solipsism* (1948, pp. 191–8). The first corresponds more or less closely to Metaphysical Solipsism and the second to Epistemological Solipsism. His distinction parallels that between the atheist who denies outright and the agnostic who merely doubts. But whichever of these forms is in question Russell held that ordinary solipsism is a pusillanimous half-measure; a consistent or honest solipsist should adhere to the more radical *solipsism of the present moment*. The crux for Russell was that to go beyond the data provided by present experience required principles of inference that could not themselves be empirically justified: the cost of abandoning this radical form of solipsism was to accept that empiricism has (clearly pretty drastic!) limits. Santayana (see SANTAYANA) saw the escape route in “animal faith”, while recognizing merit in an attitude which confines attention to the passing moment. Wittgenstein confessed a temptation to assert, “All that is real is the experience of the present moment”.

It is notable that both Russell and Wittgenstein (see WITTGENSTEIN) took solipsism to be a serious philosophical issue. (The same applies to some important continental philosophers like Husserl (see HUSSERL).) What comes into focus with Wittgenstein’s treatment is what can be called *conceptual solipsism*. Whereas the more familiar variety of solipsist denies or doubts whether there are in fact any other minds, the conceptual solipsist contends that it is inconceivable that there are any. Although the distinction between conceptual solipsism and factual versions of solipsism cuts across the distinction between dogmatic and sceptical solipsism, it remains true that a dogmatic solipsist had better be a conceptual solipsist and that a conceptual solipsist had better not be a sceptical solipsist. For the position that there are no other minds but in fact there might have been has nothing whatsoever going for it; and the conceptual solipsist cannot allow any probability at all to the proposition that other minds exist (see OTHER MINDS).

Wittgenstein famously remarked that “If one has to imagine someone else’s pain on the

model of one’s own, this is none too easy a thing to do: for I have to imagine pain I *do not feel* on the model of the pain which I *do feel*” (1953, §302). Somewhat similarly it might be contended that one cannot conceive of a mind of which one is not conscious on the model of one’s own conscious mind: hence that other conscious minds are inconceivable to one. In order, however, not to fall too easy prey to conceptual solipsism it is worth noting that even if it is conceded that I cannot conceive of another mind it by no means follows that it is inconceivable that there are other minds. This can be seen from considering Thomas Nagel’s claim that he does not know what it is like to be a bat: he implies that there exists something (the bat’s mind or experiences) of which he cannot conceive. (That said, the inability to conceive some other mind does result in some sort of quasi-solipsistic isolation from it.)

A problem for the solipsist is that he or she may be well aware that others can argue exactly the same way. Ayer (see AYER) pointed out (1986, p. 29) that philosophers are tempted by a sort of generalized solipsism for which he coined the paradoxical term *multiple solipsism*. (For example, “pain” in A’s mouth means real pain when applied to his own but pain-behaviour when applied to B’s, and the same goes for “pain” in B’s mouth.) He accused the early Wittgenstein, Carnap and his own earlier self of this type of view which he came to think was incoherent (1986, p. 38). Wittgenstein overcame his early predilection for this and other forms of solipsism by means of his private language argument, but whether Carnap ever came to see the error of his ways – if such it be – is less clear.

The project to which Carnap (1967, p. 102) (see CARNAP) applied the term “methodological solipsism” was that of each person constructing the whole language including that about physical objects, cultural objects, other minds and even their own mind on an “autopsychological” basis, where that took the form of an initially subjectless given. This basis he held was a matter of choice but it possessed the (alleged) advantage of following the epistemological order (1967, p. 101).

Bennett, however, has used the term for the acknowledgement of what he takes to be an inevitable feature of the human epistemic situation: each of us can utilize and assess evidence only according to how things seem to us (1974, pp. 66–9). The term “methodological solipsism” has been given fresh currency from an influential article which Putnam published in 1975. He used it for the assumption that “no psychological state, properly so called, presupposes the existence of any individual other than the subject to whom that state is ascribed” (p. 220). Putnam’s usage is followed by Fodor (1987, ch. 2) in proposing methodological solipsism as a research strategy in cognitive psychology. How far that project is removed from orthodox solipsism is shown in the fact that it is to be used by psychologists in studying other minds.

See also ARGUMENT FROM ANALOGY; OTHER MINDS; WITTGENSTEIN.

#### BIBLIOGRAPHY

- Ayer, A. J.: *Ludwig Wittgenstein* (Harmondsworth: Penguin, 1986).  
 Bennett, J.: *Kant’s Dialectic* (Cambridge: Cambridge University Press, 1974).  
 Carnap, R.: *The Logical Structure of the World* (London: Routledge, 1967).  
 Fodor, J.: *Psychosemantics* (Cambridge, MA: MIT Press, 1987).  
 Putnam, H.: *Philosophical Papers*, vol. 2 (Cambridge: Cambridge University Press, 1975).  
 Russell, B.: *Human Knowledge* (London: Allen and Unwin, 1948).  
 Wittgenstein, L.: *Philosophical Investigations* (Oxford: Blackwell, 1953).

CLIVE BORST

**Spinoza, Benedict (also Baruch) (1632–77)** Dutch philosopher. Spinoza’s epistemology is both a response to his seventeenth-century context (see RATIONALISM) and an integral part of his philosophy. His 1663 *Descartes’s “Principles of Philosophy”*, which is a reformulation of the first two parts of Descartes’ famous work (see DESCARTES)

into the axiomatized “geometrical order” of presentation modelled on Euclid, manifests a deep understanding of Cartesian epistemology. Spinoza’s own epistemology is presented primarily, though not exclusively, in Part II of his *Ethics* and in an earlier unfinished work, *Treatise on the Emendation of the Intellect*, both of which were first published in the *Opera postuma* of 1677. Although he follows Descartes in emphasizing the distinction between intellect and imagination as representational faculties, his epistemology nevertheless differs from that of Descartes in a number of fundamental respects.

One such difference – and one on which Spinoza particularly insists – lies in his rejection of Descartes’ account of error, according to which error is the result of the will’s freely affirming ideas from which it has the power to withhold assent. On Spinoza’s deterministic alternative, every idea naturally involves affirmation of its own content; an idea can and will be denied or doubted only when the mind also has another idea which contradicts the first or calls it into question (*Ethics* 2 p. 49).

An equally fundamental difference lies in Spinoza’s doctrine that truth “requires no sign” (*TEI* 34–6) and is instead “its own standard” (*Ethics* 2 p. 435). Descartes treats “clarity and distinctness” as an introspectible characteristic of ideas, a characteristic that can and must – after an appropriate process of vindication – be used as a criterion for the quite separate characteristic of truth. Spinoza rejects the need for any such separate criterion, and hence also for any such process of vindication. Indeed, since the truth of true ideas can be discerned directly for Spinoza – “the true thought is distinguished from a false one not only by an extrinsic, but chiefly by an intrinsic denomination” (*TEI* 69) – Cartesian methodological scepticism becomes unnecessary. Doubt, he holds, arises only when the mind has two related ideas, at least one of which is obscure (*TEI* 78). Its proper remedy is to proceed in the correct order, deducing effects from causes, so that the mind will never have ideas whose bearing on one another is uncertain (*TEI* 80). Thus, to begin as Descartes does, by searching out

grounds for doubt, is from Spinoza's point of view simply to guarantee that one will proceed in the wrong order, and thereby to put as many obstacles as possible in the way of one's ultimate success.

Partly because of his claim that truth is "chiefly an internal denomination", some commentators have maintained that Spinoza held a "coherence" theory of truth; others, citing his axiom that "a true idea must agree with its object" (*Ethics* 1a7), have maintained that he held a "correspondence" theory of truth (see TRUTH). In fact, however, his well-known doctrine that modes of thought are identical with their objects (*Ethics* 2p7s) arguably allows him to hold that truth consists *both* in what he calls the internal "adequacy" of an idea, *and* in what he calls its "agreement" with its object, on the grounds that these two characteristics of ideas are really the same characteristic considered in two different ways. His monism requires that there be only one thinking substance, God, of which human minds are modes. He asserts that all ideas are true as they exist in God; however, they may be mutilated and confused, and hence false, as they exist in individual human minds (*Ethics* 2 pp. 32–5).

Spinoza distinguishes three kinds of knowledge (*Ethics* 2 p.40s2). The first and lowest is opinion (*opinio*), consisting of "random" sensory experience (i.e. experience not determined by the intellect) and external reports; this kind of knowledge is imaginative and inadequate. The second is reason (*ratio*), which involves adequate intellectual knowledge of things through their properties, as opposed to their essences. This knowledge includes and is largely based on the "common notions", which are ideas of characteristics that are "common to all things" and are "equally in the part and in the whole". He gives few examples of such characteristics, but he does cite, as giving rise to common notions, the fact that all bodies involve the concept of the same attribute (i.e. extension), the fact that all bodies are susceptible to motion and rest, and the fact that all bodies are susceptible to degrees of motion (Lemma 2 after *Ethics* 2 p. 13). Thus it appears that, for Spinoza, at least some knowledge of physics can be of the

second kind, based on the knowledge of universal properties of bodies. (Knowledge of the second kind is, however, by no means restricted to physics; it is also possible in psychology at least.) The third and highest kind of knowledge is intuition (*scientia intuitiva*), which "proceeds from an adequate idea of the formal essence of certain attributes of God to the adequate knowledge of the essence (as opposed to the mere properties) of things". Knowledge of both the second and third kinds is thus "adequate" and intellectual. In Spinoza's ethical theory, adequate knowledge, and especially knowledge of the third kind, constitutes the greatest good of life.

#### WRITINGS

- Spinoza Opera*, ed. C. Gebhardt, 4 vols (Heidelberg: Carl Winter, 1923).  
*The Collected Works of Spinoza*, ed. E. M. Curley, 2 vols (Princeton, NJ: Princeton University Press, 1985), vol. 1.

#### BIBLIOGRAPHY

- Curley, E. M.: "Experience in Spinoza's Theory of Knowledge," in *Spinoza: A Collection of Critical Essays*, ed. M. Grene (Garden City, NY: Anchor, 1973), 25–59.  
 Gueroult, M.: *Spinoza*, 2 vols (Paris: Aubier, 1974), vol. 2, *L'Ame*.  
 Matheron, A.: "Spinoza and Euclidean Arithmetic: The Example of the Fourth Proportional," in *Spinoza and the Sciences*, ed. M. Grene and D. Nails (Dordrecht: Reidel, 1986), 125–50.  
 Parkinson, G. H. R.: *Spinoza's Theory of Knowledge* (Oxford: Clarendon Press, 1954).  
*Studia Spinozana* 2 (1986): issue devoted to "Spinoza's Epistemology."

DON GARRETT

**Stoic epistemology** The Stoic school was founded in Athens around the end of the fourth century BC by Zeno of Citium (335–263 BC). The school takes its name from the Stoa *poikile*, or painted colonnade, where its members met to discuss and expound their views. Following Zeno, orthodox Stoics divided

philosophy into three parts: logic, physics and ethics. Epistemological issues were a concern of logic, which studied *logos*, reason and speech, in all of its aspects, not, as we might expect, only the principles of valid reasoning – these were the concern of another division of logic, dialectic. The epistemological part, that concerned with canons and criteria, belongs to logic conceived in this broader way because it aims to explain how our cognitive capacities make possible the full realization of reason in the form of wisdom, which the Stoics, in agreement with Socrates, equated with virtue and made the sole sufficient condition for human happiness. Indeed, they went so far as to identify virtue and its exercise as the sole good, and their opposites, vice and vicious action, as the sole evil. At the same time, they held that items popularly supposed to be good or evil – fame, wealth and pleasure or ill repute, poverty and pain – are strictly speaking indifferent. The attitude of superiority required by this view towards the indifferents, especially to the so-called evils, is behind our use of the expressions “stoic” and “stoical”. This attitude is not, however, as we would expect, an affective, but a cognitive condition; and much of the emphasis in the Stoics’ position on knowledge satisfying the most exacting standards is due to the overriding moral importance they attach to wisdom.

But if the Stoics’ motivations are in some ways unfamiliar, their epistemological position, in its broad outlines at any rate, is much less so; it is in many ways similar to positions familiar to us from modern philosophy, far more so than the views of Plato (see PLATO) or Aristotle. This is not surprising, as the ancient authors who preserve most of what we know about Stoic epistemology, Cicero, Diogenes Laertius and Sextus Empiricus (see SEXTUS EMPIRICUS), exerted considerable influence, directly and indirectly, on the philosophers of the early modern period. Hume’s acknowledged debt to Cicero is only the most conspicuous example.

The Stoics are foundationalists (see FOUNDATIONALISM); according to them, we must start with a stock of secure, directly given

impressions which, without standing in need of any support themselves, in one way or another somehow support everything else we know. And their position is an empiricist one (see EMPIRICISM); for the ground-level impressions of which it makes use are furnished by the senses. But before going on to consider how these impressions do what is required of them, we should take note of an element of complexity in the Stoic account. According to it, the impressions of adult human beings, unlike those of children and animals, are rational. This does not mean their judgements, which arise according to the Stoa when assent is given to an impression, cannot be criticized as irrational. Rather, what is meant is that these impressions have a rational structure; i.e. they have propositional content; they are impressions that something is the case, which represent the world as being a certain way by means of a stock of general concepts (see SELLARS; SENSATION/COGNITION). And the Stoics’ empiricism first makes its presence felt in their account of how these concepts are acquired. Indeed Zeno seems to deserve credit for being the first to have compared the mind at birth to a blank slate (see TABULA RASA). Into this slate, he maintains, experience gradually inscribes the so called common notions, which will eventually make possible rational thought by means of rational impressions. The process begins with impressions of simple perceptual features, repeated experience of which results in memories; these in turn furnish the basis for the common notions – a developmental process not unlike that described by Aristotle. The process is completed at age seven or fourteen – our sources disagree – and thought is conducted thereafter by means of rational impressions.

As we noted, this is not to say that human beings who have reached this stage cannot have false beliefs, reason invalidly or form irrational attachments. Indeed, according to the Stoa, such is the corrupt condition of mankind that all human beings are thoroughly and viciously irrational in this sense. But they are now in principle capable of realizing fully the potential for reason which nature has given them.

Reason is fully realized as knowledge, which the Stoics define as secure and firm cognition unshakeable by argument. According to them, no one except the wise man can lay claim to this condition. He is armed by his mastery of dialectic against fallacious reasoning, which might lead him to draw a false conclusion from sound evidence, and thus possibly force him to relinquish the assent he has already properly conferred on a true impression. Hence, as long as he does not assent to any false ground-level impressions, he will be secure against error, and his cognitions will have the security and firmness required of knowledge. Everything depends, then, on his ability to avoid error in his ground-level perceptual judgements. To be sure, the Stoics do not claim that the wise man can distinguish true from false perceptual impressions; that is beyond even his powers. But they do maintain that there is a kind of true perceptual impression, the so-called cognitive impression, by confining his assent to which the wise man can avoid giving error a foothold.

An impression is cognitive when it is (a) from what is (the case), (b) stamped and impressed in accordance with what is, and (c) such that could not arise from what is not. And because all of our knowledge depends directly or indirectly on it, the Stoics make the cognitive impression the criterion of truth. It makes possible a secure grasp of the truth not only by guaranteeing the truth of its own propositional content, which in turn supports the conclusions that can be drawn from it; even before we become capable of rational impressions, nature must have arranged for us to discriminate in favour of cognitive impressions so that the common notions we end up with will be sound. And it is by means of these concepts that we are able to extend our grasp of the truth through inferences beyond what is immediately given; for this reason, the Stoics sometimes also speak of two criteria: cognitive impressions and common notions.

These Stoic claims about the cognitive impression elicited a strong sceptical response from Plato's successors in the Academy (see *ACADEMY*), who produced a mass of arguments intended to suggest that, contrary to the

Stoic view, it is always possible in the case of any true impression, however clear and forceful, that another impression indistinguishable from the first could have arisen in circumstances that make it false. If so, there are no true impressions such that they could not be false as far as an intrinsic, discriminable feature of theirs is concerned, thus no guarantee in any given case that assent has not been conferred on a false impression. These arguments – including those based on the power of the gods to induce true-seeming false impressions and the inability of dreamers and madmen to distinguish true from false impressions – were to exert a tremendous influence on the later history of epistemology. We should not suppose because of our greater familiarity with the sceptics' case, however, that the Stoics were not able to mount a powerful defence of their position.

The details of that defence are beyond our present scope, but there are two features of the Stoic position which must not be overlooked if we are to avoid giving a seriously misleading picture of it. First, the emphasis placed upon the security and certainty of the wise man's judgements by the Stoa does not exclude judgement under conditions of uncertainty. The wise man will sometimes, e.g., choose actions on the basis of expectations about the future which will not be fulfilled. To be sure, he does not assent to impressions which later developments will show to have been false, but he does sometimes judge that certain expectations are reasonable. And this judgement will not be falsified by later developments. Second, the Stoics are empiricists not only regarding concept formation. Though on their view every event is the necessary outcome of antecedent causal conditions, a grasp of whose nature should in principle support a rational, conceptually sanctioned inference predicting that event, even the wise are very rarely in a position to achieve such a grasp or draw such an inference. Instead, they must rely on observed relations of conjunction and sequence between events the underlying causal connections between which remain unknown. A large part of the knowledge on which the Stoic wise man must rely will also be empirical in this sense.



## WRITINGS

Fragments of the Stoics can be found in two places:

- Arnim, H. von, ed.: *Stoicorum Vetera Fragmenta*, 3 vols (Leipzig: Teubner, 1903–5); vol. 4, indices by M. Adler (Leipzig: Teubner, 1924).  
 Long, A. A. and Sedley, D. N., ed. and trans.: *The Hellenistic Philosophers*, 2 vols (Cambridge: Cambridge University Press, 1987).

There are texts with English translation of the major ancient sources in the Loeb Classical Library series:

- Bury, R. G., ed. and trans.: *Sextus Empiricus: Outlines of Pyrrhonism and Against the Mathematicians*, 4 vols (London: Heinemann, 1933–49).  
 Hicks, R. D., ed. and trans.: *Diogenes Laertius: Lives of Eminent Philosophers*, 2 vols (London: Heinemann, 1925), vol. 2, bk VII.  
 Rackham, H., ed. and trans.: *Cicero: De Natura Deorum and Academica* (London: Heinemann, 1933).

## BIBLIOGRAPHY

- Annas, J.: “Stoic Epistemology,” in *Epistemology*, ed. S. Everson (Cambridge: Cambridge University Press, 1990), 184–203.  
 Frede, M.: “Stoics and Sceptics on Clear and Distinct Ideas,” in *The Sceptical Tradition*, ed. M. F. Burnyeat (Berkeley, CA: University of California Press, 1983), 65–93.  
 Sandbach, F. H.: “Phantasia kataleptike” and “Ennoia and prolepsis,” in *Problems in Stoicism*, ed. A. A. Long (London: Athlone, 1971), 9–21, 22–37.  
 Striker, G.: “Kriterion tes aletheias,” *Nachrichten der Akademie der Wissenschaften in Göttingen Phil.-hist. klasse*, Nr 2 (1974), 47–110.

J. V. ALLEN

**Strawson, Peter Frederick (1919–2006)**

P. F. Strawson spent most of his career at the University of Oxford, becoming Waynflete Professor of Metaphysical Philosophy in 1968. He published extensively, principally in metaphysics and philosophical logic. Though he did not offer a general theory of knowledge, a dominant aim was that of

devising a response to scepticism. Strawson offered different reasons for rejecting scepticism, but he also thought that it was wrong to provide “justifications” for our opinions resting on a supposedly privileged foundation. In 1952 he argued that inductive scepticism has its source in confusion, and no justification of our practices is required (see PROBLEMS OF INDUCTION). His basic reason, though, was that our fundamental objective categories cannot be abandoned at any stage in our thinking, nor can we even describe our experiences without them (see Strawson, 1979). Strawson’s chief preoccupation as an epistemologist was to substantiate and explore the consequences of this.

In *Individuals* and *The Bounds of Sense*, he tried to show that scepticism is involved in a deep incoherence, because the intelligibility of the concepts which the sceptic himself employs requires acceptance of things the sceptic is doubtful of. Arguments to show this are called transcendental arguments (see TRANSCENDENTAL ARGUMENTS).

In *Individuals*, Strawson argued that for us space-occupying bodies are the basic objects of thought. We also think about possessors of consciousness, persons, and they must, at least characteristically, have both psychological and material attributes. The chief epistemological thesis is that it is a condition for having such a conceptual scheme that the basis we standardly regard as sufficient for certain judgements must in fact be sufficient. Thus, our normal ways of telling that another is in a given psychological state must be “logically adequate kinds of criteria” (Strawson, 1959, p. 105).

In *The Bounds of Sense* Strawson further developed such arguments by defending many of Kant’s central views (see KANT) after detaching them from transcendental idealism. In particular, Strawson supported a version of Kant’s thesis that for a sequence of experiences to belong to a self-conscious subject they must constitute experience of an objective world. The central idea is that a substantive concept of an experiencing subject has employment only where we can apply the distinction between how in those experiences things seem to be and how they really are.

This distinction can apply only if the experience is thought of as being of an objective world.

Debate about these arguments has centred on whether the conceptual dependencies are strongly enough established, and also on what the requirement of objective experience amounts to.

In *Scepticism and Naturalism* Strawson presented a new response, with its roots in Hume and Wittgenstein (see HUME; WITTGENSTEIN). The thesis is that sceptical arguments should be dismissed as idle, since they cannot persuade us. Counter-arguments, even of a transcendental type, are not needed, though they may reveal conceptual connections. The reason for this is that we cannot help believing in (say) bodies and other minds.

While the psychological claim seems indisputable, criticisms of Strawson's naturalism have focused on whether the psychological facts justify a dismissal of sceptical arguments. It is widely felt that a rejection of scepticism is unsatisfactory unless, at the very least, some mistake can be located in such arguments.

See also OTHER MINDS; PROBLEMS OF INDUCTION; TRANSCENDENTAL ARGUMENTS.

#### WRITINGS

- Introduction to Logical Theory* (London: Methuen, 1952).  
*Individuals* (London: Methuen, 1959).  
*The Bounds of Sense* (London: Methuen, 1967).  
*Scepticism and Naturalism* (London: Methuen, 1985).  
 "Perception and Its Objects," in *Perception and Identity*, ed. G. F. Macdonald (London: Macmillan, 1979).

#### BIBLIOGRAPHY

- Ayer, A. J.: "The Concept of a Person," in his *The Concept of a Person and Other Essays* (London: Macmillan, 1963).  
 Evans, G.: "Things without the Mind – a Commentary upon Chapter Two of Strawson's *Individuals*," in his *Collected Papers* (Oxford: Clarendon Press, 1985).

Stroud, B.: "Transcendental Arguments," *Journal of Philosophy* 65 (1968), 241–56.

P. F. SNOWDON

**strong programme** In the sociology of knowledge the "strong programme" embodies the aim of sociologically explaining not just inadequately grounded or biased beliefs, but also true and rationally held beliefs. The hope is to provide not just a sociology of error, but a sociology of genuine knowledge. The "weak programme" represents the more limited aim under which the sociologist studies *either* the generalized preconditions for the growth of knowledge, *or* the specific sources of distortion, but treats the rational acts of apprehending inferences and evidence as falling outside the scope of the discipline. Adopting the strong programme does not mean denying the role either of sense experience or of our natural reasoning propensities, but it does mean seeing them as inadequate by themselves to account for the phenomenon called "knowledge". In particular it means appreciating the way inputs from these sources must be organised and interpreted in the course of constructing a shared representation of reality that is held as a convention. One formulation of the strong programme declares its explanatory aims to be:

1. *Causal*: it would be concerned with the conditions that bring about belief – though, of course, sociological causes will always cooperate with others, e.g. psychological processes.
2. *Impartial*: both true and false, rational and irrational beliefs would require explanation.
3. *Symmetrical*: both sides of the above dichotomies would call for causal explanations of the same type, e.g. in terms of socialization, responsiveness to interests, the pursuit of a multitude of practical goals, the need to construct conventions and sustain consensus, to integrate new knowledge in existing culture, etc.
4. *Reflexive*: in principle its patterns of explanation would have to be applicable to itself. (If they were not, it would be a standing refutation of its own theories.)

Naturally this programme has proven controversial.

See also SOCIOLOGY OF KNOWLEDGE.

#### BIBLIOGRAPHY

- Barnes, B.: *Scientific Knowledge and Sociological Theory* (London: Routledge and Kegan Paul, 1974).  
 Barnes, B.: *T. S. Kuhn and Social Science* (London: Macmillan, 1982).  
 Bloor, D.: *Knowledge and Social Imagery*, 2nd edn (Chicago, IL: University of Chicago Press, 1991).  
 Brown, J. R., ed.: *Scientific Rationality: The Sociological Turn* (Dordrecht: Reidel, 1984).  
 Brown, J. R.: *The Rational and the Social* (London: Routledge, 1989).  
 Collins, H.: *Sociology of Scientific Knowledge: A Source Book* (Bath: Bath University Press, 1982).

DAVID BLOOR

**Suárez, Francisco (1548–1617)** Spanish philosopher and theologian known as Doctor Eximius. Born in Granada, he joined the Society of Jesus in 1564, studied at Salamanca, and taught law, theology and philosophy there and also at Rome, Coimbra and other leading universities. Apart from his many theological works, he wrote four important philosophical treatises. *De legibus* (1612) deals with traditional legal problems and with issues arising from the discovery of America, *De Deo uno et trino* (1606) is concerned with the nature of the Trinity but has implications for philosophical theology, and *De anima* (1621) concentrates on psychology and epistemology. His most important work, however, is the *Disputationes metaphysicae* (1597), which is the first systematic and comprehensive work of metaphysics written in the West that is not a commentary on Aristotle's *Metaphysics*. It is divided into fifty-four disputations that cover every metaphysical topic known at the time. Its influence was immediate and lasting. Within a few years of its publication, it had become the standard text in the field in continental Europe. Its impact can be seen in the thought

of Descartes, Leibniz, Wolff, Schopenhauer and others.

Suárez's sources are vast, ranging from antiquity to his contemporaries, but it is in Aristotle and Thomas Aquinas (see ARISTOTLE; AQUINAS) that he finds most frequent inspiration. His main contributions to philosophy occur in metaphysics and law, although his thought is also relevant in many other areas, including epistemology. Suárez followed the scholastic procedure of raising epistemological issues in the context of metaphysics; thus preestablished metaphysical categories determined to a great extent the type of epistemological questions he raised. Three Suárezian doctrines are of particular interest for epistemology. They concern the object of metaphysics, knowledge of the singular, and middle knowledge. The first two serve as a bridge between scholastic and modern philosophy.

According to Suárez, the proper object of metaphysics is the objective concept of being (*DM* II, 1). An objective concept is a thing as conceived in the mind rather than the thing itself (*res ipsa*) or the act of understanding (*conceptus formalis*) whereby the mind understands. This doctrine has been interpreted as a step toward the progressive mentalization of metaphysics characteristic of modern philosophy, in which the object of the discipline is not reality (*res*), but a mental representation of it.

Suárez's understanding of the knowledge of the singular is another example of how he may be considered a bridge between scholastic and modern philosophy. He holds that our intellect knows the singular through a proper and separate concept (*DA* IV, 3, 3) without having to resort to reflection (IV, 3, 7). This thesis lends support to an empiricist point of view (see EMPIRICISM) and undermines the position, accepted by Thomists among others, that knowledge of singulars is only indirect, through universals. Moreover, it contributes to the erosion of any kind of realistic epistemology where real natures are the object of science.

Middle knowledge (*scientia media*) is the knowledge that God has of what every free creature would freely do in every situation

where the creature could possibly be. It is a 'middle' knowledge because it stands somewhere in between the knowledge God has of the merely possible and the knowledge he has of the actual. Suárez used this notion to explain how God can control human action without violating free will (*De gratia*, prol. 2. 7.1).

See also AQUINAS; CONCEPTS; OCKHAM.

#### WRITINGS

*Opera omnia*, 28 vols; ed. C. Berton (Paris: Vivès, 1856–61).

#### BIBLIOGRAPHY

- Adams, R. M.: "Middle Knowledge and the Problem of Evil," *American Philosophical Quarterly* 14 (1977), 109–17.
- Cronin, T. J.: *Objective Being in Descartes and Suárez* (New York: Garland, 1987).
- Gracia, J. J. E. and Davis, D.: *The Metaphysics of Good and Evil according to Suárez* (Munich and Vienna: Philosophia Verlag, 1989).
- Peccorini, F. L.: "Knowledge of the Singular: Aquinas, Suárez and Recent Interpreters," *The Thomist* 38 (1974), 606–55.

JORGE J. E. GRACIA

**subjectivism** What is a subjective account of epistemic rationality? This is none too easy to state precisely (it is no easier to say what subjectivism in ethics is), but as a rough working definition it will do to say that an account is subjective just in case it implies that the standards of rational belief are those that the individual believer or the individual's community or the human community at large would either approve of or take for granted in so far as their ends are intellectual. Subjectivism, so understood, can take an individualistic or intersubjective form, but either way a subjective account will imply that the standards of rational belief are somehow the products of our beliefs, our dispositions, or our practices.

For example, if the standards that my beliefs must meet in order to be rational,

according to an account, are those that I myself would regard as intellectually sound were I to be suitably reflective, then the account is a subjective one. Similarly, if an account implies that the relevant standards are those that would be approved of by most members of my community or those that are presupposed by the intellectual practices of my community, then once again the account can be classified as a subjective one, albeit one with a social rather than an individualistic cast to it. Finally, if an account implies that the standards of rational belief are those that would be endorsed or presupposed by the human community at large, then the account yet again can be plausibly regarded as a subjective one, provided that there is nothing else in the account to guarantee that adhering to these standards is a reliable way to acquire true beliefs. Indeed, this last qualification provides a good negative test of subjectivity. If an account of epistemic rationality implies that by being rational, individuals are assured of being reliable (or at least more reliable than they would be if they were irrational), then the account is not a subjective one.

Thus reliabilist accounts of rational belief are paradigmatically objective (see RELIABILISM). Classical foundationalist accounts are also objective (see FOUNDATIONALISM). What about coherentist accounts (see COHERENTISM)? It can be tempting to think they are best classified as subjective, since what it is rational for us to believe, according to coherentists, is in large part a function of what we happen to believe. But any account of rational belief will allow for subjective inputs of some sort, whether they be in the form of experiences or beliefs or whatever. The crucial question is whether the standards that relate these inputs to rational belief are objective or subjective, and in the case of coherentism the standards are typically explicated in a thoroughly objective manner. Coherentists ordinarily insist that if my beliefs are to be coherent and hence rational, they must at a minimum be consistent. It does not matter whether I think that inconsistency is always and everywhere to be avoided, and it does not matter whether I would think this were I

to be reflective. Likewise, it does not matter whether the members of my community or the human community at large think this or whether their intellectual practices presuppose this. According to coherentists, inconsistency implies incoherence, and it is always irrational for us to be incoherent, regardless of our own subjective standards. (Similar things can be said about Bayesian approaches to epistemology (see BAYESIAN EPISTEMOLOGY in Part I).)

One simple consideration in favour of subjectivism is that many of our judgements of rationality cannot be plausibly understood in an objective manner. For example, when we are assessing the beliefs of individuals from a different culture, perhaps far removed from us in both time and place, it often seems appropriate to do so in terms of their own standards or at least in terms of some standard that is relative to their community. This will seem particularly attractive when the individuals belong to a culture that we judge to be less advanced than our own. The intellectual methods and practices used in this culture may be at odds with those that we now think are best, but we may think it is unfair to evaluate their beliefs in terms of our more sophisticated methods. However, it is not unfair, we think, to evaluate their beliefs in terms of the methods that were standard in their community.

Indeed, if we were not prepared to evaluate these individuals in terms of their own standards and to give expression to our evaluations using the language of rationality, we would find ourselves in a dilemma. Either we would have to find some way of insisting that their methods really are our methods, the ones that we take to be reliable, or we would be forced to say that there is no interesting sense in which they are rational. In short, we would be forced either to make them into us or to dismiss them as irrational.

Another consideration in favour of subjectivism appeals to first-person rather than third-person considerations. Suppose that we are under the control of an evil demon who widely deceives us without our being aware of it. The demon deprives us of an opportunity of being reliable and in so doing deprives us of knowledge, but it does not seem as if the

demon thereby also deprives us of any chance of being rational. In particular, if there is nothing that indicates to us that we are under the control of this demon, it is rational for us to believe what our senses naturally incline us to believe. And yet, in this demon world, doing so is unreliable. Thus, there is at least one sense of rational belief that is not tied intimately to truth or even to likely truth. But then, the question is, what is this sense of rational belief? One natural suggestion is that it is a subjective sense. In the demon world, it is rational for us to trust our senses because, complications aside, doing so conforms to our own deep epistemic standards: even if we were deeply reflective, we would tend to think that this is part of a reliable intellectual strategy for us.

Why, then, have subjective accounts of epistemic rationality so often been considered non-starters? Part of the answer is that many epistemologists have been primarily concerned with the explication of knowledge, and they have assumed that knowledge is something like rational true belief (absent Gettier problems (see GETTIER PROBLEM)). Thus, they have thought that whatever else we may want to say about rational belief, it had better be the sort of thing that turns true belief into a good candidate for knowledge. But if rational belief is construed subjectively, this need not be the case.

Subjectivists, for their part, will deny this assumption. They will deny that the conditions of rational belief always turn true belief into a good candidate for knowledge (see RELIGIOUS BELIEF, EPISTEMOLOGY OF). Of course, if it suits our philosophical purposes, we can stipulate that this is so for at least one sense of rational belief, but then subjectivists will insist that there is at least one other sense of rational belief, important for our everyday evaluations of each others' beliefs and important also given the history of epistemology, that is essentially a matter of living up to our own standards. This is a sense that equates rationality with success in meeting a certain kind of criticism, either self-criticism or criticism by one's community. And in this sense, a rational true belief need not always be a good candidate for knowledge.



Another motivation for not taking subjectivism seriously in epistemology is the fear that it commits us to some unacceptable form of relativism (see RELATIVISM). But since one can be a subjectivist about questions of rational belief without being committed to any relativistic notion of truth, it is not clear that this fear has any real basis. Yet another fear has to do with irresolvable disagreements. Given subjectivism, it cannot be assumed that all theoretical disputes can in principle be settled if the disputants are equally well informed and fully rational. This is an implication of subjectivism, but it is hard to see why it is an objection to it. In the practical realm we no longer find it puzzling that two people can be in conflict despite being equally well informed and fully rational. There is no good reason for thinking that this should strike us as any more puzzling in the theoretical realm.

See also ETHICS AND EPISTEMOLOGY; EXTERNALISM/INTERNALISM; OBJECTIVE/SUBJECTIVE; OBJECTIVITY; SOCIAL SCIENCES.

## BIBLIOGRAPHY

- Foley, R.: *The Theory of Epistemic Rationality* (Cambridge, MA: Harvard University Press, 1987).  
 Goldman, A. I.: "Strong and Weak Justification," *Philosophical Perspectives* 2 (1988), 51–71.  
 Kvanvig, J.: "Subjective Justification," *Mind* 93 (1984), 71–84.  
 MacIntyre, A.: *Whose Justice? Which Rationality?* (Notre Dame, IN: University of Notre Dame Press, 1989).  
 Rorty, R.: *Philosophy and the Mirror of Nature* (Princeton, NJ: Princeton University Press, 1979).  
 Wilson, B., ed.: *Rationality* (New York: Harper and Row, 1970).

RICHARD FOLEY

**subjectivity** Subjectivity has been attributed variously to certain concepts; to certain properties of objects; and to certain modes of understanding. The overarching idea of these attributions is that the nature of the concepts, properties or modes of understanding in

question is dependent upon the properties and relations of the subjects who employ those concepts, possess the properties or exercise those modes of understanding. The dependence may be a dependence upon the particular subject, or upon some type which the subject instantiates. What is not so dependent is objective. In fact, there is virtually nothing which has not been declared subjective by some thinker or other, including such unlikely candidates as space and time (Kant (see KANT)) and the natural numbers (Brouwer). In recent years there has been a lively debate about the more plausible candidates.

There are several sorts of subjectivity to be distinguished, if subjectivity is attributed to a concept, considered as a way of thinking of some object or property. It would be much too indiscriminating to say that a concept is subjective if particular mental states are mentioned in the correct account of mastery of the concept. For instance, if the later Wittgenstein is right, the mental state of finding it natural to go on one way rather than another has to be mentioned in the account of mastery of any concept. All concepts would then be counted as subjective. We can distinguish several more discriminating criteria. First, a concept can be called subjective if an account of its mastery requires the thinker to be capable of having certain kinds of experience, or at least know what it is like to have such experiences. Variants on this criterion can be obtained by substituting other specific psychological states in place of experience. If we confine ourselves to the criterion which does mention experience, then concepts of experiences themselves plausibly meet the condition. What have traditionally been classified as concepts of secondary qualities – such as *red*, *tastes bitter*, *warm* – have also been argued to meet this criterion. The criterion does, though, also include some relatively observational shape concepts. The relatively observational shape concepts *square* and *regular diamond* pick out exactly the same shape properties, but differ in which perceptual experiences are mentioned in accounts of their mastery – different symmetries are perceived when something is

seen as a diamond from when it is seen as a square. This example shows that from the fact that a concept is subjective in this sense, nothing follows about the subjectivity of the property it picks out. Few philosophers would now count shape properties, as opposed to concepts thereof, as subjective.

Concepts with a second type of subjectivity could more specifically be called "first-personal". A concept is first-personal if, in an account of its mastery, the application of the concept to objects other than the thinker is related to the conditions under which the thinker is willing to apply the concept to himself. Though there is considerable disagreement on how the account should be formulated, many theories of the concept of belief treat it as first-personal in this sense. For example, this is true of any account which says that a thinker understands a third-personal attribution "He believes that so-and-so" by understanding that it holds, very roughly indeed, if the third person in question is in circumstances in which the thinker would himself (first-person) judge that so-and-so. It is equally true of accounts which in one way or another say that the third-person attribution is understood as meaning that the other person is in some state which stands in some specified sameness relation to the state which causes the thinker to be willing to judge "*I believe that so-and-so*" (see OTHER MINDS).

The subjectivity of indexical concepts, such as *I*, *here*, *now*, and *that* (perceptually presented) *man*, has been widely noted. The last of these is subjective in the sense of the first criterion above; but they are all subjective in that the possibility of a subject's using any one of them to think about an object at a given time depends upon his relations to that particular object then. Indexicals are thus particularly well suited to expressing a particular point of view on the world of objects, a point of view available only to those who stand in the right relations to the objects in question.

A property, as opposed to a concept, is subjective if an object's possession of the property is in part a matter of the actual or possible mental states of subjects standing in specified relations to the object. Colour properties,

secondary qualities in general, moral properties, the property of propositions of being necessary or contingent, and the property of actions and mental states of being intelligible, have all been discussed as serious contenders for subjectivity in this sense. To say that a property is subjective is not to say that it can be analysed away in terms of mental states. The mental states in terms of which subjectivists have aimed to elucidate, say, the property of being red or the property of being kind have included the mental states of experiencing something as red, and judging something to be kind, respectively. These attitudes embed reference to the original properties themselves – or at least to concepts thereof – in a way which makes eliminative analysis problematic. The same plausibly applies to a subjectivist treatment of intelligibility: here the mental state would have to be that of finding something intelligible. Even without any commitment to eliminative analysis, though, the subjectivist's claim remains substantial. The subjectivist's claim needs extensive consideration for each of the diverse areas mentioned. In the case of colour, part of the task of the subjectivist who makes his claim at the level of properties rather than concepts is to argue against those who would identify the property of being red with a physical reflectance property, or with some more complex vector of physical properties.

Suppose that for an object to have a certain property is for subjects standing in a certain relation to it to be in a certain mental state. If subjects standing in that relation to it, and in that mental state, judge the object to have the property, their judgement will be true. Some subjectivists have been tempted to work this point into a criterion of a property's being subjective. There is, though, an issue here which is not definitional. *Prima facie*, it seems that we can make sense of this possibility: that though in certain circumstances, a subject's judgement about whether an object has a property are guaranteed to be correct, it is not his judgement (in those circumstances) or anything else about his or others' mental states which *makes* the judgement correct. To many philosophers, this will seem to be the actual situation for easily decided arithmetical

propositions such as  $3 + 3 = 6$ . If this is correct, the subjectivist will have to make essential use of some such asymmetrical notion as "what makes a proposition true", or "that in virtue of which a proposition is true". Conditionals or equivalences alone, not even a priori ones, will not capture the subjectivist character of the position.

Finally, subjectivity has been attributed to modes of understanding. Elaborating a mode of understanding can in large part be seen as elaboration of the conditions of mastery of mental concepts. For instance, those who believe that some form of imagination is involved in understanding third-person ascriptions of experiences will want to write this into the account of mastery of those attributions. However, some of those who attribute subjectivity to modes of understanding include in their conception the claim that some or all mental states are themselves subjective. This can be a claim about the mental properties themselves, rather than concepts thereof; but it is not charitable to interpret it as the assertion that mental properties involve mental properties. Rather, using the distinctions we already have, it can be read as the conjunction of these two propositions: that concepts of mental states are subjective in one of the sense given above, and that mental states can only be thought about by concepts which are thus subjective. Such a position need not be opposed to philosophical materialism, since it can allow for some version of this materialism for mental states. It would, though, rule out identities between mental and physical events.

See also CONCEPTS; LINGUISTIC UNDERSTANDING; OBJECTIVE/SUBJECTIVE; OBJECTIVITY.

## BIBLIOGRAPHY

- McDowell, J.: "Non-cognitivism and Rule-following," in *Wittgenstein: To Follow a Rule*, ed. S. Holtzman and C. Leich (London: Routledge, 1981), 141–62.
- McDowell, J.: "Values and Secondary Qualities," in *Morality and Objectivity: A Tribute to J. L. Mackie*, ed. T. Honderich (London: Routledge and Kegan Paul, 1985), 110–29.

- McGinn, C.: *The Subjective View* (Oxford: Clarendon Press, 1983).
- Nagel, T.: *The View from Nowhere* (New York: Oxford University Press, 1986).
- Smith, M., Lewis, D., Johnston, M.: Respective contributions to the symposium "Dispositional Theories of Value," *Proceedings of the Aristotelian Society*, supp. vol. 63 (1989), 89–111, 113–37, 139–74.
- Wiggins, D.: *Needs, Values, Truth* (Oxford: Blackwell, 1987).
- Williams, B.: *Descartes: The Project of Pure Enquiry* (Harmondsworth: Penguin, 1978); see especially sections on "the absolute conception."
- Wright, C.: "Moral Values, Projection and Secondary Qualities," *Proceedings of the Aristotelian Society*, supp. vol. 62 (1988), 1–26.

CHRISTOPHER PEACOCKE

**supervenience** see EPISTEMIC SUPERVENIENCE.

**surprise examination paradox** A teacher announces that there will be a surprise examination next week. A clever student argues that this is impossible. "The test cannot be on Friday, the last day of the week, because it wouldn't be a surprise. We would know the day of the test on Thursday evening. This means we can also rule out Thursday. For after we learn that no test has been given by Wednesday, we would know the test is on Thursday or Friday – and would already know that it is not on Friday by the previous reasoning. The remaining days can be eliminated in the same manner."

This puzzle has over a dozen variants. The first was probably invented by the Swedish mathematician Lennart Eklom in 1943. Although the first few commentators regarded the reverse elimination argument as cogent, every writer on the subject since 1950 agrees that the argument is unsound. The controversy has been over the proper diagnosis of the flaw.

Initial analyses of the student's argument tried to lay the blame on a simple equivocation. Their failure led to more sophisticated diagnoses. The general format has been an assimilation to better known paradoxes. One tradition casts the surprise examination paradox as a self-referential problem, as fundamentally akin to the Liar, the paradox

of the knower (*see* PARADOX OF THE KNOWER), or Gödel's incompleteness theorem. The original talk of a surprise is read as a reflexive claim about unprovability. That is, the teacher's announcement is paraphrased as:

- (A) There will be an examination next week but its date cannot be deduced from the conjunction of this announcement and the record of eliminated days.

If only one day remains, the announcement implies 'The test is on Friday but its date cannot be deduced from this statement'. This consequence does bear a provocative resemblance to the liar sentence "This statement is not true". Critics concede that this self-referential interpretation yields a liar-type paradox but complain that the interpretation strays too far from the meaning of the original announcement.

The second main tradition places the surprise examination paradox among epistemological puzzles such as Moore's paradox and the lottery paradox (*see* MOORE'S PARADOX; LOTTERY PARADOX). The connection with the epistemological puzzles is made by analysing the announcement as a prediction of student ignorance. This connection can be illustrated with the two-day version of the puzzle:

- (B) Either (the test is on Thursday and the students do not believe it beforehand) or (the test is on Friday and the students do not believe it beforehand).

Notice the resemblance between each disjunct and the sentence that fascinated G. E. Moore: "I went to the pictures last Tuesday, but I don't believe I did." Although Moore's sentence is consistent, the speaker cannot justifiably believe it. However, others can believe the proposition it expresses since it merely ascribes an error to the speaker. Indeed, the speaker is free to attribute past mistakes to himself or future ones: "On June 21, 1990, Nelson Mandela wore a Yankee baseball cap but by 2010 I will no longer believe so." Also note that the speaker can justifiably believe a disjunction of Moorean sentences: "Either Chaucer was right-handed and I do not believe it, or he

was left-handed and I do not believe it." So the students can believe (B) on the teacher's authority. But after they learn that the test is not on Thursday, the announcement implies the Moorean sentence "The test is on Friday but the students do not believe it." Since the students cannot believe this consequence, they can no longer believe (B). So if the teacher gives the test on the last day, the students will not justifiably expect it.

One might object that the students would expect the test because they know the teacher prefers to give an anticipated test rather than no test at all. This objection smuggles a new condition into the scenario. The original puzzle does not credit the students with this psychological insight into the teacher's desires. But let's consider the variant where they do know his preference. Here the students could indeed predict a last day examination. But the victory is hollow because the ability spoils the first step of the clever student's elimination argument. For the effect of the extra psychological premiss is to make the teacher more of an authority on whether there will be a test than on whether an event will be a surprise. So the students cannot appeal to the teacher's authority to rule out a last day examination.

A second objection starts with the observation that the students' parents could predict a last day examination. For the parents are not the surprisees. (The Moorean consequence is about the students' beliefs, not the parents'.) So if the parents can know, why can't they inform the students? No doubt, the parents could know and could tell the students that the second disjunct of (B) is true. But Moorean sentences cannot be made credible by increasing the authority of the source. God Himself cannot make you know "The test is on Friday but you do not believe it". A good memory or careful record-keeping will be equally impotent.

Other objections to the Moorean analysis turn on its completeness; can it handle all the variants of the paradox? If it can, then the surprise examination paradox teaches an important lesson: when we traffic in Moorean sentences, normal mechanisms of knowledge transfer breakdown.

*See also* LOTTERY PARADOX; MOORE'S PARADOX.

## SURPRISE EXAMINATION PARADOX

### BIBLIOGRAPHY

Binkley, R.: "The Surprise Examination in Modal Logic," *Journal of Philosophy* 65 (1968), 127–36.  
Kaplan, D. and Montague, R.: "A Paradox Regained," *Notre Dame Journal of Formal Logic* 1 (1960), 79–90.  
Olin, D.: "The Prediction Paradox Resolved," *Philosophical Studies* 44 (1983), 225–33.

Quine, W. V.: "On a So-called Paradox," *Mind* 62 (1953), 65–7.  
Sorensen, R.: *Blindspots* (Oxford: Clarendon Press, 1988).  
Wright, C. and Sudbury, A.: "The Paradox of the Unexpected Examination," *Australasian Journal of Philosophy* 11 (1977), 41–58.

ROY A. SORESENSEN



# T

---

**tabula rasa** The theory that the mind at birth is a tabula rasa (blank writing tablet) awaiting ideas from experience is usually associated with Locke's *Essay* (see LOCKE) – as by Leibniz (*Nouveaux essais*, 1704), who preferred to think of the mind as a block of veined marble, not wholly indifferent to what it receives. The phrase (from Latin translations of Aristotle's *De Anima*, 430a) is not there – except in Coste's French translation (1700) of Locke's "white paper" (*Essay*, 2.1.2); but the theory is. Taking from the Scholastics the associated thought that "there is nothing in the intellect which was not first in the senses", Gassendi and Locke made it a cornerstone of their empiricism (see EMPIRICISM).

See also STOIC EPISTEMOLOGY.

## BIBLIOGRAPHY

Aaron, R. I.: *John Locke* (Oxford: Clarendon Press, 1971), 34–5, 114.

R. S. WOOLHOUSE

**testimony** One of the central questions in the epistemology of testimony is how, precisely, hearers acquire justified or warranted beliefs from the testimony of speakers. Answers to this question have traditionally fallen into one of two camps: *non-reductionism* or *reductionism*. According to non-reductionists – whose historical roots are standardly traced back to the work of Thomas Reid (1764) – it is a *basic* source of justification or warrant, on an epistemic par with sense perception, memory, inference, and the like. Given this, non-reductionists maintain that, so as long as there are no relevant undefeated

defeaters, hearers can be justified or warranted in accepting what they are told *merely* on the basis of the testimony of speakers.

There are two central kinds of defeaters that are typically taken to be relevant here. First, there are what we might call *psychological defeaters*. A psychological defeater is a doubt or belief that is had by S that indicates that S's belief that *p* is either false or unreliably formed or sustained. Defeaters in this sense function by virtue of being *had* by S, regardless of their truth-value or epistemic status. Second, there are what we might call *normative defeaters*. A normative defeater is a doubt or belief that S ought to have that indicates that S's belief that *p* is either false or unreliably formed or sustained. Defeaters in this sense function by virtue of being doubts or beliefs that S *should have* (whether or not S does have them) given the presence of certain available evidence. The underlying thought here is that certain kinds of doubts and beliefs contribute epistemically unacceptable *irrationality* to doxastic systems; and, accordingly, justification and warrant can be defeated or undermined by their presence.

Moreover, a defeater may itself be either defeated or undefeated. Suppose, for instance, that Harold believes that there is a bobcat in his backyard because he saw it there this morning, but Rosemary tells him, and he thereby comes to believe, that the animal is instead a lynx. Now, the justification or warrant that Harold had for believing that there is a bobcat in his backyard has been defeated by the belief that he acquires on the basis of Rosemary's testimony. But, since psychological defeaters can themselves be beliefs, they, too, are candidates for defeat. For instance, suppose that Harold consults a North American wildlife book and discovers that the white tip of the animal's tail confirms that it was

indeed a bobcat, thereby providing him with a *defeater-defeater* for his original belief that there is a bobcat in his backyard. And, as should be suspected, defeater-defeaters can also be defeated by further doubts and beliefs, which, in turn, can be defeated by further doubts and beliefs, and so on. Similar considerations involving reasons, rather than doubts and beliefs, apply in the case of normative defeaters. When one has a defeater for one's belief that *p* that is not itself defeated, one has what is called an *undefeated defeater* for one's belief that *p*. It is the presence of undefeated defeaters, not merely defeaters, that is incompatible with testimonial justification or warrant.

In contrast to non-reductionism, reductionists – whose historical roots are typically traced back to the work of David Hume (1748) – believe that, in addition to the absence of undefeated defeaters, hearers must also possess *non-testimonially based positive reasons* order to be justified or warranted in accepting the testimony of speakers. These reasons are typically the result of induction: for instance, hearers observe a general conformity between facts and reports and, with the assistance of memory and reason, they inductively infer that certain speakers, contexts, or types of reports are reliable sources of information. In this way, the justification or warrant of testimony is *reduced* to the justification or warrant for sense perception, memory, and inductive inference.

There are two different versions of reductionism. According to *global reductionism*, the justification or warrant of *testimony as a source of belief* reduces to the justification or warrant of sense perception, memory, and inductive inference. Thus, in order to be justified or warranted in accepting the testimony of speakers, hearers must possess non-testimonially based positive reasons for believing that *testimony is generally reliable*. According to local reductionism, which is the more widely accepted of the two versions, the justification or warrant of *each instance of testimony* reduces to the justification or warrant of instances of sense perception, memory, and inductive inference. So, in order to be justified or warranted in accepting the testimony of

speakers, hearers must have non-testimonially based positive reasons for accepting *the particular report in question*.

Both non-reductionism and reductionism have been subject to various objections – objections that opponents use to motivate their own preferred views. The central problem raised against non-reductionism is that it is said to sanction gullibility, epistemic irrationality, and intellectual irresponsibility (see Fricker, 1994; Faulkner, 2000; and Lackey, 2006). For, given that hearers can acquire testimonially justified or warranted beliefs in the complete absence of any relevant positive reasons on such a view, arbitrarily chosen speakers and books, random postings on the Internet, and unidentified telemarketers can be trusted, so long as there is no negative evidence against such sources. Yet surely the opponent of non-reductionism urges, accepting testimony in these kinds of cases is a paradigm of epistemic irrationality and irresponsibility.

Against reductionism, two central problems are typically raised. The first is that young children clearly acquire a great deal of knowledge from their parents and teachers, and it is said to be doubtful that they possess – or even could possess – non-testimonially based positive reasons for accepting much of what they are told (see Audi, 1997). For instance, an 18-month-old baby may come to know that the stove is hot from the testimony of her mother, but it is unclear whether she has the cognitive sophistication to have reasons for believing her mother to be a reliable source of information, let alone for believing that testimony is generally reliable. Given this, reductionists – of both the global and the local stripes – may be hard-pressed to explain how such young subjects could acquire all of the testimonial knowledge they at least seem to possess. The second problem that is said to face reductionism is that most ordinary cognitive agents do not seem to have enough information to possess relevant positive reasons in all of those cases where testimonial knowledge appears present. For instance, targeting the global version of this view, it is argued that most people have been exposed only to a very limited range of

reports from speakers in their native language in a handful of communities in their native country. This limited sample of reports provides only a fraction of what would be required to conclude legitimately that testimony is *generally* reliable (see Coady, 1992; and Lackey, 2006). Against the local version of reductionism, it is argued that most cognitive agents frequently acquire testimonial knowledge from speakers about whom they know very little. For instance, upon arriving in Chicago for a new vacation, I may receive accurate directions to Navy Pier from the first passerby I see. Most agree that such a transaction can result in my acquiring testimonial knowledge of Navy Pier's whereabouts, despite the fact that my positive reasons for accepting the directions in question – if, indeed, I possess any – are scanty at best (see Webb, 1993; Foley, 1994; Strawson, 1994; and Schmitt, 1999).

The direction that some recent work on testimony has taken is to avoid the problems afflicting non-reductionism and reductionism by developing qualified or hybrid views of either of these views. For instance, in an effort to avoid the charges of gullibility and epistemic irresponsibility, some non-reductionists emphasize that hearers must be “epistemically entitled” to rely on the testimony of speakers or that they need to “monitor” incoming reports, even though such requirements do not quite amount to the full-blown need for non-testimonially based positive reasons embraced by reductionists. (See Goldberg, 2006, and forthcoming, for these qualifications to a non-reductionist view.) And some reductionists, trying to account for the testimonial knowledge of both young children and those hearers who possess very little information about their relevant speakers, argue that positive reasons are not needed during either the “developmental phase” of a person's life – which is when a subject is acquiring concepts and learning the language, relying in large part on her parents and teachers to guide the formation of her belief system – or when hearers are confronted with “mundane testimony” – which is testimony about, for instance, a speaker's name, what she had for breakfast, the time of day, and

so on. (See Fricker, 1994 and 1995, for these modifications to reductionism.) On this view of reductionism, then, while positive reasons remain a condition of testimonial justification, such a requirement applies only to hearers in the “mature phase” of their life who are encountering “non-mundane testimony”. Such qualified or hybrid views of both non-reductionism and reductionism often encounter either variations of the very same problems that led to their development, or altogether new objections. (For these objections, see Insole, 2000; Weiner, 2003; and Lackey, forthcoming a.)

In addition to non-reductionism, reductionism, and hybrids thereof, an alternative family of views has been growing in popularity in recent work in the epistemology of testimony – one that provides a radically different answer to the question of how testimonial beliefs are justified or warranted. Though there are some points of disagreement among some of the members of this family, they are united in their commitment to at least three central theses. First, and perhaps most important, the *interpersonal relationship* between the two parties in a testimonial exchange should be a central focus of the epistemology of testimony. Second, and closely related, certain features of this interpersonal relationship – such as the speaker *offering her assurance* to the hearer that her testimony is true, or the speaker *inviting the hearer to trust her* – are (at least sometimes) actually *responsible for conferring epistemic value* on the testimonial beliefs acquired. Third, the epistemic justification or warrant provided by these features of a testimonial exchange is *non-evidential* in nature. For ease of discussion, I shall call the general conception of testimony characterized by these theses the *Interpersonal View of Testimony* (hereafter the IVT). (Proponents of the IVT include Ross, 1986; Hinchman, 2005; and Moran, 2006.)

One of the central motivations for the IVT is a perceived failure on the part of existing views of testimony – particularly those that regard a speaker's testimony that *p* as merely *evidence* for a hearer to believe that *p* – adequately to account for the import of the

interpersonal relationship between the speaker and the hearer in a testimonial exchange. According to proponents of the IVT, speakers should be regarded as agents who enter into interpersonal relationships with their hearers. For instance, according to Moran's version of the IVT, a speaker's testimony that *p* is understood as the speaker giving her *assurance* that *p* is true. Since assurance can be given only when it is freely presented as such, Moran claims that a speaker freely assumes responsibility for the truth of *p* when she asserts that *p*, thereby providing the hearer with an *additional* reason to believe that *p*, different in kind from anything given by evidence alone. In a similar spirit, Hinchman argues that when a speaker tells her hearer that *p* she acts on an intention to give him an entitlement to believe that *p* that derives, not from evidence of the truth of "*p*", but merely from his understanding of the act she performs.

A central objection facing proponents of the IVT is that the interpersonal features that lie at the heart of their views are not clearly epistemologically relevant. For instance, the mere fact that a speaker offers her assurance to a hearer does not affect the reliability, proper functioning, truth-tracking, evidential relations, or any other relevant truth-conducive feature of the testimony in question. This leads to the following dilemma: either the IVT is genuinely interpersonal but epistemologically impotent, or it is not epistemologically impotent but neither is it genuinely interpersonal. Either way, there is a worry that the IVT fails to provide a compelling alternative to existing theories in the epistemology of testimony. (See Lackey, forthcoming a, for this objection to the IVT.)

A further debate at center stage in the epistemology of testimony focuses on a widely accepted thesis, which we may call the *Transmission of Epistemic Properties* (hereafter TEP). The basic thought expressed by TEP is that a testimonial exchange involves a speaker's belief, along with the epistemic properties it possesses, being *transmitted* to a hearer. There are two dimensions to TEP; one is a necessity thesis and the other is a sufficiency thesis. More precisely:

TEP-N: For every speaker, A, and hearer, B, B knows (believes with justification/warrant) that *p* on the basis of A's testimony that *p* only if A knows (believes with justification/warrant) that *p*.

TEP-S: For every speaker, A, and hearer, B, if (1) A knows (believes with justification/warrant) that *p*, (2) B comes to believe that *p* on the basis of the content of A's testimony that *p*, and (3) B has no undefeated defeaters for believing that *p*, then B knows (believes with justification/warrant) that *p*.

Support for this view, particularly for TEP-N, derives from a purported analogy between testimony and memory. While memory is thought to be capable of only *preserving* epistemic properties from one time to another – and cannot therefore *generate* new epistemic properties – testimony is said to be capable of only *transmitting* epistemic properties from one person to another. So, for instance, just as I cannot know that *p* on the basis of memory unless I non-memorially knew that *p* at an earlier time, the thought underlying this picture of testimonial knowledge is that a hearer cannot know that *p* on the basis of testimony unless the speaker from whom it was acquired herself knows that *p*. Similarly, just as my knowing that *p* at an earlier time may be sufficient, in the absence of undefeated defeaters, for me memorially to know that *p* now, it is said that a speaker's knowing that *p* may also be sufficient, in the absence of undefeated defeaters, for a hearer to know that *p* on the basis of her testimony.

Recently, however, objections have been raised to both dimensions of TEP, thereby calling into question the widely accepted view that transmission lies at the heart of the epistemology of testimony. There are two general types of counter-examples that have been raised to TEP-N. The first type involves speakers who fail to believe, and hence to know, a proposition to which they are testifying, but nevertheless reliably convey the information in question through their testimony. So, for instance, a devout creationist who does not

believe in the truth of evolutionary theory may still be able reliably to teach to her students that *Homo sapiens* evolved from *Homo erectus*, thereby imparting knowledge (justified belief, warranted belief) to her students that she fails to possess herself (see Lackey, 1999 and forthcoming a). The second type of counter-example that has been raised to TEP-N involves speakers who have an undefeated defeater for believing a proposition to which they are testifying, but nevertheless reliably convey such a proposition through their testimony without transmitting the defeater in question to their hearers. For instance, a speaker in the grips of skeptical worries may not be properly described as knowing that there is a café around the corner, but she may still reliably communicate this information to a hearer who does not have such doubts, thereby conveying knowledge (justified belief, warranted belief) that she does not have herself (see Lackey, 1999 and forthcoming a). Both cases, if successful, show that TEP-N is false.

There are also two general types of counter-examples that have been raised to TEP-S. The first type of case shows that, for reasons having to do specifically with the *hearer*, a hearer's belief may fail to be known (justified, warranted) even though the hearer has no relevant undefeated defeaters, the speaker from whom it was acquired has the knowledge (justified belief, warranted belief) in question, and the speaker testifies sincerely. For instance, a compulsively trusting hearer – who accepts whatever she is told, regardless of the amount or kind of evidence there is to the contrary – may be so constituted as to prevent the epistemic properties of a speaker's belief from being passed to her, even when there in fact are no relevant defeaters present (see Lackey, forthcoming a and forthcoming b). The second type of counter-example to TEP-S shows that, for reasons having to do specifically with the *speaker*, a hearer's belief may fail to be known (justified, warranted) even though the hearer has no relevant undefeated defeaters, the speaker from whom it was acquired has the knowledge (justified belief, warranted belief) in question, and the speaker testifies sincerely. For instance, a speaker may in fact know that there was a

badger in the park this morning because she saw one there, but she may be such that she would have reported to her hearer that there was such a badger even if there hadn't been one. In such a case, the speaker's belief may possess all of the epistemic properties in question – e.g. justification, warrant, etc. – and yet the belief that the hearer forms on the basis of her testimony may possess none of them (see Lackey, forthcoming a and forthcoming b). Once again, to the extent that such counter-examples succeed, they show that TEP-S is false.

One of the central conclusions that the above considerations motivate is the replacement of TEP with conditions focusing on the *statements* of speakers rather than on their states of believing or knowing. More precisely, TEP may be replaced with the following *Statement View of Testimony* (SVT):

SVT: For every speaker, A, and hearer, B, B knows (believes with justification/warrant) that *p* on the basis of A's testimony that *p* only if (1) A's statement that *p* is reliable or otherwise truth-conducive, (2) B comes to believe that *p* on the basis of the content of A's statement that *p*, and (3) B has no undefeated defeaters for believing that *p* (see Lackey, forthcoming a and forthcoming b).

Besides adding a necessary condition requiring the truth of the belief in question when knowledge is at issue, further conditions may be needed for a complete view of testimonial knowledge (justification, warrant), such as the need for positive reasons embraced by reductionists. But, regardless of what is added to SVT, such a view reveals that testimony is not merely a transmissive epistemic source, as has been traditionally assumed, but that it can instead generate epistemic features in its own right.

#### BIBLIOGRAPHY

- Audi, R.: "The Place of Testimony in the Fabric of Knowledge and Justification," *American Philosophical Quarterly* 34 (1997), 405–22.



- Coady, C. A. J.: *Testimony: A Philosophical Study* (Oxford Clarendon Press, 1992).
- Faulkner, P.: "The Social Character of Testimonial Knowledge," *The Journal of Philosophy* 97 (2000), 581–601.
- Foley, R.: "Egoism in Epistemology," in *Socializing Epistemology: The Social Dimensions of Knowledge*, ed. F. F. Schmitt (Lanham, MD: Rowman and Littlefield, 1994), 53–73.
- Fricker, E.: "Against Gullibility," in *Knowing from Words*, ed. B. K. Matilal and A. Chakrabarti (Dordrecht: Kluwer, 1994), 125–61.
- Fricker, E.: "Telling and Trusting Reductionism and Anti-Reductionism in the Epistemology of Testimony," *Mind* 104 (1995), 393–411.
- Goldberg, S.: "Reductionism and the Distinctiveness of Testimonial Knowledge," in *The Epistemology of Testimony*, ed. J. Lackey and E. Sosa (Oxford: Oxford University Press, 2006), 127–44.
- Goldberg, S.: "Testimonial Knowledge in Early Childhood, Revisited," *Philosophy and Phenomenological Research*, forthcoming.
- Hinchman, E. S.: "Telling as Inviting to Trust," *Philosophy and Phenomenological Research* 70 (2005), 562–87.
- Hume, D.: *An Inquiry Concerning Human Understanding* (1748), ed. L. A. Selby-Bigge, revised P. H. Nidditch (Oxford: Oxford University Press, 1975).
- Insole, C. J.: "Off the Local Threat to Irreducible Knowledge by Testimony," *The Philosophical Quarterly* 50 (2000), 44–56.
- Lackey, J.: "Testimonial Knowledge and Transmission," *The Philosophical Quarterly* 49 (1999), 471–90.
- Lackey, J.: "It Takes Two to Tango: Beyond Reductionism and Non-reductionism in the Epistemology of Testimony," in *The Epistemology of Testimony*, ed. J. Lackey and E. Sosa (Oxford: Oxford University Press, 2006), 160–89.
- Lackey, J.: *Learning from Words* (Oxford: Oxford University Press, forthcoming a).
- Lackey, J.: "Learning from Words," *Philosophy and Phenomenological Research* (forthcoming b).
- Moran, R.: "Getting Told and Being Believed," in *The Epistemology of Testimony*, ed. J. Lackey and E. Sosa (Oxford: Oxford University Press, 2006), 272–306.
- Reid, T.: *An Inquiry into the Human Mind*, in his *Inquiry and Essays* (1764), ed. R. E. Beanblossom and K. Lehrer (Indianapolis, IN: Hackett, 1983).
- Ross, A.: "Why Do We Believe What We Are Told?," *Ratio* 28 (1986), 69–88.
- Schmitt, F.: "Social Epistemology," in *The Blackwell Guide to Epistemology*, ed. J. Greco and E. Sosa (Oxford: Blackwell Publishers, 1999), 354–82.
- Strawson, P. F.: "Knowing from Words," in *Knowing from Words*, ed. B. Matilal and A. Chakrabarti (Dordrecht: Kluwer, 1994), 23–7.
- Webb, M. O.: "I Know about as Much as You: A Reply to Hardwig," *The Journal of Philosophy* 110 (1993), 260–70.
- Weiner, M.: "Accepting Testimony," *The Philosophical Quarterly* 53 (2003), 256–64.

JENNIFER LACKEY

**theory** In philosophy of science, a generalization or set of generalizations purportedly making reference to unobservable entities, e.g. atoms, genes, quarks, unconscious wishes. The ideal gas law, for example, refers only to such observables as pressure, temperature, and volume; the molecular-kinetic theory refers to molecules and their properties. Although an older usage suggests lack of adequate evidence in support thereof ("merely a theory"), current philosophical usage does *not* carry that connotation. Einstein's special theory of relativity, for example, is considered extremely well founded.

There are two main views on the nature of theories. According to the "received view", theories are partially interpreted axiomatic systems; according to the semantic view, a theory is a collection of models (Suppe, 1974).

See also AXIOMATIZATION, AXIOMATICS.

## BIBLIOGRAPHY

- Colodny, R. G., ed.: *The Nature and Function of Scientific Theories* (Pittsburgh, PA: University of Pittsburgh Press, 1970).
- Suppe, F., ed.: *The Structure of Scientific Theories* (Urbana, IL: University of Illinois Press, 1974).

WESLEY C. SALMON

**transcendental arguments** These were brought into philosophical prominence by the work of Kant, who used them to establish the role of certain fundamental a priori concepts in our knowledge of the external world (see KANT). The role which transcendental arguments show these concepts to play is,

Kant claims, that they make experience possible, and that therefore sceptical doubts about our possession of them are empty. It is the scepticism-refuting promise of transcendental arguments that has provoked most interest among subsequent philosophers.

Kant's central question concerns the possibility of synthetic a priori knowledge. An answer to it presupposes an answer to the more general question, what are the necessarily presupposed grounds of experience in general? In arguing for the categories, and for space and time as pure forms of sensibility, Kant is arguing that there can be experience only under certain conditions, and this strategy displays the characteristic form of a transcendental argument. Simply in respect of its form, the argument is that there must be something Y if there is something X of which Y is a necessary condition. In crude outline, therefore, the transcendental strategy consists in the search for key necessary conditions of some given region of discourse or experience.

Reasoning of this character is not unique to Kant, nor do all the varieties of it found in the philosophical literature model themselves exactly on Kant's procedure. Wittgenstein (see WITTGENSTEIN) in the *Philosophical Investigations* and *On Certainty* argues transcendently about the impossibility of private language and the possibility of knowledge respectively. J. L. Austin (see AUSTIN) argues transcendently in formulating a theory of truth by distinguishing between demonstrative and descriptive conventions in language, his point being that some such distinction is required for a certain other concept – that of truth as correspondence – to have application. A different example is supplied by Gilbert Ryle in his use of “polar concept” arguments (see RYLE). The sceptical suggestion that we might undetectably be in error on any given occasion is refuted, Ryle claims, by the fact that just as we cannot have counterfeit coins unless there are genuine ones, so we cannot have a concept of error unless we have the concept of being right, and therefore we must sometimes know we are right. Once again, the strategy is to argue that X cannot be the case unless Y is the case; and since X is indeed the case, Y must be the case also.

These examples show that transcendental-style arguments are commonplace in philosophy. But there is a marked difference between reasonings of these familiar sorts and the weightier, more ambitious transcendental arguments found in Kant and others. One way to mark the distinction is to say that the familiar kinds of argument just sketched share their methodology with transcendental arguments proper, but not their scope; for the latter are not concerned merely with local conceptual polarities like “illusory-veridical”, but with the conditions under which highly general concepts – the concept of other minds, or the concept of an independently existing reality – have application, given that such concepts play a key role in entire discourses and are central to our experience as a whole. So put, the aim of transcendental arguments is to establish the conditions necessary for experience, or experience of a certain kind, in general; and, at their most controversial, to establish conclusions about the nature and existence of an external world, or other minds, derived from paying attention to what has to be the case for there to be experience, or for experience to be as it is.

Classic contemporary examples of transcendental arguments are to be found in P. F. Strawson's *Individuals* (see STRAWSON). One of them is aimed at establishing the vacuity of sceptical doubt about the existence of other minds, and proceeds as follows. One can ascribe states of consciousness to oneself only if one is able to ascribe them to others, Strawson argues, because to doubt the existence of other minds one must employ the concept of other minds itself, and this can only be done if one can distinguish between “my states of consciousness” and “others' states of consciousness”. But this can only be done, in turn, if others exist, because the identification of conscious states can only be effected by reference to particulars of a certain kind, namely, persons; the concept of which – in turn again – demands that there be criteria for distinguishing one person from another, for otherwise the identification of states of consciousness would be impossible. So one can talk of “my experiences” only if one can talk of “others' experiences”; this is

possible only if there are criteria for distinguishing between persons; and since one can indeed talk significantly of one's own experiences, such criteria must exist. But if they exist they constitute logically adequate grounds for ascription of states of consciousness to others. Hence sceptical doubts about the existence of other minds are vacuous, because the sceptic cannot so much as formulate them without employing the discourse whose very conditions of employment legitimize what he wishes to question.

Another transcendental argument in Strawson's *Individuals* concerns the perception-independent existence of material particulars, and it has been much discussed as a result of Barry Stroud's criticisms of it. The subsequent debate can be said to turn on the following points.

One of the crucial questions about transcendental arguments concerns what they might hope to establish. The options, simply put, are that either they establish the existence of something (an external world, other minds), or they establish that certain concepts are necessary to our conceptual scheme. Clearly these are quite different results, and the latter involves the further problem of whether our conceptual scheme is the only possible one, for if not the terminus of a transcendental argument is strictly relative.

To grasp the difference between these options it is useful to consider them as responses to sceptical challenges over our belief in the existence of an external world. On the first option, call it option A, the aim is to establish the existence of the external world. On the second, call it option B, the aim is to show that we must believe in the existence of an external world as a condition of the coherence of experience. To settle the sceptic's doubts outright under B one might have further to show that our conceptual scheme is the only conceivable one.

The chief difficulty faced by option A is that even if one could show that it is a necessary condition of our having coherent experience that we possess and apply a concept of an independently existing world, it still needs to be shown that something "out there" answers to that concept; in other

words, that it is a necessary condition of our having the concept of an external world that an external world exists. It is one thing to argue that we must have and employ concepts of space, time, causality and particulars conditioned by them, and another to show that there exist things corresponding to these concepts and existentially prior to their use.

The difficulty here is that it does not look possible to argue from the fact that there is experience, or from some richer premise characterising that experience, to talk about the way things are independently of experience, without either supplementing the argument with additional factual premisses, or arguing that it is somehow constitutive of the concept-introducing terms employed that they have empirical conditions of application under which, and only under which, they can be known to be truly applied.

Neither move looks promising. Additional factual premisses will be unacceptable from the sceptic's point of view because precisely the same doubts about entitlement to them will arise as with the belief in the existence of the external world itself. And a verification principle, if it were not anyway highly debatable in itself, would make the apparatus of transcendental arguments unnecessary, because one could rely on the verification principle alone to overcome scepticism (see VERIFICATIONISM). Even in one were to weaken a verification principle to a general "principle of significance" stating that there can only be legitimate or meaningful employment of concepts if they are governed by experiential conditions of application, it would still be a question whether the satisfaction of such conditions guaranteed that a concept succeeded in applying to something existing independently of its use, the reason being that a verification principle can be satisfied even in an ideal universe in which systematic criteria exist for distinguishing between what counts as "objective" and "subjective" among a mind's experiences – perhaps, on Hume's lines (see HUME), by counting as objective all those ideas which are "forceful" and "vivid", and as subjective all those which are otherwise.

Option B transcendental arguments look, by comparison, like relatively mundane affairs.

There is no special difficulty in exploring the necessary conditions of the experience we enjoy; we are, at very least, entitled to claim attention for investigations into the role of our concepts of objects, space, time and causality, given that it at least seems to us that ours is a world of causally interactive spatiotemporal things. If we have a concept of objects, we must have a concept of the continued unperceived existence of objects, because this is necessary to the concept of a single spatio-temporal world; and it is precisely a belief in the continued unperceived existence of objects that the sceptic asks us to justify. Showing that we must have such a belief as a condition of experience is not the same as proving that such objects exist. One is stating what we must believe, not how things are; but since the sceptic wishes us to justify the belief, doing so – the argument goes – is enough to put an end to scepticism.

The sceptic, however, can in response shift his attack to a more general level, by arguing that there might be a certain parochial interest in pointing out what a given form of experience requires as its fundamental concepts, but that this does not settle doubts about the general validity of those concepts. What if there are forms of experience which are non-spatial, or non-causal, or in which there is no need to distinguish experience from its objects? To counter scepticism at this more general level B-type transcendental arguments require supplementation by antirealist arguments, themselves a form of transcendental argument designed to show that the concepts required for our experience are the concepts required for any experience. That, clearly, is in its different way as ambitious a project as option A transcendental arguments themselves.

The primary importance of transcendental arguments resides in the fact that we have to reflect on our concepts and beliefs from the internal perspective of having to use them even as we investigate them. There is no external point from which we can view our conceptual scheme; like Neurath's sailors we have to rebuild our ship at sea. Transcendental arguments offer ways of reflecting on our concepts which at least promise to tell us much about their nature and interrelations,

and about the degree of strength possessed by sceptical challenges to our use of them.

See also SCEPTICISM, CONTEMPORARY; VERIFICATIONISM.

#### BIBLIOGRAPHY

- Davidson, D.: "On the Very Idea of a Conceptual Scheme," in his *Inquiries into Truth and Interpretation* (Oxford: Oxford University Press, 1984), 183–98.
- Grayling, A. C.: *The Refutation of Scepticism* (London: Duckworth, 1985).
- Griffiths, A. Phillips and Macintosh, J. J.: "Transcendental Arguments," *Proceedings of the Aristotelian Society* supp. vol. 43 (1969), 165–180.
- Harrison, R.: *On What There Must Be* (Oxford: Clarendon Press, 1974).
- Kant, I.: *Critique of Pure Reason* (1781), trans. N. Kemp Smith (London: Macmillan, 1964).
- Rosenberg, T.: "Transcendental Arguments Revisited," *Journal of Philosophy* 25 (1975), 611–24.
- Schaper, E.: "Arguing Transcendentially," *Kant-Studien* 63 (1972), 101–16.
- Strawson, P. F.: *Individuals* (London: Methuen, 1959).
- Strawson, P. F.: *The Bounds of Sense* (London: Methuen, 1965).
- Stroud, B.: "Transcendental Arguments," *Journal of Philosophy* 65 (1968), 241–56.
- Walker, R. C. S.: *Kant* (London: Routledge and Kegan Paul, 1978).
- Wilkerson, T. E.: "Transcendental Arguments," *Philosophical Quarterly* 20 (1970), 200–12.

A. C. GRAYLING

**tripartite definition of knowledge** The tripartite definition of knowledge states that the propositional knowledge, i.e. knowledge that *p*, has three individually necessary and jointly sufficient conditions: justification, truth and belief (see PROPOSITIONAL KNOWLEDGE). In short, propositional knowledge is justified true belief. The belief condition requires that anyone who knows that *p* believe that *p*. The truth condition requires that any known proposition be true. And the justification condition requires that any known proposition be adequately justified, warranted or evidentially supported. Plato

(see PLATO) appears to be considering the tripartite definition in the *Theaetetus* (201c–202d), and to be endorsing its jointly sufficient conditions for knowledge in the *Meno* (97e–98a). This definition has come to be called ‘the standard analysis’ of knowledge, and has received a serious challenge from Edmund Gettier’s counterexamples in 1963.

See GETTIER PROBLEM; KNOWLEDGE AND BELIEF; PROPOSITIONAL KNOWLEDGE.

#### BIBLIOGRAPHY

- Gettier, E. L.: “Is Justified True Belief Knowledge?,” *Analysis* 23 (1963), 121–3.  
 Moser, P. K.: *Knowledge and Evidence* (Cambridge: Cambridge University Press, 1989).  
 Moser, P. K. and vander Nat, A., eds: *Human Knowledge: Classical and Contemporary Approaches* (New York: Oxford University Press, 1987).  
 Pappas, G. S. and Swain, M., eds: *Essays on Knowledge and Justification* (Ithaca, NY: Cornell University Press, 1978).  
 Shope, R. K.: *The Analysis of Knowing* (Princeton, NJ: Princeton University Press, 1983).

PAUL K. MOSER

**truth, theories of** The notion of truth occurs with remarkable frequency in our reflections on language, thought and action. We are inclined to suppose, for example, that *truth* is the proper aim of scientific inquiry, that *true* beliefs help us to achieve our goals, that to understand a sentence is to know which circumstances would make it *true*, that reliable preservation of *truth* as one argues from premises to a conclusion is the mark of valid reasoning, that moral pronouncements should not be regarded as objectively true, and so on. In order to assess the plausibility of such theses, and in order to refine them and to explain why they hold (if they do), we require some view of what truth is – a theory that would account for its properties and its relations to other matters. Thus there can be little prospect of understanding our most important faculties in the absence of a good theory of truth.

Such a thing, however, has been notoriously elusive. The ancient idea that truth is some sort of “correspondence with reality” has still never been articulated satisfactorily: the nature of the alleged “correspondence” and the alleged “reality” remain objectionably obscure. Yet the familiar alternative suggestions – that true beliefs are those that are “mutually coherent”, or “pragmatically useful”, or “verifiable in suitable conditions” – have each been confronted with persuasive counterexamples. A twentieth-century departure from these traditional analyses is the view that truth is not a *property* at all – that the syntactic form of the predicate, “is true”, distorts its real semantic character, which is not to *describe* propositions but to *endorse* them. But this radical approach is also faced with difficulties and suggests, somewhat counter-intuitively, that truth cannot have the vital theoretical role in semantics, epistemology and elsewhere that we are naturally inclined to give it. Thus truth threatens to remain one of the most enigmatic of notions: an explicit account of it can appear to be essential yet beyond our reach. However, recent work provides some grounds for optimism.

#### TRADITIONAL THEORIES

The belief that snow is white owes its truth to a certain feature of the external world: namely, to the fact that snow is white. Similarly, the belief that dogs bark is true because of the fact that dogs bark. This sort of trivial observation leads to what is perhaps the most natural and popular account of truth, the *correspondence theory*, according to which a belief (statement, sentence, proposition, etc.) is true just in case there exists a fact corresponding to it (Wittgenstein, 1922; Austin, 1950). This thesis is unexceptionable in itself. However if it is to provide a rigorous, substantial and complete theory of truth – if it is to be more than merely a picturesque way of asserting all equivalences of the form:

The belief *that p* is true  $\leftrightarrow p$



then it must be supplemented with accounts of what *facts* are, and what it is for a belief to *correspond* to a fact; and these are the problems on which the correspondence theory of truth has foundered. For one thing, it is far from clear that any significant gain in understanding is achieved by reducing “the belief that snow is white is true” to “the fact that snow is white exists”; for these expressions seem equally resistant to analysis and too close in meaning for one to provide an illuminating account of the other. In addition, the general relationship that holds in particular between the belief that snow is white and the fact that snow is white, between the belief that dogs bark and the fact that dogs bark, and so on, is very hard to identify. The best attempt to date is Wittgenstein’s (1922) so-called “picture theory” (see WITTGENSTEIN), whereby an elementary proposition is a configuration of terms, an atomic fact is a configuration of simple objects, an atomic fact *corresponds* to an elementary proposition (and makes it true) when their configurations are identical and when the terms in the proposition refer to the similarly placed objects in the fact, and the truth value of each complex proposition is entailed by the truth values of the elementary ones. However, even if this account is correct as far as it goes, it would need to be completed with plausible theories of “logical configuration”, “elementary proposition”, “reference” and “entailment”, none of which is easy to come by.

A central characteristic of truth – one that any adequate theory must explain – is that when a proposition satisfies its “conditions of proof (or verification)” it is regarded as true. To the extent that the property of *corresponding with reality* is mysterious, we are going to find it impossible to see why what we take to verify a proposition should indicate the possession of that property. Therefore a tempting alternative to the correspondence theory – an alternative which eschews obscure, metaphysical concepts and which explains quite straightforwardly why verifiability implies truth – is simply to *identify* truth with verifiability (Peirce, 1932). This idea can take on various forms. One version involves the further assumption that verification is holistic

– i.e. that a belief is justified (i.e. verified) when it is part of an entire system of beliefs that is consistent and “harmonious” (Bradley, 1914; Hempel, 1935) (see HOLISM). This is known as the *coherence theory* of truth. Another version involves the assumption that there is, associated with each proposition, some specific procedure for finding out whether one should believe it or not. On this account, to say that a proposition is true is to say that it would be verified by the appropriate procedure (Dummett, 1978; Putnam, 1981). In the context of mathematics this amounts to the identification of truth with provability.

The attractions of the verificationist account of truth are that it is refreshingly clear compared with the correspondence theory, and that it succeeds in connecting truth with verification (see VERIFICATIONISM). The trouble is that the bond it postulates between these notions is implausibly strong. We do indeed take verification to indicate truth. But also we recognize the possibility that a proposition may be false in spite of there being impeccable reasons to believe it, and that a proposition may be true even though we aren’t able to discover that it is. Verifiability and truth are no doubt highly correlated; but surely not the same thing.

A third well-known account of truth is known as pragmatism (James, 1909; Papineau, 1987) (see PRAGMATISM). As we have just seen, the verificationist selects a prominent property of truth and considers it to be the essence of truth. Similarly the pragmatist focuses on another important characteristic – namely, that true beliefs are a good basis for action – and takes this to be the very nature of truth. True assumptions are said to be, by definition, those which provoke actions with desirable results. Again we have an account with a single attractive explanatory feature. But again the central objection is that the relationship it postulates between truth and its alleged analysans – in this case, utility – is implausibly close. Granted, true beliefs tends to foster success. But it happens regularly that actions based on true beliefs lead to disaster, while false assumptions, by pure chance, produce wonderful results.

One of the few uncontroversial facts about truth is that the proposition that snow is white is true if and only if snow is white, the proposition that lying is wrong is true if and only if lying is wrong, and so on. Traditional theories acknowledge this fact but regard it as insufficient and, as we have seen, inflate it with some further principle of the form "X is true if and only if X has property P" (such as corresponding to reality, verifiability, or being suitable as a basis for action), which is supposed to specify what truth is. Some radical alternatives to the traditional theories result from denying the need for any such further specification (Ramsey, 1927; Strawson, 1950; Quine, 1990). For example, one might suppose that the basic theory of truth contains nothing more than equivalences of the form, "The proposition *that p* is true if and only if *p*" (Horwich, 1990).

This sort of proposal is best presented in conjunction with an account of the *raison d'être* of our notion of truth: namely, that it enables us to express attitudes towards those propositions we can designate but not explicitly formulate. Suppose, for example, you are told that Einstein's last words expressed a claim about physics, an area in which you think he was very reliable. Suppose that, unknown to you, his claim was the proposition that quantum mechanics is wrong. What conclusion can you draw? Exactly which proposition becomes the appropriate object of your belief? Surely not that quantum mechanics is wrong; because you are not aware that that is what he said. What is needed is something equivalent to the infinite conjunction:

If what Einstein said was that  $E = mc^2$ , then  $E = mc^2$ , and if what he said was that quantum mechanics is wrong, then quantum mechanics is wrong, . . . and so on

that is, a proposition, K, with the following properties: that from K and any further premise of the form, "Einstein's claim was the proposition *that p*" you can infer "*p*", whatever it is. Now suppose, as the deflationist says,

that our understanding of the truth predicate consists in the stipulative decision to accept any instance of the schema, "The proposition *that p* is true if and only if *p*". Then your problem is solved. For if K is the proposition, "Einstein's claim is true", it will have precisely the inferential power that is needed. From it and "Einstein's claim is the proposition that quantum mechanics is wrong", you can use Leibniz' law to infer "The proposition that quantum mechanics is wrong is true", which, given the relevant axiom of the deflationary theory, allows you to derive "Quantum mechanics is wrong". Thus one point in favour of the deflationary theory is that it squares with a plausible story about the function of our notion of truth: its axioms explain that function without the need for any further analysis of "what truth is".

Not all variants of deflationism have this virtue. According to the redundancy/performative theory of truth, the pair of sentences, "The proposition *that p* is true" and plain "*p*", have exactly the same meaning and express the same statement as one another; so it is a syntactic illusion to think that "is true" attributes any sort of *property* to a proposition (Ramsey, 1927; Strawson, 1950). But in that case it becomes hard to explain why we are entitled to infer "The proposition that quantum mechanics is wrong is true" from "Einstein's claim is the proposition that quantum mechanics is wrong" and "Einstein's claim is true". For if truth is not a property, then we can no longer account for the inference by invoking the law that if X is identical with Y then any property of X is a property of Y, and vice versa. Thus the redundancy/performative theory, by *identifying* rather than merely *correlating* the contents of "The proposition *that p* is true" and "*p*", precludes the prospect of a good explanation of one of truth's most significant and useful characteristics. So it is better to restrict our claim to the weak, equivalence schema: The proposition *that p* is true if and only if *p*.

Support for deflationism depends upon the possibility of showing that its axioms – instances of the equivalence schema – unsupplemented by any further analysis, will suffice to explain all the central facts about

truth: for example, that the verification of a proposition indicates its truth, and that true beliefs have a practical value. The first of these facts follows trivially from the deflationary axioms. For given our a priori knowledge of the equivalence of “*p*” and “The proposition *that p* is true”, any reason to believe that *p* becomes an equally good reason to believe that the proposition that *p* is true. The second fact can also be explained in terms of the deflationary axioms, but not quite so easily. Consider, to begin with, beliefs of the form.

- (B) If I perform act A, then my desires will be fulfilled.

Notice that the psychological role of such a belief is, roughly, to cause the performance of A. In other words, given that I do have belief (B), then typically:

I will perform act A

And notice also that when the belief is true then, given the deflationary axioms, the performance of A will in fact lead to the fulfillment of one's desires. i.e.

- If (B) is true, then if I perform A, my desires will be fulfilled.

Therefore,

- If (B) is true, then my desires will be fulfilled

So it is quite reasonable to value the truth of beliefs of that form. But such beliefs are derived by inference from other beliefs and can be expected to be true if those other beliefs are true. So it is reasonable to value the truth of any belief that might be used in such an inference.

To the extent that such deflationary accounts can be given of *all* the facts involving truth, then the explanatory demands on a theory of truth will be met by the collection of all statements like “The proposition that snow is white is true if and only if snow is white”, and the sense that some deep analysis of truth is needed will be undermined.

However, there are several strongly felt objections to deflationism. One reason for dissatisfaction is that the theory has an infinite number of axioms, and therefore cannot be completely written down. It can be *described* (as the theory whose axioms are the propositions of the form “*p* if and only if it is true *that p*”), but not explicitly formulated. This alleged defect has led some philosophers to develop theories which show, first, how the truth of any proposition derives from the referential properties of its constituents; and, second, how the referential properties of primitive constituents are determined (Tarski, 1943; Davidson, 1969). However, it remains controversial to assume that *all* propositions – including belief attributions, laws of nature and counterfactual conditionals – depend for their truth values on what their constituents refer to. Moreover, there is no immediate prospect of a decent, finite theory of reference. So it is far from clear that the infinite, list-like character of deflationism can be avoided.

Another source of dissatisfaction with this theory is that certain instances of the equivalence schema are clearly false. Consider.

- (a) THE PROPOSITION EXPRESSED BY THE SENTENCE IN CAPITAL LETTERS IS NOT TRUE.

Substituting this into the schema one gets a version of the “liar” paradox: specifically,

- (b) The proposition *that the proposition expressed by the sentence in capital letters is not true* is true if and only if the proposition expressed by the sentence in capital letters is not true,

from which a contradiction is easily derivable. (Given (b), the supposition that (a) is true implies that (a) is not true, and the supposition that it is not true implies that it is.) Consequently, not every instance of the equivalence schema can be included in the theory of truth; but it is no simple matter to specify the ones to be excluded (*see* Kripke, 1975). Of course, deflationism is far from alone in having to confront this problem.

A third objection to the version of the deflationary theory presented here concerns its reliance on *propositions* as the basic vehicles of truth. It is widely felt that the notion of proposition is defective and that it should not be employed in semantics. If this point of view is accepted then the natural deflationary reaction is to attempt a reformulation that would appeal only to *sentences*: for example,

“p” is true if and only if *p*.

But this so-called “disquotational theory of truth” (Quine, 1990) comes to grief over indexicals, demonstratives and other terms whose referents vary with the context of use. It is not the case, for example, that *every* instance of “I am hungry” is true if and only if *I* am hungry. And there is no simple way of modifying the disquotational schema to accommodate this problem. A possible way out of these difficulties is to resist the critique of propositions. Such entities may well exhibit an unwelcome degree of indeterminacy, and may well defy reduction to familiar items. However, they do offer a plausible account of belief (as relations to propositions) and, in ordinary language at least, they are indeed taken to be the primary bearers of truth (*see* BELIEF).

#### THE ROLE OF TRUTH IN METAPHYSICS AND EPISTEMOLOGY

It is commonly supposed that problems about the nature of truth are intimately bound up with questions as to the accessibility and autonomy of facts in various domains: questions about whether the facts can be known, and whether they can exist independently of our capacity to discover them (Dummett, 1978; Putnam, 1981). One might reason, for example, that if “T is true” means nothing more than “T will be verified”, then certain forms of scepticism (specifically, those that doubt the correctness of our methods of verification) will be precluded, and that the facts will have been revealed as dependent on human practices. Alternatively, it might be said that if truth were an inexplicable, primitive, non-epistemic property, then the fact that T is

true would be completely independent of us. Moreover, we could, in that case, have no reason to assume that the propositions we believe actually have this property; so scepticism would be unavoidable. In a similar vein, it might be thought that a special (and perhaps undesirable) feature of the deflationary approach is that truth is deprived of any such metaphysical or epistemological implications.

On closer scrutiny, however, it is far from clear that there exists *any* account of truth with consequences regarding the accessibility or autonomy of non-semantic matters. For although an account of truth may be expected to have such implications for facts of the form “T is true”, it cannot be assumed without further argument that the same conclusions will apply to the fact, T. For it cannot be assumed that T and “T is true” are equivalent to one another given the account of “true” that is being employed. Of course, if truth is defined in the way that the deflationist proposes, then the equivalence holds by definition. But if truth is defined by reference to some metaphysical or epistemological characteristic, then the equivalence schema is thrown into doubt pending some demonstration that the truth predicate, in the sense assumed, will satisfy it. In so far as there are thought to be epistemological problems hanging over T that do not threaten “T is true”, it will be difficult to give the needed demonstration. Similarly, if “truth” is so defined that the fact, T, is felt to be more (or less) independent of human practices than the fact that “T is true”, then again it is unclear that the equivalence schema will hold. It would seem, therefore, that the attempt to base epistemological or metaphysical conclusions on a theory of truth must fail because in any such attempt the equivalence schema will be simultaneously relied on and undermined.

*See also* COHERENTISM; LINGUISTIC UNDERSTANDING; VERIFICATIONISM.

#### BIBLIOGRAPHY

Austin, J. L.: “Truth,” *Proceedings of the Aristotelian Society* supp. vol. 24 (1950), 111–28.

- Bradley, F. H.: *Essays on Truth and Reality* (Oxford: Clarendon Press, 1914).
- Davidson, D.: "Truth and Meaning," *Synthese* 17 (1967), 304–23.
- Davidson, D.: "True to the Facts," *Journal of Philosophy* 66 (1969), 748–64.
- Davidson, D.: "Truth and Meaning," *Synthese* 17 (1967), 304–23.
- Davidson, D.: "The Structure and Content of Truth," *Journal of Philosophy* 87 (1990), 279–328.
- Dummett, M.: *Truth and Other Enigmas* (Oxford: Clarendon Press, 1978).
- Hempel, C.: "On the Logical Positivist's Theory of Truth," *Analysis* 2 (1935), 49–59.
- Horwich, P. G.: *Truth* (Oxford: Blackwell, 1990).
- James, W.: *The Meaning of Truth* (New York: Longmans Green, 1909).
- Kripke, S.: "Outline of a Theory of Truth," *Journal of Philosophy* 72 (1975), 690–716.
- Papineau, D.: *Reality and Representation* (Oxford: Blackwell, 1987).
- Peirce, C. S.: *Collected Papers*, vols 2–4 (Cambridge, MA: Harvard University Press, 1932).
- Putnam, H.: *Reason, Truth and History* (Cambridge: Cambridge University Press, 1981).
- Quine, W. V.: *Pursuit of Truth* (Cambridge, MA: Harvard University Press, 1990).
- Ramsey, F.: "Facts and Propositions," *Proceedings of the Aristotelian Society* supp. vol. 7 (1927), 153–70.
- Strawson, P.: "Truth," *Proceedings of the Aristotelian Society* supp. vol. 24 (1950), 125–56.
- Tarski, A.: "The Semantic Conception of Truth," *Philosophy and Phenomenological Research* 4 (1943), 341–75.
- Wittgenstein, L.: *Tractatus Logico-Philosophicus* (London: Routledge and Kegan Paul, 1922).

PAUL HORWICH

**truths of reason/truths of fact** This distinction is associated with Leibniz, who declares that there are only two kinds of truths – truths of reason and truths of fact (see LEIBNIZ). The former are all either explicit identities, i.e. of the form "A is A", "AB is B", etc., or they are reducible to this form by successively substituting equivalent terms. Leibniz dubs them "truths of reason" because the explicit identities are self-evident a priori truths, whereas the rest can be converted to such by purely rational operations. Because their denial involves a demonstrable contradiction, Leibniz also says that truths of reason "rest on the principle of contradiction, or

identity" and that they are necessary propositions, which are true of all possible worlds. Some examples are "All equilateral rectangles are rectangles" and "All bachelors are unmarried": the first is already of the form "AB is B" and the latter can be reduced to this form by substituting "unmarried man" for "bachelor". Other examples, or so Leibniz believes, are "God exists" and the truths of logic, arithmetic and geometry.

Truths of fact, on the other hand, cannot be reduced to an identity and our only way of knowing them is a posteriori, or by reference to the facts of the empirical world. Likewise, since their denial does not involve a contradiction, their truth is merely contingent: they could have been otherwise and hold of the actual world, but not of every possible one. Some examples are "Caesar crossed the Rubicon" and "Leibniz was born in Leipzig", as well as propositions expressing correct scientific generalizations. In Leibniz's view, truths of fact rest on the principle of sufficient reason, which states that nothing can be so unless there is a reason why it is so. This reason is that the actual world (by which he means the total collection of things past, present and future) is better than any other possible world and was therefore created by God.

In defending the principle of sufficient reason, Leibniz runs into serious problems. He believes that in every true proposition, the concept of the predicate is contained in that of the subject. (This holds even for propositions like "Caesar crossed the Rubicon": Leibniz thinks anyone who did not cross the Rubicon would not have been Caesar!) And this containment relationship – which is eternal and unalterable even by God – guarantees that every truth has a sufficient reason. If truth consists in concept containment, however, then it seems that all truths are analytic and hence *necessary*; and if they are all necessary, surely they are all truths of reason. Leibniz responds that not every truth can be reduced to an identity in a finite number of steps; in some instances revealing the connection between subject and predicate concepts would require an infinite analysis. But while this may entail that we cannot prove such propositions a priori, it does not appear to



show that the proposition could have been false. Intuitively, it seems a better ground for supposing that it is a necessary truth of a special sort. A related question arises from the idea that truths of fact depend on God's decision to create the best world: if it is part of the concept of this world that it is best, how could its existence be other than necessary? Leibniz answers that its existence is only hypothetically necessary, i.e. it follows from God's decision to create this world, but God had the power to decide otherwise. Yet God is necessarily good, so how *could* he have decided to do anything else? Leibniz says much more about these matters, but it is not clear whether he offers any satisfactory solutions.

*See also* A PRIORI/A POSTERIORI; ANALYTICITY; LEIBNIZ; NECESSARY/CONTINGENT.

## BIBLIOGRAPHY

- Blumenfeld, D.: "Leibniz on Contingency and Infinite Analysis," *Philosophy and Phenomenological Research* 45 (1985), 483–514.
- Copleston, F.: *A History of Modern Philosophy: Descartes to Leibniz*, vol. 4 (Garden City, NY: Doubleday, 1963).
- Leibniz, G. W.: *Philosophical Essays*, ed. and trans. R. Ariew and D. Garber (Indianapolis, IN: Hackett, 1989).
- Mates, B.: *The Philosophy of Leibniz: Metaphysics and Language* (Oxford: Oxford University Press, 1986).
- Rescher, N.: *Leibniz: An Introduction to His Philosophy* (Totowa, NJ: Rowman and Littlefield, 1979).
- Wilson, M. D.: *Leibniz' Doctrine of Necessary Truth* (New York: Garland, 1990).

DAVID BLUMENFELD

# U

---

**underdetermination of theory** It is often our position that several theories are compatible with the available evidence: we then hope that more evidence will resolve the matter. The underdetermination of theory by evidence holds that logically incompatible theories may fit all possible evidence. Alternatively, there may be pairs of empirically equivalent theories which, while not contradicting each other, use radically different theoretical notions. Examples of such underdetermination have been proposed which embarrass empiricist philosophers (Quine, 1990, pp. 95–101) who must either deny that such theories are in competition or find an empirical basis for preferring one of the pair.

*See also* HOLISM.

## BIBLIOGRAPHY

Quine, W. V.: *Theories and Things* (Cambridge, MA: Harvard University Press, 1981).

Quine, W. V.: *Pursuit of Truth* (Cambridge, MA: Harvard University Press, 1990).

CHRISTOPHER HOOKWAY

**use/mention** In the following sentence, the word “cat” is used:

My cat is on the couch.

In the next sentence, the word “cat” is mentioned:

The word “cat” has three letters.

To make explicit that a word is mentioned rather than used, it must be put between quotation marks.

## BIBLIOGRAPHY

Quine, W. V.: *Mathematical Logic* (New York: Harper and Row, 1962), 23–6.

MATTHIAS STEUP

# V

---

**verificationism** Any view according to which the conditions of a sentence's or a thought's being meaningful or intelligible are equated with the conditions of its being verifiable or falsifiable. An explicit defence of the position would be a defence of the verifiability principle of meaningfulness (see LOGICAL POSITIVISM). Implicit verificationism is often present in positions or arguments which do not defend that principle in general, but which reject suggestions to the effect that a certain sort of claim is unknowable or unconfirmable on the sole ground that it would therefore be meaningless or unintelligible. Only if meaningfulness or intelligibility is indeed a guarantee of knowability or confirmability is the position sound. If it is, nothing we understand could be unknowable or unconfirmable by us.

See also PHILOSOPHICAL KNOWLEDGE; REALISM; TRANSCENDENTAL ARGUMENTS; TRUTH.

BARRY STROUD

**Vico, Giambattista (1668–1744)** Italian philosopher, historian and rhetorician. Born and educated in Naples, Vico graduated in law in 1694 and was Professor of Rhetoric at the University from 1699 to 1741. He was disappointed by his failure to gain the chair of Civil Law in 1723, after which he concentrated upon working out his theories of the historical development of nations, expressed in the various editions of *The New Science*.

His interest as an epistemologist lies in two features of his work. The first is his rejection of Cartesianism (see CARTESIANISM) in favour of a purely constructivist theory of knowledge, known as the *verum-factum* theory. The second is his later attempt to modify this in order to show how historical knowledge,

usually thought to be less certain than physical knowledge, could become more certain.

A Cartesian for his first forty years, Vico accepted that the only things which could be known were those which were in principle deducible a priori. A growing interest in history, however, led to dissatisfaction with this theory, according to which the past was not a possible object of knowledge. In his *On the Study Methods of Our Time* (1709) Vico first attacked it on the ground of its inadequacy with respect to mathematics and physics: the truths of mathematics can be known not, as Descartes claimed (see DESCARTES), because they consist in clear and distinct ideas, but because they follow solely from axioms and methods which are of human construction. The projected Cartesian extension of the geometrical method to the physical world is accordingly impossible since the latter is not of human construction. In *On the Most Ancient Wisdom of the Italians* (1710) this insight became the basis of the *verum-factum* theory, that we can know only what we have made, and was used to reject all tenets of Cartesianism which depended upon the theory of clear and distinct ideas. Even *cogito ergo sum* (see COGITO) was dismissed on the ground that since the mind does not create itself, it cannot know the mode of its own construction.

Although the *verum-factum* theory undermined Descartes' account of the main areas of knowledge, it did little to rehabilitate history, for Vico could not see how it could be applied to the realm of human conduct, given the presence there of chance and contingency. It was only when, after further historical study, he came to believe that underlying the histories of different societies lay a common developmental pattern, arising from the operation of laws of historical and social growth, that

he found a subject matter in history stable enough to be an object of knowledge.

So the first crucial claim of Vico's later theory of knowledge, expounded in *The Principles of a New Science of the Nature of Nations* of 1725 and the more complex Second and Third Editions of 1730 and 1744, is that the historical world has been made by men and that its principles must therefore be discovered within the modifications of the human mind. To this Vico adds the idea that the histories of different nations are to be understood as exemplifications of a single pattern, the 'ideal eternal history'. Understanding of this pattern comes not from historical research but from philosophy, which demonstrates the necessity of its various features by reflection upon the growth of mind. Thus history becomes scientific because it is pursued in the light of philosophical theory, while philosophy relates to the real through its capacity to ground the principles of historical interpretation. The result is that historical events cease to be merely contingent, being manifestations of necessary aspects of the ideal eternal history.

The theory gives rise to two problems. The first concerns the question whether it is sufficiently constructivist to achieve the degree of certainty which Vico claims. The initial *verum-factum* theory is plausible, in relation to the world of mathematics at least, since the conventional nature of mathematics can be shown by the fact that its content can be altered by an alteration in its axioms and assumptions. This is not true, however, of the world of history, where Vico does not deny that the historian can alter his theories without thereby altering the activities in which historical reality consists. The real past thus remains as external to the Vichian historian as do the clear and distinct ideas, against which Vico first reacted, to the Cartesian mathematician. The second related problem arises from the necessity to treat historical reality as the exemplification of laws, in order to benefit from the combination of historical research and philosophical reflection. For even if these laws operate through human activities, they are not themselves of human making. It is difficult to see

therefore how the fact that Vico's historian makes history, in the sense of making theories about the necessary conditions of the development of mind and applying them to historical evidence, can be a reason for holding that this entitles him to a knowledge of past reality which is more certain than that which, by similar methods, could be gained in any other law-governed field.

See also HISTORICAL KNOWLEDGE.

#### WRITINGS

*De antiquissima Itolorum sapientia ex linguae latinae originibus eruenda* (Naples: 1710), trans. L. M. Palmer as *On the Most Ancient Wisdom of the Italian* (Ithaca, NY: Cornell University Press, 1988).

*De nostri temporis studiorum ratione* (Naples, 1709), trans. E. Gianturco as *On the Study Methods of Our Time* (Indianapolis, IN: Bobbs-Merrill, 1968).

*Princi di una Scienza Nuova Dintorno alla Natura delle Nazioni* (Naples: 1725), abridged and trans. L. Pompa as *Vico: Selected Writings* (Cambridge: Cambridge University Press, 1982).

*Principi di Scienza Nuova d'Intorno alla Comune Natura delle Nazioni* (Naples: 1730, 1744), trans. T. G. Bergin and M. H. Fisch as *The New Science of Giambattista Vico* (Ithaca, NY: Cornell University Press, 1984).

*Vita di Giambattista Vico Scritta de Se Medesimo* (Naples: 1818), trans. T. G. Bergin and M. H. Fisch as *The Autobiography of Giambattista Vico* (Ithaca, NY: Cornell University Press, 1975).

#### BIBLIOGRAPHY

Bedani, G.: *Vico Revisited* (Oxford: Berg, 1989).  
Berlin, I.: *Vico and Herder* (London: The Hogarth Press, 1976).

Croce, B.: *La Filosofia di Giambattista Vico* (Bari: 1911); trans. R. G. Collingwood as *The Philosophy of Giambattista Vico* (London: Howard Latimer, 1913).

Pompa, L.: *Vico: A Study of the "New Science"*, 2nd edn (Cambridge: Cambridge University Press, 1990).

Verene, D. P.: *Vico's Science of Imagination* (Ithaca, NY: Cornell University Press, 1981).

LEON POMPA

**Vienna Circle** A group of philosophers, scientists and mathematicians, often known collectively as the logical positivists (*see* LOGICAL POSITIVISM), who met in Vienna under the leadership of Moritz Schlick (*see* SCHLICK) in the late 1920s and 1930s. The general philosophical stance of the Circle was anti-metaphysical and “scientific”, and it was noted for its advocacy of the so-called “Verification Principle” (*see* VERIFICATIONISM). Members included Carnap, Neurath, Herbert Feigl, Kurt Gödel and Friedrich Waismann (*see* CARNAP; NEURATH). Popper and Reichenbach were associated with the

Circle, and Ayer and Quine also attended meetings (*see* POPPER; REICHENBACH; AYER; QUINE). The Circle were heavily influenced by their understanding (or misunderstanding) of Wittgenstein’s early philosophy (*see* WITTGENSTEIN). Schlick was assassinated by a deranged student in 1936, Popper (on his own account) “murdered” the verification principle, and Hitler’s annexation of Austria in 1938 brought the meetings to an end. By the time of the Second World War most members had emigrated to Britain or the United States.

ANDY HAMILTON



# W

---

## **Whitehead, Alfred North (1861–1947)**

Mathematician (*A Treatise on Universal Algebra*, 1898), logician (*Principia Mathematica*, with Bernard Russell, 1910–13), philosopher of science (*An Enquiry Concerning the Principles of Natural Knowledge*, 1919; *The Concept of Nature*, 1920; *The Principle of Relativity*, 1922), educator (*The Aims of Education and Other Essays*, 1929), and metaphysician (*Science and the Modern World*, 1925; *Process and Reality*, 1929; *Adventures of Ideas*, 1933). Whitehead begins the preface to his *magnum opus*, *Process and Reality*, as follows: “These lectures are based upon a recurrence to that phase of philosophic thought which began with Descartes and ended with Hume.” This piece of stage-setting reflects Whitehead’s conviction that epistemology has been the dominant dimension of philosophy in recent centuries because of Descartes’ exclusion of mind from nature, focusing subsequent attention on the issue of how mind can then know anything about the nature from which it has been divorced. The work of Hume and Santayana (see HUME; SANTAYANA) exhibits for Whitehead the insuperable character of the epistemological problems which the Cartesian ontological commitment entails. Whitehead’s response is that “all difficulties as to first principles are only camouflaged metaphysical difficulties. Thus also the epistemological difficulty is only solvable by an appeal to ontology” (*Process and Reality*, p. 189).

His thinking enriched by the twentieth-century revolutions in science as well as by the Theory of Evolution, Whitehead replaces the Cartesian dualistic ontology with a monistic version of a philosophy of organism, which is now popularly known as process philosophy. The new ontology, bringing together what Descartes cast asunder, includes a novel

theory of propositions according to which “there is a ‘correspondence’ theory of the truth and falsehood of propositions, and a ‘coherence’ theory of the correctness, incorrectness and suspension of judgments” (ibid., p. 191). Whitehead’s conviction is that “At the end, in so far as the [ontological] enterprise has been successful, there should be no problem of space-time, or of epistemology, or of causality, left over for discussion. The scheme should have developed all those generic notions adequate for the expression of any possible interconnection of things” (ibid., p. xii).

## WRITINGS

*Process and Reality*, corrected edn, ed. D. R. Griffin and D. W. Sherburne (New York: The Free Press, 1978).

DONALD W. SHERBURNE

## **Wittgenstein, Ludwig (1889–1951)**

Austrian-born philosopher who spent much of his working life at Cambridge.

The self-criticism of thought lacks the appeal of speculation, and critical philosophy has sometimes struck people as an evasion with no right to the title “philosophy”. The impression of irrelevance became much stronger in this century when the direct critique of thought was replaced by a critique of the language expressing the thought. Wittgenstein, the greatest philosopher in this school, once said that the subject on which he worked might be called “one of the heirs of the subject which used to be called ‘philosophy’” (1958, p. 28). However, his writings have seldom been dismissed as irrelevant. He draws the line between the meaningful and the

meaningless, but he draws it imaginatively and without the flat repressiveness of Positivism (see LOGICAL POSITIVISM), and he never conveys the feeling that the horizon has contracted. His work has even achieved an unusual popularity among people with no training in philosophy who find that it has something to say to them in spite of the difficulty of its interpretation.

His philosophy falls into two periods separated by an interval during which he turned to other things. Born in Vienna, he studied engineering in Berlin and Manchester, became interested in the foundations of mathematics and from 1911, when he went to Cambridge to work with Russell, until the end of the First World War, in which he served in the Austrian army and was taken prisoner, he developed the theory of language and logic which he published in his first book, *Tractatus Logico-Philosophicus* (1922). After that he abandoned philosophy and did not return to it until the end of the 1920s, when he discussed his book with some of the philosophers of the Vienna Circle, became for a short time Professor of Philosophy at Cambridge, and continued to elaborate a largely new set of ideas until his death. The first result of his later work was his posthumous *Philosophical Investigations* (1953), which has been followed by the publication of many volumes of his notes.

His aim in both periods of his philosophy was to determine what language can and cannot do. This is a more radical enterprise than fixing the scope and limits of human knowledge, because knowledge is expressed in sentences, which have to achieve sense before they can achieve truth. He develops his critique of language in two quite different ways in the two periods of his philosophy. In the *Tractatus* a general theory of language is used to fix the bounds of sense, while in *Philosophical Investigations* no such theory is offered, and the line between sense and senselessness is drawn not on any general principle but with an eye on the special features of each case that is reviewed. If his first book is like a map with a superimposed grid, his second book is like the diary of a journey recording all the deviations which looked so

tempting but would have ended in the morass of senselessness. He himself drew attention to the analogy between this later work and psychotherapy (*PI*, pp. 47, 91).

The theory that fixes the bounds of sense in the *Tractatus* assimilates sentences to pictures. A sentence can achieve sense only in one of two ways; either it will picture a fact, or else it will be analysable into further, more basic sentences which picture facts. The sense-giving relation between language and the world is called "picturing" because the words in a basic (elementary) sentence are supposed to stand for objects in the same way that points on the surface of a picture stand for points in physical space. Everything that we can say, and equally everything that we can think, must be a projection of a possible arrangement of objects.

The Picture Theory draws the boundary of sense very tightly around factual sentences, excluding moral, aesthetic and religious discourse. The *Tractatus* was, therefore, naturally adopted by the philosophers of the Vienna Circle as a model for Logical Positivism. However, it is a work which presents many facets to the world and there is another, more sympathetic way of reading it: the excluded types of discourse are not eliminated, because they are preserved by the very different roles that they play in our lives. The assimilation of their roles to the role of factual discourse is a misunderstanding, whether it is intended to preserve them (as in H. Spencer's scientific ethics) or to destroy them (as in the writings of Positivists). The *Tractatus* gives factual language (ordinary and scientific) the central place but resists Scientism.

Consistently with their general interpretation of the *Tractatus*, the Vienna Circle construed the Picture Theory as an empiricist theory of meaning (see EMPIRICISM) and identified Wittgenstein's "objects" with sense-data (see SENSE-DATA). But he himself had avoided any such identification. His "objects" were, by definition, simple (i.e. they had no internal structure) and he deduced from the Picture Theory that they must exist at some level of analysis in order to give factual sentences their senses. However, he never claimed to have discovered them or even to know their

category, and he never shared Russell's view (see RUSSELL) that when we acquire factual language, we find that there are words whose meanings we can learn only through acquaintance with the things designated by them and that those things are simple objects (see Russell, 1956, pp. 193–5) (see KNOWLEDGE BY ACQUAINTANCE/BY DESCRIPTION). That was an empirical argument for an empiricist account of the foundations of language, and he neither followed that route nor arrived at that destination.

The *Tractatus* is a work of rare originality and the history of philosophy shows that such books often produce effects which do not square with their authors' intentions. Fortunately, in this case we have the journal (Wittgenstein, 1961) which allows us to follow the development of the ideas that went into it, his own, Russell's and Frege's (see FREGE). This is a much better guide to its interpretation than the reactions of its early readers. It should be taken as an abstract treatise on language and, though its results affect other disciplines and even the nature of philosophy itself, it is not based on the doctrines of any particular school of philosophy.

Its main achievement is a uniform theory of language which yields an explanation of logical truth. A factual sentence achieves sense by dividing the possibilities exhaustively into two groups, those that would make it true and those that would make it false. A truth of logic does not divide the possibilities but comes out true in all of them. It, therefore, lacks sense and says nothing, but it is not nonsense. It is a self-cancellation of sense, necessarily true because it is a tautology, the limiting case of factual discourse, like the figure 0 in mathematics.

It was precisely the uniformity of this theory of language that Wittgenstein found unacceptable when he took up philosophy in the late 1920s for the second time. Language takes many forms and even factual discourse does not consist entirely of sentences like "The fork is to the left of the knife". However, the first thing that he gave up was the idea that this sentence itself needed further analysis into basic sentences mentioning simple objects with no internal structure (see Wittgenstein,

1929; 1975, pp. 105–14, 317). He now conceded that a descriptive word will often get its meaning partly from its place in a system, and he applied this idea to colour-words, arguing that the essential relations between different colours do not indicate that each colour has an internal structure which needs to be taken apart. On the contrary, analysis of our colour-words would only reveal the same pattern – ranges of incompatible properties – recurring at every level, because that is how we carve up the world.

This may look like a small change, a footnote to the *Tractatus*, but in fact it was the first sign of a revolution in his philosophy. He ceased to believe that the actual structure of our discourse can only be explained as the confused manifestation of a deeper structure, which will be revealed by an analysis yet to be successfully completed. "We make our moves in the realm of the grammar of our ordinary language, and this grammar is already there. Thus we already have everything and need not wait for the future" (1979, p. 183).

Indeed, it may even be the case that the grammar of our ordinary language is created by moves which we ourselves make. If so, the philosophy of language will lead into the philosophy of action. Certainly there is a close connection between the meaning of a word and the applications of it which its users intend to make. There is also an obvious need for people to understand each other's meanings and that requires agreement in the applications of their words. There are many links between the philosophy of language and the philosophy of mind and it is not surprising that the impersonal examination of language in the *Tractatus* was replaced by a very different, anthropocentric treatment in *Philosophical Investigations*.

If the logic of our language is created by moves which we ourselves make, various kinds of realism (see REALISM) are threatened. First, the way in which our descriptive language carves up the world will not be forced on us by the natures of things, and the rules for the application of our words, which feel like external constraints, will really come from within us (PI, §§130–242). That is a

concession to nominalism which is, perhaps, readily made. The idea that logical and mathematical necessity are also generated by what we ourselves do is more paradoxical. Yet that is the conclusion of Wittgenstein (1956) and (1976), and here his anthropocentrism has carried less conviction. However, paradox is not a sure sign of error and it is possible that what is needed here is a more sophisticated concept of objectivity than Platonism provides (see OBJECTIVITY).

In his later work Wittgenstein brings the great problems of philosophy down to earth and traces them to very ordinary origins. His examination of the concept of *following a rule* takes him back to a fundamental question about counting things and sorting them into types: “What qualifies as doing the same again?” Of course, an expert would regard this question as extraneous rather than fundamental, and would suggest that we forget it and get on with the subject. But Wittgenstein’s question is not so easily dismissed. It has the naive profundity of questions that children ask when they are first taught a new subject. Such questions remain unanswered without detriment to their learning, but they point the only way to a complete understanding of what is learned (see Wittgenstein, 1974, pp. 381–2).

The philosophy of mind, recessive in the *Tractatus*, dominates his later work, both because of its connections with meaning and necessity and in its own right. Solipsism (see SOLIPSISM), for example, is important, not, of course, because many philosophers have adopted it but because it may well be the unavoidable consequence of more moderate and, therefore, more popular theories of mind, and it is discussed briefly in the *Tractatus* and at length in his later writings. His critique of solipsism should be contrasted with Russell’s. Russell argued that the solipsist has overwhelming inductive reasons for believing in the existence of things outside his own mind and of other people like himself observing them (Russell, 1956, pp. 125–74). Wittgenstein’s criticism was the more radical one, that when the solipsist cuts himself off from the external world, he deprives himself of any way of giving sense to his thoughts about his

private world (see *Tract.*, 5.6–5.641; *PI* §§243ff; and Wittgenstein, 1968).

This is his famous private language argument (see PRIVATE LANGUAGE ARGUMENT), which in exegeses of his ideas is often presented as an isolated, self-sufficient refutation of any philosophy with a purely mentalistic basis. In fact, it is part of a more general critique of the kind of intellectualism which assumes that a philosopher can stand in the position of the solipsist and still bring all the resources of his mind to bear on his restricted set of data. But if he really turns his mind back on the physical world, how will he ever have acquired and maintained the concept of himself as a person? Any sentence beginning “I . . .” will pose an insoluble problem for his understanding. If the sentence continues “. . . am in pain”, the understanding of those words is not a purely intellectual achievement for us, because they replace our natural pain-behaviour, but he will not have that resource (see *PI*, §§244–5). Similarly, our vocabulary for identifying points in local space was not correlated with the facts directly by the intellect working alone, but based on a picture of the world which antedated the advent of language and in which we could already find our way around. Has the philosopher who experimentally adopts the standpoint of the solipsist and so deprives himself of all such resources any right to assume that he can keep meaning safe?

Wittgenstein does not confine himself to showing that the solipsist’s position is untenable, but also offers a diagnosis of the mistakes that led him into it. The first, perhaps the most important, mistake was uncritical scepticism about the physical world. This is dismissed with characteristic brevity in the *Tractatus*: “Scepticism is *not* irrefutable, but obviously nonsensical, when it tries to raise doubts where no questions can be asked.” Towards the end of his life he returned to this topic and developed a more detailed and satisfying critique of the sceptic’s voracious doubts. The leading idea of this later treatment was holistic: some statements about the world achieve immunity from doubt not because they are more thoroughly confirmed, but because they provide the framework within

which alone other statements can be questioned and confirmed or rejected (see Wittgenstein, 1969).

See also OTHER MINDS; PRIVATE LANGUAGE ARGUMENT; PROBLEM OF RULE-FOLLOWING; SOLIPSISM.

#### WRITINGS

*Tractatus Logico-Philosophicus* trans. C. K. Ogden (London: Routledge and Kegan Paul, 1922); trans. D. F. Pears and B. F. McGuinness (London: Routledge and Kegan Paul, 1961; 2nd edn 1974).

"Some Remarks on Logical Form," *Proceedings of the Aristotelian Society* supp. vol. 9 (1929), 162–71; reprinted in *Essays on Wittgenstein's Tractatus*, ed. R. Beard and I. Copi (London: Routledge and Kegan Paul, 1966).

*Philosophical Investigations*, trans. G. E. M. Anscombe (Oxford: Blackwell, 1953).

*Remarks on the Foundations of Mathematics*, trans. G. E. M. Anscombe (Oxford: Blackwell, 1956).

*The Blue and Brown Books* (Oxford: Blackwell, 1958).

*Notebooks 1914–1916*, trans. G. E. M. Anscombe (Oxford: Blackwell, 1961).

Notes for Lectures on "Private Experience" and "Sense Data," *Philosophical Review* 77 (1968), 275–320.

*On Certainty* (Oxford: Blackwell, 1969).

*Philosophical Grammar*, trans. A. Kenny (Oxford: Blackwell, 1974).

*Philosophical Remarks*, trans. R. Hargreaves and R. White (Oxford: Blackwell, 1975).

*Lectures on the Foundations of Mathematics*, ed. C. Diamond (Hassocks: Harvester, 1976).

*Ludwig Wittgenstein and the Vienna Circle: Conversations Recorded by Friedrich Waismann*, trans. J. Schulte and B. McGuinness (Oxford: Blackwell, 1979).

#### BIBLIOGRAPHY

Hacker, P. M. S.: *Wittgenstein's Place in Twentieth-century Analytic Philosophy* (Oxford: Blackwell, 1996).

Kenny, A.: *Wittgenstein* (London: Penguin, 1973).  
McGuinness, B.: *Wittgenstein: A Life. Young Ludwig 1889–1921* (London: Duckworth, 1988).

Monk, R.: *Ludwig Wittgenstein: The Duty of Genius* (London: Cape, 1990).

Pears, D. F.: *The False Prison*, 2 vols (Oxford: Oxford University Press, 1987, 1988).

Russell, B.: *Logic and Knowledge: Essays 1901–1950*, ed. R. C. Marsh (London: Allen and Unwin, 1956).

Stroll, Avrum: *Moore and Wittgenstein on Certainty* (New York: Oxford University Press, 1994).

DAVID PEARS



# Z

---

**zebras and cleverly disguised mules** An alleged counter-example to epistemic closure owing to Fred Dretske (see FRED DRETSKE self-profile in Part II) and an epistemological puzzle in its own right. Suppose I am at a zoo standing before a well-marked zebra exhibit in broad daylight and clearly see a black-and-white-striped equine animal ten feet in front of me. In such a situation, it seems clear that:

- (1) I know that the animal before me is a zebra.  $[Kz]$

But now consider the following skeptical hypothesis: In an effort to save money, the zoo director has filled the zebra exhibit exclusively with white mules whose hair has been dyed meticulously in stripes so as exactly to resemble the coat of zebras. Sometimes people engage in such fraudulent behavior. Do I know that the animal before me is not a cleverly disguised mule  $[CDM]$ ? If it were a CDM, it would have black – and white stripes and look exactly as it does. Since I haven't done the genetic tests needed to rule out the CDM scenario, it seems true that:

- (2) I do not know that the animal before me is *not* a cleverly disguised mule.  $[\sim K\sim c]$

From here, it's two short steps to paradox. I know that zebras are not mules of any stripe. So:

- (\*) I know that if the animal before me is a zebra, then the animal before me is not a cleverly disguised mule.  $[K(z \rightarrow \sim c)]$

Now, consider a simplified version of the highly plausible principle of epistemic closure:

- (PEC) If I know that  $p$  and I know that  $p$  implies  $q$ , then I know that  $q$ .  $[(Kp \ \& \ K(p \rightarrow q)) \rightarrow Kq]$

A precise articulation of the closure principle would require that the epistemic subject competently deduce  $q$  from her knowledge that  $p$  and  $(p \rightarrow q)$ , though nothing turns on that refinement here. (For additional needed refinements, see Hawthorne, 2005.) The rationale for (PEC) is straightforward: Intuitively, it seems we know what we recognize to be the obvious logical consequences of our knowledge.

Despite the intuitiveness of (PEC), if (1), (2) and (\*) are all true, we have a counter-example to (PEC), namely:  $Kz \ \& \ K(z \rightarrow \sim c) \ \& \ \sim K\sim c$ . While Dretske's example has failed to provide a *decisive* counter-example to closure, it has left an epistemological paradox in its wake, arising from the seeming plausibility of each of the following three propositions:

- (1)  $Kz$   
 (2)  $\sim K\sim c$   
 (3)  $\sim K\sim c \rightarrow \sim Kz$

In light of the obviousness of " $z \rightarrow \sim c$ ", (3) functions as an abbreviated stand-in for (PEC): Given  $K(z \rightarrow \sim c)$ , closure entails that if I don't know that the animal is not a CDM, then I don't know that it is a zebra. Each of (1)–(3) strikes us as intuitively obvious, and yet one of them must be false. This inconsistent triad forms *the zebra paradox*. Six main responses to this paradox have been proposed.

*The Skeptical Response* is to accept (2) and (3) and reject (1). Skeptics insist that I don't know that  $z$  unless I can eliminate those possibilities I know to be incompatible with the animal's being a zebra (Lehrer, 1971: 295;

Stroud, 1984: 16–17). One such  $\sim z$ -possibility is  $c$ . Skeptics contend I don't know that  $\sim c$  because my perceptual evidence is consistent with  $c$ : If the animal were a CDM, my visual experience would be just as it is. Appealing to both  $\sim K\sim c$  and closure, skeptics then derive  $\sim K\sim z$ .

The *Moorean Response* turns the skeptic's reasoning on its head. The Moorean accepts (1) and (3) and then denies (2) on that basis. Appealing to common sense, the Moorean insists that I do know that  $z$ , for I am perceiving the zebra in good lighting conditions and I know the unmistakable look of zebras. The Moorean then uses closure to conclude that I also know that  $\sim c$  (Moore, 1962: 144–8 and 220–2; Pollock, 1986: 3–7).

Both of these responses have been criticized for failing to explain why we were led into paradox in the first place. An adequate solution to the paradox must not only identify the false statement in the triad, it must also explain why we found that false statement so compelling initially. Both responses just considered utterly fail in the latter regard.

The *Closure Denial Response*, advocated in Dretske (1970), maintains that (1) and (2) are true and (3) is false. According to Dretske,  $S$  knows that  $p$  only if  $S$  has a conclusive reason for  $p$ , where  $R$  is a *conclusive reason* for  $p$  just in case if  $p$  were false then  $R$  would be false (Dretske, 1971). To assess this counterfactual, one must go to the closest  $\sim p$ -worlds to see whether  $R$  is true in any of these worlds. If  $R$  is true in any of the closest  $\sim p$ -worlds, then  $R$  is not a conclusive reason for  $p$ .

In the envisaged zebra scenario, I do have a conclusive reason for  $z$ , namely:

$L$ : The animal looks like a zebra.

$L$  is a conclusive reason for  $z$ , because in these circumstances (reputable zoo, normal lighting conditions, etc.), if  $z$  were false,  $L$  would also be false: In the closest  $\sim z$ -worlds, the animal would be some other zoo animal (e.g. a giraffe) and would not look like a zebra. However, I do *not* have a conclusive reason for  $\sim c$ . If the animal in the exhibit were a CDM, the animal would look exactly like it does. Accordingly, if  $c$  were true (i.e. if

$\sim c$  were false),  $L$  would still be true. Hence,  $L$  is not a conclusive reason for  $\sim c$ . On the assumption that conclusive reasons are both necessary and sufficient for knowledge, since I have a conclusive reason for  $z$  but lack a conclusive reason for  $\sim c$ , we get:

(1)  $Kz$

and

(2)  $\sim K\sim c$ .

Hence,

(3)  $\sim K\sim c \rightarrow \sim Kz$

is false.

Closure advocates stress that denying closure has highly counter-intuitive results (Hawthorne, 2005; DeRose, 1995). One such result is the following: On Dretske's account,  $L$  is not only a conclusive reason for  $z$ ;  $L$  is also a conclusive reason for believing that the animal is not a mule [ $\sim m$ ], because, if  $m$  were true,  $L$  would be false. The fact that I have a conclusive reason for  $\sim m$  but lack a conclusive reason for  $\sim c$  commits Dretske to the following abominable conjunction: I know that the animal is not a mule, but I don't know that it's not a cleverly disguised mule.

The *Contextualist Response* asserts that which proposition of the paradox we should deny depends on how the conversational context we are in fixes the meaning of the word "know". The semantic standards governing knowledge ascriptions, according to contextualism, are a function of the error possibilities salient in the context of ascription. As more error possibilities become *salient* in a given conversation, the standards for truly ascribing "knows" rise, often to the point where we fail to meet those standards. Contextualists exploit this idea to resolve the zebra paradox as follows: In ordinary contexts where far-fetched error possibilities like the CDM possibility are out – of sight and out – of mind, the standards for truly ascribing "knows" tend to be low. Relative to such a low standards context, (1) is true. When we initially confront (1), we are in an ordinary

low-standards context, and we rightly judge (1) to be true. Then we're presented with (2), which makes salient for us the CDM possibility. Once the CDM possibility becomes salient, the standards for zebra "knowledge" are so high that I fail to meet them, and now we rightly judge that (2) and (3) are true. While (1) was true in the original low-standards context, (1) is false in this new higher-standards context. The paradox is resolved by calling our attention to the context sensitivity of the word "know" and by alerting us to the fact that all three sentences simultaneously seem true only when we fail to attend to these context shifts and thereby equivocate between these different senses of "know" (Cohen, 1988; DeRose, 1995; Lewis, 1996) (*see also* CONTEXTUALISM in Part I).

The contextualist response stands or falls on its proposed semantics for the word "know". One problem for such a semantics emerges when we consider what happens in a conversation between a Moorean and a skeptic, neither of whom adjusts to the other's use of "know". The contextualist has a limited number of ways of explaining what's going on in such an "uncooperative" conversation, none of which is attractive. Option 1: Semantic chaos ensues, and none of the knowledge ascriptions being uttered by either party expresses a proposition (*see* DeRose, 2004). Option 2: The mere mentioning of a skeptical error possibility immediately makes that error possibility salient (Lewis, 1996), thereby automatically raising the standards for knowledge, in which case the Moorean simply isn't a competent language-user. Option 3: When the Moorean refuses to play along with the skeptic and simply dismisses the error possibility out of hand, the standards stay low and the Moorean automatically wins (*see* Cohen, 1999: 85). If the skeptic continues to insist that I don't know that the animal is a zebra in the face of such a Moorean veto, then the skeptic is an incompetent language-user. It is highly implausible that either the Moorean or the skeptic is an incompetent language-user.

*The Ambiguity Response* maintains that the paradox arises because the word "know" is ambiguous and is resolved by recognizing

and removing this ambiguity. The ambiguity arises because there are two principal senses of knowledge – infallible knowledge and fallible knowledge. According to infallibilism, *S* knows that *p* only if *S*'s justification for *p* entails *p*. According to fallibilism, *S* knows that *p* only if *S*'s justification renders *p* extremely probable (but it need not entail *p*). On fallibilist accounts, justification sufficient for knowing *p* is compatible with  $\sim p$ , though such justification renders  $\sim p$  highly unlikely. Ambiguity theorists maintain that unqualified knowledge ascriptions are genuinely ambiguous between infallibilist and fallibilist readings at the time of utterance. On the infallibilist reading, (2) and (3) are true, and (1) is false. On the fallibilist reading, (1) and (2) are true, and (3) is false. There is no univocal reading of (1)–(3) such that they are all true. All three sentences only seem true when we equivocate between fallibilist and infallibilist readings (Feldman, 1986; Engel, 2004; Steup, 2005).

The ambiguity response has some affinities with the contextualist response, but there is one crucial difference. On a contextualist semantics, when a knowledge ascription is uttered in a given conversation, it has a uniquely determined univocal meaning on that occasion. While that same sentence would be ambiguous (or meaningless) outside a context of ascription, inside a context of ascription its meaning is fixed and univocal. Whereas, on the ambiguity response, when a knowledge ascription is uttered in a conversation, that sentence remains ambiguous at the time of utterance. This residual ambiguity allows the ambiguity theory to explain both the Moorean and skeptical reactions to the paradox. The Moorean interprets each ambiguous sentence in the triad according to ordinary fallibilistic standards. That is why she rejects (2). The skeptic interprets each sentence in the triad according to infallibilistic standards, and so she rejects (1). Each is correct for her respective interpretation. Neither is an incompetent language-user. They may be speaking past each other, but they are still making sense.

Critics of the ambiguity response insist that the zebra paradox iterates when we use

fallible knowledge [ $\text{know}_f$ ] throughout, i.e. each of the following statements seems true:

- (1')  $K_f(z)$
- (2')  $K_f(z \rightarrow \sim c)$
- (3')  $\sim K_f(\sim c)$

Ambiguity theorists must reject (3') and insist that I do have knowledge<sub>f</sub> that the animal is not a CDM. The challenge, of course, is to explain how it is that I know<sub>f</sub> that  $\sim c$ , when I don't possess any evidence of the animal's  $\sim$ CDM status. The animal looks exactly like what a CDM would look like, and I've done no other tests to rule out the animal's CDM status. How, then, is it that I know<sub>f</sub> that  $\sim c$ ?

*The Unhappy-face Resolution Response*, due to Schiffer, contrasts sharply with the extant happy-face solutions discussed above. *Happy-face* solutions purport to identify which proposition in the triad is false. In contrast, an *unhappy-face* resolution maintains that the problem doesn't lie with one of the seemingly inconsistent propositions, but rather lies with the *concept of knowledge* itself. On this view there is a "deep-seated incoherence" in our concept of knowledge, because "The criteria we employ to tell us when we know something conflict, and the concept contains no higher criterion whose application can resolve the conflict" (Schiffer, 1996: 330). That incoherence can be demonstrated by noting three principal theses that seem constitutive of our concept of knowledge.

- Perception*: We gain knowledge of the world through the proper exercise of our perceptual faculties in situations where those faculties are reliable.
- Closure*: We can properly expand our knowledge by competently deducing the consequences of other propositions that we know.
- Sensitivity*: A person knows that  $p$  only if she is sensitive to  $p$ 's truth value. If she would believe  $p$  even if  $p$  were false, then it is, in some sense, *accidental* or

*epistemically lucky* that she happens to be believing  $p$  when  $p$  is true, and thus she fails to know that  $p$ .

The problem is that no coherent concept of knowledge can incorporate all three theses, and yet each thesis seems to be part of the underived conceptual role of the concept of knowledge.

If our concept of knowledge does incorporate *perception*, *closure*, and *sensitivity*, there is no happy-face solution to the zebra paradox. Given *perception*, we must judge that (1) is true. Given *closure*, we must judge that (3) is true. And, given *sensitivity*, we must judge that (2) is true. So, if *perception*, *closure*, and *sensitivity* are indeed essential ingredients of our concept of knowledge, the zebra paradox has no solution but brings to the fore that our concept of knowledge is deeply incoherent.

See also CONTEXTUALISM in Part I; FRED DRETSKE self-profile in Part II; MOORE; NOZICK.

#### BIBLIOGRAPHY

- Cohen, S.: "How to Be a Fallibilist," *Philosophical Perspectives* 2, Issue Epistemology (1988), 91–123.
- Cohen, S.: "Contextualism, Skepticism, and the Structure of Reasons," *Philosophical Perspectives, Epistemology* (1999), 57–89.
- DeRose, K.: "Solving the Skeptical Problem," *The Philosophical Review* 104 (1995), 17–52.
- DeRose, K.: "Single Scoreboard Semantics," *Philosophical Studies* 119 (2004), 1–21.
- Dretske, F.: "Conclusive Reasons," *Australasian Journal of Philosophy* 49 (1971), 1–22.
- Dretske, F.: "Epistemic Operators," *Journal of Philosophy* 67 (1970), 1007–23.
- Engel, M.: "What's Wrong with Contextualism, and a Noncontextualist Resolution of the Skeptical Paradox," *Erkenntnis* 61 (2004), 203–31.
- Hawthorne, J.: "The Case for Closure," in *Contemporary Debate in Epistemology*, ed. M. Steup and E. Sosa (Malden, MA: Blackwell, 2005), 26–43.
- Feldman, F.: *A Cartesian Introduction to Philosophy* (New York: McGraw-Hill 1986).
- Klein, P.: "Contextualism and the Real Nature of Academic Skepticism," *Philosophical Issues* 10, *Skepticism* (2000), 108–16.

- Klein, P.: "Skepticism and Closure: Why the Evil Genius Argument Fails," *Philosophical Topics* (1995), 213–36.
- Lehrer, K.: "Why Not Scepticism?," *The Philosophical Forum* 2 (1971), 283–98.
- Lewis, D.: "Elusive Knowledge," *Australasian Journal of Philosophy* 74 (1996), 549–67.
- Moore, G. E.: *Philosophical Papers* (New York: Collier Books, 1962).
- Pollock, J.: *Contemporary Theories of Knowledge* (Totawa, NJ: Rowman and Littlefield, 1986).
- Schiffer, S.: "Contextualist Solutions to Scepticism," *Proceedings of the Aristotelian Society* 96 (1996), 317–33.
- Steup, M.: "Contextualism and Conceptual Disambiguation," *Acta Analytica* 20 (2005), 3–15.
- Stroud, B.: "The Problem of the External World," in *Epistemology: An Anthology*, ed. E. Sosa and J. Kim (Malden, MA: Blackwell Publishers, 1984).

MYLAN ENGEL



# INDEX

---

*Note:* Page references in **bold** indicate corresponding essays; page references within parentheses, following the indexed name of a contemporary epistemologist, refer to an essay or entry written by that epistemologist.

- a priori/a posteriori 43–53, **219–21**, 330, 394, 547, 548, 660, 778
- “A Priori Knowledge” 38–9, **43–53**, 109, 116–7, 148, 154, 219–21, 229, 326, 455, 475, 487, 501, 505–6, 519, 527, 546–7, 548, 642, 653, 663, 707, 722
- abduction **221**, 445, 449–50, 580–1, 600
- absurdity **221**
- Academy (Plato) **221–2**, 393, 752
- act/object analysis **222**, 358–61, 586, 702–5, 732
- adverbial theory **222–3**, 237, 239, 360–1, 586, 704–5
- agnosticism **223**
- Alston, William P. 40–1, 58, **223–4**, 333–5, (382–5), 431, (433), (438), (444), (614–5), 682
- analyticity 43–53, 154, **224–30**, 330, 387, 395, 437, 455, 462–3, 475, 503, 505–6, 525, 547, 550, 566, 593, 642, 655, 657, 662, 723, 778
- anamnesis **230**, 455, 548
- antinomy **230–1**, 572
- apodeictic **231**
- aporia **231**
- apperception **231**, 422, 460, 475, 726
- Aquinas, Thomas **231–5**, 259, 448, 501, 562, 755–6
- argument **235**
  - from analogy **235–6**, 388, 567, 571, 617, 651, 749
  - from illusion 64, **236–40**, 248, 250, 308, 321, 361, 400, 621, 630, 634, 732
- Aristotle 77–8, 210, 222, 231–5, **240–4**, 319–21, 447–8, 471, 548, 562–3, 608, 665, 755
- Armstrong, David M. **244–5**, 319, 478, 569, 681
- association **245**, 526
- ataraxia **245**, 655, 736
- Audi, Robert (**109–14**), (223–4), (369–70), 533, (558–9), (672–3), 717
- Augustine, St. **245–7**
- Austin, John L. **247–9**, 250, 252, 301, 769
- avowals **249**, 568, 616, 619, 621, 724
- axiomatization, axiomatics **249–50**, 292, 393, 659, 768
- Ayer, A.J. 45, 46, 49–50, 247–9, **250–2**, 326, 330, 398–9, 517, 547, 708, 782
- Bacon, Francis **253**, 429, 719
- basic belief 29–31, 62–75, 112, **253–4**, 695
- “Bayesian Epistemology” **93–105**, 201, 390, 410, 443, 510, 542, 625, 757
- behaviourism **254**, 283, 322, 478, 620, 649–50, 711
- belief 29–31, 98, 100, **254–9**, 282, 478, 644, 711, 776
- belief in and belief that **259–60**, 696
- Bergmann, Gustav **260–1**
- Berkeley, George **261–4**, 282, 318–9, 326, 328, 426, 429, 528, 586–7, 611, 630, 659, 668, 703, 713, 720
- Blanshard, Brand **264–5**
- BonJour, Laurence (33–43), 44, 52, 60–1, 64–6, (**114–8**), 164, 219, (364–8), 464–5, 478, (638–43), 685
- Brentano, Franz **265–6**, 423, 431, 553
- burden of proof **266**, 277, 451
- Carnap, Rudolf 224–30, 244, **267–8**, 326, 330, 357, 391, 420, 503, 506, 517, 538–40, 550, 649, 722, 748, 782
- Cartesianism 3–4, 158–61, 196–8, 207, 249, **268**, 273–4, 291, 358, 460, 580, 599, 724, 780
- see also* Descartes

- causal theories in epistemology 102, 245, 258,  
**268–72**, 332, 397, 403, 437, 471, 521,  
523, 646, 680, 690, 702  
*see also* reasons/causes
- certainty 265, **272–5**, 305, 309, 398, 400, 406,  
476, 489, 538, 546, 563, 644, 677, 715,  
716, 732
- Chisholm, Roderick 45, 53–62, 65, 146, 169,  
(265–6), **275–7**, 283–4, 298, 304, 328,  
349, 380, 384, 489, 588–9, 626–8, 645,  
679, 693
- circular reasoning 25, 28–9, 30–1, 84, 266,  
**277**, 453
- closure, *see* “Skepticism and Closure”
- cogito **277–8**, 308–9, 394, 462, 600, 720, 780
- cognition, *see* sensation/cognition
- Cohen, Stewart 7–8, 16, (**118–23**), 124,  
(678–81), 686, 688
- coherentism 62, 102, 114–6, 163–4, 171, 223,  
242, 254, 276, **278–81**, 287, 295, 300,  
303, 364–5, 391, 397, 398–400, 406,  
442, 447, 471, 489, 526, 535–6, 584, 623,  
652–3, 675, 722, 756–7, 776  
*see also* “Foundations and Coherence”
- collective belief **281–2**, 82–93
- “The Common Sense Tradition” **53–62**
- commonsensism and critical cognitivism 53–62,  
260, 277, **282–5**, 298, 523, 530, 536,  
580, 627–8, 676, 679, 721
- concepts 229, 281, **285–7**, 368, 494–6, 528,  
702, 731, 756, 760
- Conee, Earl 13, 34, (**123–30**), 342–3, 349,  
(604–6), 686
- “Contextualism” **12–22**, 118–23, 294, 339,  
509–10, 680, 789–91
- Continental epistemology **287–92**, 379, 382,  
389, 407, 514, 524, 551
- contingent, *see* necessary/contingent
- convention **292–3**, 518–9
- conversational implicature **293–4**
- criteria and knowledge 277, **294–8**, 305, 350,  
523, 571, 628
- criterion, canon 294, **298–9**
- critical cognitivism, *see* commonsensism and  
critical cognitivism
- Davidson, Donald 255–8, **300–1**, 434, 492,  
613, 707, 718, 742
- death of epistemology 291, **301–4**, 544, 604,  
678, 707
- defeasibility 113, 156–8, 295, **304–5**, 354, 569,  
614, 645
- definition 292, **305–6**, 566
- Derrida, Jacques 290, **306**
- Descartes, René 93, 231, 257, 268, 277, 294,  
**306–10**, 332, 353, 365, 380, 383, 393,  
398, 419, 425, 438, 455, 457, 462, 500–1,  
544, 567, 600, 610, 659–61, 692–4, 707,  
712, 714–5, 715, 719–20, 747, 749, 780  
*see also* Cartesianism
- Dewey, John 264, 301, **310–2**, 385, 470,  
602–3
- dialectic (Hegel) **312–3**, 407
- dialectic (Plato) **313**, 596
- different constructions in terms of “knows”  
12–22, 70–2, **313–8**, 414, 485, 710–1
- direct realism 34–5, 67–70, 222, 236–40, 245,  
**318–22**, 361, 502, 513, 584–6, 629–31,  
634, 702, 705, 713
- disposition 79–80, 257–8, **322**, 711
- dogmatism **322**, 598, 677
- doxastic voluntarism 77–8, **322–5**
- Dretske, Fred 5–7, 10, (**130–4**), 206, 268,  
(581–6), 646–7, 682, 689, (729–32),  
788–92
- Dutch book argument, *see* “Bayesian  
Epistemology;” probability
- empiricism 43, 47–8, 52, 53–4, 109, 254, 260,  
264, 267, 282, **326–31**, 331, 376, 398,  
429, 437, 455, 469–70, 488, 500, 505–6,  
515–7, 524, 528, 550, 565, 603–4, 616,  
649, 658, 663, 674, 713, 722, 732, 743,  
751, 755, 784
- Epicurus **331–2**
- epistemic deontology 128, **332–6**
- epistemic luck **336–40**, 344, 471, 735
- epistemic supervenience 26–7, 175–6, 276–7,  
**340–3**, 349, 370, 428, 543, 668
- epistemic virtue 147, 339, **343–7**, 348–9, 351,  
368, 532, 696  
*see also* “Virtue Epistemology”
- epoche 291, **347**, 423, 498, 590
- essence (Husserl) **347**, 423
- essence (Plato) **348**
- essentialism, *see* necessary/contingent
- ethics and epistemology 277, **348–9**, 368, 696,  
758
- evidence 86, 88–90, 96, 101, 123–30, 265,  
305, **349–51**, 471, 510, 648, 693,  
695
- evolutionary argument against naturalism  
**351–4**
- evolutionary epistemology **354–6**, 389–90, 392,  
598, 603–4
- existence 267, **356–7**, 540, 563, 676, 713

- experience, theories of 23–4, 43, 62–3, 220–1, 222, 254, 321, **357–61**, 400, 431, 482, 525, 586, 603–4, 634, 705, 726, 731, 732
- explanation **361–4**, 391, 410, 415–6, 447, 512, 542
- external world, *see* problem of the external world
- externalism/internalism 5, 67–72, 130–2, 146, 164–5, 173, 223, 239, 271, 272, 276, 280, 294, 296, **364–8**, 406, 471, 478, 556, 558, 583, 586, 587, 594, 647, 673, 690, 696, 698, 700, 702, 717, 734–5, 758
- see also* “Recent Work on the Internalism–Externalism Controversy”
- fact/value **369–70**, 392, 544, 740
- fallibilism 251, **370–5**, 406, 437, 444, 527, 555–6, 580, 598, 674
- Feldman, Richard (12–22), 34, 86, (**123–30**), 146, 334, 342–3, (349–51), 396, 686
- feminist epistemology 291, **375–80**
- first philosophy **380**, 543, 545
- Firth, Roderick **380–1**, 590
- Foley, Richard (**134–9**), (275–7), (756–8)
- for itself, *see* in itself/for itself
- Foucault, Michel 290, **381–2**, 741
- foundationalism 62, 110, 114–6, 139–44, 163–4, 223, 232, 241, 251, 254, 276, 302–3, 350, 364, 380, **382–5**, 406, 447–8, 456, 471, 482, 489, 535–6, 543, 550, 587, 616, 621, 649, 652, 675, 694, 709, 722, 732, 751, 756
- “Foundations and Coherence” **22–33**, 196–7, 254, 281, 385
- Frege, Gottlob 45, 48–9, 224–30, **385–7**, 423, 436, 444, 461–2, 472, 492, 506, 516, 649, 785
- Freud, Sigmund 362, **387–8**, 430, 650–2, 724, 741, 743
- Fumerton, Richard (**139–44**), (236–40), (445–7), (586–90), 686
- Gadamer, Hans-Georg 289, **389**, 741
- genetic epistemology 356, **389–92**, 604, 747
- genetic fallacy **393**
- geometry 250, **393–5**, 418, 473, 502, 516, 519, 538, 663, 674, 722
- Gettier problem 134, 156–8, 170–2, 251, 281, 296, 305, 336–40, 373, **395–7**, 471, 556, 644–6, 648, 680, 682, 716, 718, 757, 772
- given, the 141, 252, 254, 385, **397–400**, 409, 431, 447, 457, 460, 482, 489, 616, 705, 722, 723, 726, 727–8, 731, 732
- Goldman, Alvin I. 36–9, (82–93), 126, 132, (**144–52**), 164, 269–72, 344–6, 351, 545, (681–92), 717
- Goodman, Nelson **400–2**, 440, 538–9, 628, 642, 643, 666
- Grice, H. Paul 144, 293–4, **402–3**, 530, 664
- Habermas, Jürgen 289, **404–5**, 741, 743
- Harman, Gilbert (**152–6**), (438–44), 689
- Hegel, G.W.F. 287–8, 291, 312–3, 389, **405–8**, 417, 432–3, 669
- Heidegger, Martin 400, **408–9**, 424, 590, 713
- Hempel, Carl Gustav 362–3, 391, **409–10**, 410–2, 415, 517, 540, 666
- Hempel’s paradox of the ravens 244, 409, **410–2**, 539, 666
- hermeneutics 289, 389, 400, 408–9, **412–3**, 651, 741
- Hintikka, Jaakko (313–8), **413–4**, 476, (483–6), 531
- historical knowledge 364, **414–6**, 500, 523, 743, 781
- historicism 291, **416–7**, 598
- Hobbes, Thomas 330, **417–20**, 662, 720
- holism 30, 47–8, 304, 328, 330, **420**, 539, 700–1, 728, 742–3, 779
- Hume, David 53–4, 85, 206, 208, 238, 245, 250–1, 255, 282–3, 294, 326–7, 330, **420–3**, 456, 458, 472, 489, 505, 517, 526, 538, 586–7, 626–7, 638–9, 643, 652, 659, 665, 674, 675, 720–1, 725, 754, 764, 770, 783
- Husserl, Edmund 291, 306, 347, **423–4**, 436, 524, 567, 590, 649, 713, 748
- idea 261–3, **425**, 616, 713, 732
- idealism 264, 264–5, 311, 319, 404, 406, 407, **425–9**, 475, 524, 530, 553, 601, 616, 668–9, 671, 706, 707, 713
- ideology 288, **429–31**, 741, 743
- illusion, *see* argument from illusion
- immediacy, presence **431–2**
- in itself/for itself 289, 291, 407, **432–3**, 475, 513, 554, 559, 561, 713, 721
- incorrigibility 384, 398, **433**, 444
- indeterminacy of reference **433**, 564
- indeterminacy of translation 433, **434**, 658, 742
- Indian epistemology **434–8**, 459, 556
- indubitability 384, 398, **438**, 444

- induction: enumerative and hypothetical 221, 235, 244, 253, 401, **438–44**, 447, 527, 535, 580, 600
  - induction, problem of, *see* problems of induction
- infallibility 223, 384, 398, 433, **444**, 732
- inference 28–9, 30, 238–9, **444–5**, 461
  - to the best explanation 244, 279, 364, 384, 439, 443, **445–7**, 449–50, 604, 617, 632, 634, 651, 704
- infinite regress argument 22, 114–6, 161–2, 232, 241, 254, 350, 368, 383, 385, 398, **447–50**, 489
- informal fallacies 235, 266, 277, 393, **450–4**
- innate ideas **454–6**, 488, 500
- internalism, *see* externalism/internalism
  - see also* “Recent Work on the Internalism–Externalism Controversy”
- introspection 26, 165, 245, **456–60**, 482, 568, 571, 617, 621, 711, 723, 726
- intuition and deduction 229, 330, 384, 387, 444, **460–3**, 503, 506, 527, 533, 580, 600
- intuition in epistemology 53–62, 148–9, **464–8**, 502
- isostheneia **468**, 655–6, 736–7
- James, William 301, 326, **469–70**, 479–82, 524, 581, 599, 601–3
- judgement **470–1**
- justification 3–12, 22–33, 33–43, 57–61, 62–75, 109–14, 117–8, 118–9, 123–5, 137–8, 150, 185–90, 195, 202, 270–2, 275–7, 278–81, **471**, 682–4, 687–8, 692–7, 732–6, 753
- Kant, Immanuel 43–53, 191, 193, 219, 224–30, 231, 287, 322, 330, 393–4, 398, 405–6, 408, 413–4, 422, 427, 429, 432–3, 455, 457–8, 470–1, **472–6**, 505, 516–7, 525, 533, 538, 552–4, 652, 660, 665, 674, 707, 713, 721, 722, 726, 758, 768–9
- KK-thesis 413, 435, **476**
- Klein, Peter 8, (**156–63**), (272–5), (714–5), (715–9)
- knower paradox, *see* paradox of the knower
- knowledge
  - and belief 245, **476–9**, 644, 648, 718, 772
  - by acquaintance/by description 236, 312, 317, 400, 431, 470, 474, **479–82**, 709, 722, 723, 785
  - how, who, why, etc., *see* different constructions in terms of “knows”
  - seeking by questioning 413–4, **483–6**
  - see also* causal theories in epistemology; coherentism; different constructions in terms of “knows;” foundationalism; Gettier problem; knowledge and belief; propositional knowledge; reliabilism
- Kornblith, Hilary 125, (**163–8**), 351, 465–6, (543–6), 689
- Kripke, Saul 5, 9, 45, 132–3, 143, 286, 338–9, 635, 637–8
- Kuhn, Thomas 84, 103, 355, 541–2, 739, 743, 746
- Lehrer, Keith 59–60, 157, (**168–73**), 257, (278–281), 341–2, 396, 476, 644–5, (674–6), 685
- Leibniz, Gottfried Wilhelm 231, 250, 265, 426, 455, 472, 474, **487–8**, 546–7, 615, 659, 661–2, 738, 777–8
- Lewis, C.I. 264, 326, 330, 380, 398, **489**, 510, 517, 589, 603
- limits of human knowledge **490–2**
- linguistic understanding 287, **492–7**, 566, 569, 731, 760, 776
- literature and knowledge **497–500**
- Locke, John 224–30, 245, 262–4, 321, 326, 330, 332–3, 384, 425, 429, 455, 457, 462, 474, 488, **500–3**, 528, 610–1, 659, 692–4, 703, 707, 720, 763
- logical construction 250, 267, **503–4**, 538, 567, 708
- logical empiricism, *see* logical positivism; Reichenbach
- logical positivism 250, 330, 409, 463, **504–6**, 517, 540, 547, 549, 587, 590, 591, 649, 674, 722, 740, 780, 782, 784
- logicism 386, 463, **506**, 516
- lottery paradox 24, 170, 374, **506–11**, 532, 572, 605, 666, 735, 761
- luck, epistemic, *see* epistemic luck
- Marxism 288, 404, 429–31, **512–4**, 741, 743
- mathematical knowledge 45–9, 229, 326, 395, 413, 475, 506, **514–20**, 527, 560, 600, 722
- memory 78, 112, 125–6, 272, 414, 416, 489, **520–4**, 695
- mention, *see* use/mention
- Merleau-Ponty, Maurice 424, **524–5**, 590
- methodology **525**
- Mill, John Stuart 45, 47, 235, 326, 328, 516, **525–7**, 567, 587, 739
- modal knowledge, *see* necessity, modal knowledge
- Molyneux’s problem 262, **527–8**, 613

- Montaigne, Michel de **528–9**, 719, 737
- Moore, G.E. 8–9, 53–62, 206, 209, 222, 244, 260, 282–4, 427, 474, **529–31**, 533, 553, 574–5, 627, 787–91
- Moore's paradox 413, 529, **531–2**, 572, 761
- moral epistemology 277, **532–7**, 676
- myth of the given, *see* foundationalism; the given; Sellars
- naïve realism, *see* argument from illusion; direct realism; problem of the external world
- natural science, epistemology of 253, 267, 390, **538–42**, 598, 653, 673, 739, 743, 747
- naturalism 145–6, 148, 163, 174–8, 249, 404, 406, 408, 429, 515, 525, **542–3**, 613, 635, 689
- naturalized epistemology 229, 272, 304, 328, 355–6, 370, 376, 392, 403, 530, **543–6**, 550, 593, 598, 604, 652, 657–8, 690, 709
- necessary/contingent 43–5, 306, 330, 488, **546–7**, 547, 598, 778
- necessity, modal knowledge 43–5, 547, **547–8**
- Neoplatonism 425, **548–9**
- Neurath, Otto 506, **549–51**, 649, 722, 771, 782
- Nietzsche, Friedrich 288, 291, 382, 400, **551–2**, 652, 741
- nihilism, *see* death of epistemology
- noumenal/phenomenal 433, 475, **552–4**, 559, 561, 613, 671, 721
- Nozick, Robert 4–5, 9, 132, 338–9, **554–6**, 647, 681, 717, 732–4, 791
- Nyāya 434–8, **556–7**
- objective/subjective 244, 274, **558–9**, 561, 613, 758, 760
- objectivity 378, 389, 405, 429, 512, 559, **559–62**, 613, 671, 758, 760, 786
- Ockham, William of 402, **562–3**, 756
- ontological commitment 357, **563–4**, 564, 658
- ontological relativity 433, 563, **564**, 658, 678, 742
- ostensive definition 496, **564–6**, 616–9, 621, 638
- other minds 54–7, 235–6, 245, 249, 250, 263, 388, **566–71**, 617, 621, 695, 711, 713, 748–9, 754, 759, 787
- paradox 120–1, 204, **572**, 577, 606, 760–1
- paradox of the knower 532, 572, **572–4**, 761
- paradoxes of analysis 572, **574–7**, 588
- paranormal knowledge **577–9**
- Peirce, Charles S. 221, 370, 427, 445, 469–70, 526, **580–1**, 599–603
- perceptual knowledge 78, 112, 125, 130–4, 241–2, 272, 321, 385, 525, 528, **581–6**, 621, 651, 705
- see also* “The Power of Perception”
- perspectivism, *see* Nietzsche
- phenomenalism 238–9, 247–9, 250–2, 264, 277, 283, 328, 380, 400, 470, 489, 503, 527, 538, 550, 553, 567, **586–90**, 632–4, 649, 674, 709
- see also* noumenal/phenomenal
- phenomenology 220, 289, 347, 408, 423, 524, **590**, 713
- philosophical knowledge 229, 506, **590–3**, 658, 780
- Plantinga, Alvin 137, (**173–8**), 219, (223), (322), 333–4, 351–4, 384, 546, (692–7), 697
- Plato 77, 221–2, 230, 240–2, 246, 254–5, 313, 348, 393, 455, 477, 515, 548–9, **593–7**, 608, 665, 771–2
- Pollock, John (**178–85**), 304–5, 605, 645–6
- Popper, Karl 156, 355, 362, 417, 539, **597–9**, 640, 666, 738, 742, 782
- positivism, *see* logical positivism
- “The Power of Perception” **62–75**
- pragmatism 149, 155, 264, 273, 301, 310–1, 385, 469–70, 471, 517–8, 524, 527, 551, 580–1, **599–604**, 655, 657–8, 705–6, 707, 773
- preface paradox 508, 532, 572, **604–6**
- presence, *see* immediacy
- presocratic epistemology **606–9**
- prima facie reasons 69–70, 110, 266, 305, **609**, 614, 673, 739
- primary and secondary qualities 262–3, 264, 321, 501, 528, 552, 554, 559, **609–13**, 671, 703
- principle
- of charity 532, 564, **613**, 742
  - of contradiction 230, **613–4**
  - of credulity **614–5**, 676
  - of identity **615–6**
- private language argument 85, 236, 296, 330, 385, 460, 566, 567, 571, **616–21**, 634, 638, 651, 786–7
- probability, theories of 7, 23, 93–105, 156, 201, 256–7, 265, 267, 507, 539, 542, **622–6**, 739



- problem
  - of the criterion 277, 283–4, 298, 326, 407, 523, 536, **626–8**, 678
  - of the external world 54–7, 62–75, 239, 251, 321, 327, 447, 502, 587, 590, **628–34**, 703, 709
  - of induction, *see* problems of induction
  - of other minds, *see* other minds
  - of rule-following 296, 444, **634–8**, 787
- problems of induction 244, 327, 401, 422, 440, 443, 505, 526, 538–9, 542, 580, 597, 614, **638–43**, 643, 665, 674, 753–4
  - see also* induction
- projection, projectibility 401, 440, 642, **643**
- proof 235, 444, **643–4**
- propositional knowledge 374, 396–7, 471, 478, 556, **644–8**, 718, 771–2
- protocol sentences 267, 400, 505, 550, **649**
- psychologism 386, 423, 436, 526, 545, **649**
- psychology and epistemology 254, 392, 403, 545, 593, **649–54**, 658, 667
- Putnam, Hilary 45, 47, 51, 224–30, 407, 604, **654–5**, 718, 743
- Pyrrhonism 245, 529, **655–6**, 714–5, 720, 736–7
  
- Quine, W.V. 45, 47–8, 50–1, 146, 224–30, 255–6, 301, 303–4, 306, 326, 328, 355, 357, 380, 420, 433, 434, 474, 503, 517–8, 539, 543–5, 547, 550, 563, 564, 603, 642, 652–3, **657–8**, 707, 738, 742–3, 782
  
- rationalism 109–14, 116–7, 329, 393, 406, 407, 422, 437, 475, 487, 500, 533, 603, 616, **659–63**, 713, 743, 749
- rationality 90–1, 109, 134–9, 178–85, 404, 510, 545, **663–8**, 378, 740
- realism 111, 241–2, 244–5, 255–6, 303, 311, 318, 355, 406, 429, 470, 475, 496, 560–1, 581, 604, 654–5, **668–72**, 674, 702, 707, 712, 713, 780, 785
- reasons/causes 362, 651, 653, **672–3**
- “Recent Work on the Internalism–Externalism Controversy” **33–43**, 67–72, 114, 272, 690
  - see also* externalism/internalism
- regress argument, *see* infinite regress argument
- Reichenbach, Hans 395, 539, 603, 624, 639–40, 642, **673–4**, 683, 782
- Reid, Thomas 53–62, 63–4, 224, 277, 282–4, 627, **674–6**, 721, 763
- reification, hypostatization **676**
- relativism 84, 405, 407, 447, 471, 551, 564, **676–8**, 742, 747, 758
- relevant alternatives 144, 251, 269, **678–81**
- reliabilism 28–9, 34–6, 63, 127–8, 144–8, 244, 251, 270–2, 296, 327, 344–5, 351, 365, 384, 397, 471, 522–3, 535, 545, 556, 558, 563, 647, 673, 680, **681–92**, 718, 735, 756
- religious belief, epistemology of 223, 260, **692–7**, 757
  - recent developments **697–9**
- representation 304, 361, 616, **699–702**
- representative realism 64–7, 222, 239, 304, 318, 321, 361, 403, 470, 502–3, 513, 584–6, 589, 630, 634, **702–5**, 732
- Rescher, Nicholas 604, **705–6**
- Rorty, Richard 84, 301–2, 311, 417, 603, **706–7**, 743
- rule-following, *see* problem of rule-following
- Russell, Bertrand 45, 46, 130–1, 140, 143, 244, 267, 273, 317, 320, 326, 327, 329, 414, 474, 479, 481–2, 489, 503, 506, 516, 522, 538, **707–9**, 712, 747–8, 785
- Ryle, Gilbert 247, 258, 283, 322, 458–60, 568, 593, **709–11**, 724, 769
  
- safety, *see* sensitivity and safety
- Santayana, George 709, **712–3**, 748, 783
- Sartre, Jean-Paul 289, 321, 424, 432–3, **713–4**
- scepticism 19–20, 22, 113, 120–1, 126–7, 174, 191–3, 194–9, 204–10, 238, 241–3, 245, 249, 263, 272–5, 288, 294, 300, 302, 326, 336–7, 355, 407, 447, 475, 543–4, 556, 604, 633–4, 656, 679–80, 685, **714–5**, 715, 718, 735, 736–7, 788–9
  - contemporary 158–61, 310, 368, 374, 530, 545, 634, 655, 715, **715–9**, 771
  - modern 419, 422, 502, **719–21**, 737
  - see also* “Skepticism and Closure”
- Schlick, Moritz 267, 506, 550, 649, **722–3**, 782
- scientia media, *see* Suárez
- self-consciousness 460, **723**
- self-evidence 109, 250, **723–4**
- self-knowledge and self-identity 245, 249, 277, 306, 310, 422, 460, 475, 571, 621, 723, **724–6**
- self-presenting **726–7**
- Sellars, Wilfrid 170, 199, 301, 306, 330, 385, 398–9, 603–4, 707, **727–9**, 751
- sensation/cognition 321, 361, 400, 525, 586, 726, **729–32**, 751
- sense-data 56–7, 62–75, 222, 236, 239, 248–9, 250, 306, 320, 358–9, 361, 403, 481, 503, 567, 586, 649, 708, 731, **732**, 784

- sensitivity and safety 4–5, 9, 96, 338–9, 344, 556, 681, **732–6**, 791
- Sextus Empiricus 243, 245, 298, 332, 407, 468, 528, 628, 656, 718, 719, 721, **736–8**, 751
- simplicity 440–2, **738–9**
- skepticism, *see* scepticism
- “Skepticism and Closure” **3–12**, 14–16, 120–1, 132–4, 338, 508–9, 556, 718
- “Social Epistemology” **82–93**, 103, 112, 149–50, 690
- social sciences, epistemology of 88–90, 364, 382, 389, 404, 405, 412, 416, 431, 500, 542, 564, 613, 653, 673, **739–44**, 758
- sociology of knowledge 390, 405, 430, 542, 544, 678, **744–7**
- Socrates, *see* Plato
- solipsism 236, 267, 512, 571, 617, **747–9**, 786–7
- Sosa, Ernest 8–9, 52, 75–82, 86, (**185–90**), 212–3, 344–6, 349, 645, 681, 690, 734–5
- Spinoza, Benedict 250, 659, 661, **749–50**
- Stoic epistemology 222, 655, 736–7, **750–3**, 763
- Strawson, P.F. 475, 538, 567–8, 639, 641, **753–4**, 769–70
- strong programme 83, 678, 744, **754–5**
- Stroud, Barry (**190–4**), (245), (420–3), (504–6), (609–13), 770, (780)
- Suárez, Francisco **755–6**
- subjective, *see* objective/subjective
- subjectivism 289, 407, 471, 678, **756–8**
- subjectivity 415, 460, 559, 561, 613, **758–60**
- supervenience, *see* epistemic supervenience
- surprise examination paradox 510, 532, 572, **760–2**
- tabula rasa 500, 751, **763**
- testimony 80–1, 85–7, 90, 112, 246, 523, 614, 676, **763–8**
- theory 249–50, **768**
- transcendental arguments 191, 405, 475, 571, 634, 678, 718, 753–4, **768–71**, 780
- tripartite definition of knowledge 132, 391, 395, 536, 644, **771–2**
- truth, theories of 12–22, 87–8, 99, 118–9, 264, 281, 290, 304, 493, 496, 655, 671, 750, **772–7**, 780
- truths of reason/truths of fact 662, **777–8**
- underdetermination of theory 4, 657, 745, **779**
- use/mention 305, **779**
- value, *see* fact/value
- verificationism 250, 267, 409, 469, 505, 587, 591–3, 654, 702, 722, 739, 770–1, 773, 776, **780**, 782
- Vico, Giambattista 416, **780–1**
- Vienna Circle 260, 267, 506, 722, **782**, 784
- “Virtue Epistemology” **75–82**, 210–5, 223, 690, 696  
*see also* epistemic virtue
- Whitehead, Alfred North 506, **783**
- Williams, Michael 84, (**194–9**), (264–5), (301–4), (706–7)
- Williamson, Timothy (**199–204**), 259, 374, 681
- Wittgenstein, Ludwig 84, 208, 235–6, 249, 273, 286, 301, 330, 413, 444, 459, 492, 506, 519, 530, 536, 565, 567–71, 621, 634, 638, 649, 678, 722, 724, 746, 748–9, 754, 769, 773, 782, **783–7**
- Wright, Crispin 8, (**204–10**)
- Zagzebski, Linda (**210–5**), 345–6, 687
- zebras and cleverly disguised mules 5–6, 338, 680, **788–92**