

Contemporary Debates In Philosophy

Second Edition

# Contemporary Debates in Epistemology

*Edited by*

*Matthias Steup,  
John Turri,  
and Ernest Sosa*

**WILEY** Blackwell



## Praise for *Contemporary Debates in Epistemology 2e*

"When Blackwell published the first edition of *Contemporary Debates in Epistemology* in 2005, that volume very quickly became epistemology's superego: it expressed the ideals that were implicit in the best recent epistemological work, and it served to guide both practicing and apprenticing epistemologists to the questions that mattered most to the field back then. Of course, the questions that matter most in 2013 are not exactly the same as those that mattered most in 2005; thus, the need for a new edition. I expect that this new edition – which contains units on the now widely discussed issues of whether knowledge is epistemologically fundamental, whether practical concerns encroach on epistemic status, whether evidential justification is permissive, what sort of epistemic luck (if any) is incompatible with knowledge – will serve to guide epistemological practice for the next several years, and with at least as much authority as the first edition. The contributors are among the most prominent in the field, and their contributions represent some of the best work being done on the topics under discussion. If you want to contribute to the most important epistemological conversations today, you will need to read this book."

– Ram Neta, *University of North Carolina, Chapel Hill*

## Praise for the First Edition

"This book is packed with cutting-edge epistemology by excellent contributors to the field. It is both comprehensive and admirably brief."

– Robert Audi, *University of Notre Dame*

"What are the burning problems of today's epistemology? What are the most promising solutions to these problems? They are all in this timely volume, explained and debated by leading authorities."

– Alvin Goldman, *Rutgers University*

"With leading and emerging figures in epistemology debating some of its most fundamental questions, this volume will be required reading for anyone interested in where the theory of knowledge has been and where it is going. A superb collection."

– Paul Boghossian, *New York University*

## Contemporary Debates in Philosophy

In teaching and research, philosophy makes progress through argumentation and debate. Contemporary Debates in Philosophy provides a forum for students and their teachers to follow and participate in the debates that animate philosophy today in the western world. Each volume presents pairs of opposing viewpoints on contested themes and topics in the central subfields of philosophy. Each volume is edited and introduced by an expert in the field, and also includes an index, bibliography, and suggestions for further reading. The opposing essays, commissioned especially for the volumes in the series, are thorough but accessible presentations of opposing points of view.

1. Contemporary Debates in Philosophy of Religion  
*edited by Michael L. Peterson and Raymond J. Vanarragon*
2. Contemporary Debates in Philosophy of Science  
*edited by Christopher Hitchcock*
3. Contemporary Debates in Epistemology  
*edited by Matthias Steup and Ernest Sosa*
4. Contemporary Debates in Applied Ethics  
*edited by Andrew I. Cohen and Christopher Heath Wellman*
5. Contemporary Debates in Aesthetics and the Philosophy of Art  
*edited by Matthew Kieran*
6. Contemporary Debates in Moral Theory  
*edited by James Dreier*
7. Contemporary Debates in Cognitive Science  
*edited by Robert Stainton*
8. Contemporary Debates in Philosophy of Mind  
*edited by Brian McLaughlin and Jonathan Cohen*
9. Contemporary Debates in Social Philosophy  
*edited by Laurence Thomas*
10. Contemporary Debates in Metaphysics  
*edited by Theodore Sider, John Hawthorne, and Dean W. Zimmerman*
11. Contemporary Debates in Political Philosophy  
*edited by Thomas Christiano and John Christman*
12. Contemporary Debates in Philosophy of Biology  
*edited by Francisco J. Ayala and Robert Arp*
13. Contemporary Debates in Bioethics  
*edited by Arthur L. Caplan and Robert Arp*
14. Contemporary Debates in Epistemology, Second Edition  
*edited by Matthias Steup, John Turri, and Ernest Sosa*

# Contemporary Debates in Epistemology

SECOND EDITION

Edited by

*Matthias Steup*

*John Turri*

*Ernest Sosa*

**WILEY** Blackwell

This edition first published 2014  
© 2014 John Wiley & Sons, Inc  
Contemporary Debates in Epistemology, 1st edition, published in 2005

Wiley-Blackwell is an imprint of John Wiley & Sons, formed by the merger of Wiley's global Scientific, Technical and Medical business with Blackwell Publishing.

*Registered Office*

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

*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 Matthias Steup, John Turri, and Ernest Sosa to be identified as the authors of the editorial material in this work has been asserted in accordance with the UK 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.

**Limit of Liability/Disclaimer of Warranty:** While the publisher and author(s) have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. It is sold on the understanding that the publisher is not engaged in rendering professional services and neither the publisher nor the author shall be liable for damages arising herefrom. 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*

Contemporary debates in epistemology / edited by Matthias Steup, John Turri,  
Ernest Sosa. – Second Edition.

pages cm. – (Contemporary debates in philosophy ; 2408)

Includes bibliographical references and index.

ISBN 978-0-470-67209-9 (pbk. : alk. paper) 1. Knowledge, Theory of. I. Steup, Matthias,  
editor of compilation.

BD161.C6545 2013

121-dc23

2013006412

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

Cover design by Cyan Design: [www.cyandesign.co.uk](http://www.cyandesign.co.uk)

Set in 10/12.5pt Rotis Serif by SPi Publisher Services, Pondicherry, India

# Contents

Notes on Contributors	viii
Preface to the Second Edition	xiii
Preface to the First Edition	xiv
<b>1 Should Knowledge Come First?</b>	<b>1</b>
Knowledge First <i>Timothy Williamson</i>	1
What Is Knowledge-first Epistemology? <i>Trent Dougherty</i> <i>and Patrick Rysiew</i>	10
Experience First <i>Trent Dougherty and Patrick Rysiew</i>	17
Knowledge Still First <i>Timothy Williamson</i>	22
Still Nowhere Else to Start <i>Trent Dougherty and Patrick Rysiew</i>	25
<b>2 Is Knowledge Closed under Known Entailment?</b>	<b>27</b>
The Case against Closure <i>Fred Dretske</i>	27
The Case for Closure <i>John Hawthorne</i>	40
Reply to Hawthorne <i>Fred Dretske</i>	56
<b>3 Is Knowledge Contextual?</b>	<b>60</b>
Contextualism Contested <i>Earl Conee</i>	60
Contextualism Defended <i>Stewart Cohen</i>	69
Contextualism Contested Some More <i>Earl Conee</i>	75
Contextualism Defended Some More <i>Stewart Cohen</i>	79
<b>4 Do Practical Matters Affect Whether You Know?</b>	<b>84</b>
Practical Matters Affect Whether You Know <i>Jeremy Fantl</i> <i>and Matthew McGrath</i>	84
Practical Matters Do Not Affect Whether You Know <i>Baron Reed</i>	95

<b>5 Can Skepticism Be Refuted?</b>	<b>107</b>
The Refutation of Skepticism <i>Jonathan Vogel</i>	108
The Challenge of Refuting Skepticism <i>Richard Fumerton</i>	120
<b>6 Are Intellectually Virtuous Motives Essential to Knowledge?</b>	<b>133</b>
Knowledge Need Not Be Virtuously Motivated <i>Jason Baehr</i>	133
Knowledge and the Motive for Truth <i>Linda Zagzebski</i>	140
Reply to Zagzebski <i>Jason Baehr</i>	146
Reply to Baehr <i>Linda Zagzebski</i>	149
<b>7 Can Knowledge Be Lucky?</b>	<b>152</b>
Knowledge Cannot Be Lucky <i>Duncan Pritchard</i>	152
Knowledge Can Be Lucky <i>Stephen Hetherington</i>	164
<b>8 Is There a Priori Knowledge?</b>	<b>177</b>
In Defense of the a Priori <i>Laurence Bonjour</i>	177
There Is No a Priori <i>Michael Devitt</i>	185
Reply to Devitt <i>Laurence Bonjour</i>	195
Reply to Bonjour <i>Michael Devitt</i>	197
Last Rejoinder <i>Laurence Bonjour</i>	200
<b>9 Is There Immediate Justification?</b>	<b>202</b>
There Is Immediate Justification <i>James Pryor</i>	202
There Is No Immediate Justification <i>Juan Comesaña</i>	222
Reply to Comesaña <i>James Pryor</i>	235
Reply to Pryor <i>Juan Comesaña</i>	239
<b>10 Can Belief Be Justified Through Coherence Alone?</b>	<b>244</b>
Non-foundationalist Epistemology: Holism, Coherence, and Tenability <i>Catherine Z. Elgin</i>	244
Why Coherence Is Not Enough: A Defense of Moderate Foundationalism <i>James Van Cleve</i>	255
Reply to Van Cleve <i>Catherine Z. Elgin</i>	267
Reply to Elgin <i>James Van Cleve</i>	271
<b>11 Is Infitism the Solution to the Regress Problem?</b>	<b>274</b>
Infitism Is the Solution to the Regress Problem <i>Peter Klein</i>	274
Infitism Is Not the Solution to the Regress Problem <i>Carl Ginet</i>	283
Reply to Ginet <i>Peter Klein</i>	291
Reply to Klein <i>Carl Ginet</i>	295
<b>12 Can Evidence Be Permissive?</b>	<b>298</b>
Evidence Can Be Permissive <i>Thomas Kelly</i>	298
Evidence Cannot Be Permissive <i>Roger White</i>	312



13 Is Justification Internal?	324
Justification Is Not Internal <i>John Greco</i>	325
Justification Is Internal <i>Richard Feldman</i>	337
14 Is Truth the Primary Epistemic Goal?	351
Truth Is Not the Primary Epistemic Goal <i>Jonathan L. Kvanvig</i>	352
Truth as the Primary Epistemic Goal: A Working Hypothesis <i>Marian David</i>	363
Index	378

# Notes on Contributors

**Jason Baehr** is Associate Professor of Philosophy at Loyola Marymount University in Los Angeles. He works mainly at the intersection of virtue theory and epistemology and is author of *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology* (Oxford University Press, 2011). He is currently directing the “Intellectual Virtues and Education Project,” a three-year grant project involving the application of philosophical models of intellectual virtue to educational theory and practice.

**Laurence Bonjour** is Professor of Philosophy at the University of Washington. He has published numerous articles in epistemology and the philosophy of mind and is the author of *The Structure of Empirical Knowledge* (Harvard University Press, 1985), *In Defense of Pure Reason* (Cambridge University Press, 1998), *Epistemology: Classic Problems and Contemporary Responses* (Rowman & Littlefield, 2002, 2010), and (together with Ernest Sosa) *Epistemic Justification: Internalism versus Externalism, Foundationalism versus Virtues* (Blackwell, 2003).

**Stewart Cohen** is Professor of Philosophy at the University of Arizona and the University of St Andrews. He is also the editor of *Philosophical Studies* and has published numerous articles in epistemology.

**Juan Comesaña** received his PhD from Brown University, taught for several years at the University of Wisconsin-Madison, and is now Associate Professor of Philosophy at the University of Arizona. He works mainly in epistemology, and has published papers on reliabilism, safety, disjunctivism, the internalism/externalism debate, and disagreement, among other topics.

**Earl Conee** is a Professor of Philosophy at the University of Rochester. He is the author, with Richard Feldman, of *Evidentialism: Essays in Epistemology* (Oxford University Press, 2004), and in addition to advocating invariantism in the present work, he defends other old-fashioned epistemological views in other writings.

**Marian David** is Professor for Theoretical Philosophy at the Karl-Franzens Universität Graz, Austria. He has published articles in the philosophy of language and in epistemology. He is the author of *Correspondence and Disquotation: An Essay on the Nature of Truth* (Oxford 1994) and co-editor of the journal *Grazer Philosophische Studien*.

**Michael Devitt** is a Distinguished Professor of Philosophy at the Graduate Center at CUNY. His main research interests are in the philosophy of language and linguistics, and in issues of realism. He is the author of *Designation* (Columbia University Press, 1981), *Coming to Our Senses* (Cambridge University Press, 1996), *Realism and Truth* (Princeton University Press, 1997), *Language and Reality* (with Kim Sterelny, MIT Press, 1999), *Ignorance of Language* (Oxford University Press, 2006), and the editor (with Richard Hanley) of the *Blackwell Guide to the Philosophy of Language* (Blackwell, 2006). His most recent book is *Putting Metaphysics First* (Oxford University Press, 2010).

**Trent Dougherty** is Assistant Professor of Philosophy at Baylor University, Texas. He publishes frequently in epistemology and philosophy of religion. His central interest in philosophy is taking probability and personhood seriously, separately and together. He is editor of *Evidentialism and Its Discontents* (Oxford University Press, 2011) and (with Justin McBrayer) *Skeptical Theism: New Essays* (Oxford University Press, 2012).

**Fred Dretske** is Professor Emeritus, Stanford University and the University of Wisconsin. He is presently Research Scholar (occasionally doing some teaching) at Duke University. His interest in the past 10 years has centered on the intersection of epistemology and philosophy of mind – in particular, self-knowledge. A collection of his essays on these topics, *Perception, Knowledge, and Belief*, was published in 2000 by Cambridge University Press.

**Catherine Z. Elgin** is professor of the philosophy of education at Harvard Graduate School of Education. She is the author of *Considered Judgment* (Princeton University Press, 1996), *Between the Absolute and the Arbitrary* (Cornell University Press, 1997), *With Reference to Reference* (Hackett, 1983), and co-author with Nelson Goodman of *Reconceptions* (Hackett, 1998). Her current research investigates how scientific models and other representations that are not and are known not to be true of the phenomena they pertain to nonetheless figure in our understanding of those phenomena.

**Jeremy Fantl** is Associate Professor of Philosophy at the University of Calgary. He has published numerous articles in epistemology and is co-author, with Matthew McGrath, of *Knowledge in an Uncertain World* (Oxford University Press, 2009). He is also co-editor of *Epistemology: An Anthology* (Blackwell, 2008).

**Richard Feldman** is Professor of Philosophy at the University of Rochester. His interests are in epistemology and metaphysics. His publications include *Reason and Argument* (Prentice Hall, 1993; 2nd edition, 1999), *Epistemology* (Prentice Hall, Foundations of Philosophy Series, 2003), *Evidentialism* (with Earl Conee, Oxford University Press, 2004),

*The Good, The Right, Life and Death* (edited with Jason Raibly, Kris McDaniel, and Michael Zimmerman, Ashgate, 2006), and *Disagreement* (edited with Ted A. Warfield, Oxford 2010).

**Richard Fumerton** is Professor of Philosophy at the University of Iowa. He has published numerous articles in epistemology and is the author of *Knowledge, Thought, and the Case for Dualism* (Cambridge, forthcoming), *Epistemology*, (Blackwell, 2006), *Realism and the Correspondence Theory of Truth* (Rowman & Littlefield, 2002), *Metaepistemology and Skepticism* (Rowman & Littlefield, 1996), and *Metaphysical and Epistemological Problems of Perception* (University of Nebraska, 1985).

**Carl Ginet** is Professor of Philosophy Emeritus at Cornell University. He has published *Knowledge, Perception, and Memory* (D. Reidel, 1975) and *On Action* (Cambridge University Press, 1990) and articles on epistemology and free will and action. He is currently giving his thought primarily to the topic of a priori justification.

**John Greco** is the Leonard and Elizabeth Eslick Chair in Philosophy at Saint Louis University. His previous publications include *Putting Skeptics in Their Place: The Nature of Skeptical Arguments and Their Role in Philosophical Inquiry* (Cambridge University Press, 2000) and *Achieving Knowledge: A Virtue Theoretic Account of Epistemic Normativity* (Cambridge University Press, 2010). He is also editor of *The Oxford Handbook of Skepticism* (Oxford University Press, 2008), *Sosa and His Critics* (Blackwell, 2004), *The Blackwell Guide to Epistemology* (1999), and *Virtue Epistemology: Contemporary Readings* (2012).

**John Hawthorne** is Waynflete Professor of Metaphysical Philosophy at Magdalen College, Oxford, and the author of *Knowledge and Lotteries* (Oxford University Press, 2004).

**Stephen Hetherington** is Professor of Philosophy at the University of New South Wales. His monographs include *Good Knowledge, Bad Knowledge* (Oxford University Press, 2001) and *How To Know* (Wiley-Blackwell, 2011). His introductory books include *Reality? Knowledge? Philosophy!* (Edinburgh University Press, 2003), *Self-Knowledge* (Broadview, 2007), and *Yes, But How Do You Know?* (Broadview, 2009). His edited volumes include *Epistemology Futures* (Oxford University Press, 2006) and *Epistemology: The Key Thinkers* (Continuum, 2012).

**Thomas Kelly** is Professor of Philosophy at Princeton University. His publications include "The Epistemic Significance of Disagreement" (*Oxford Studies in Epistemology* 1 2005), the entry on "Evidence" in the *Stanford Encyclopedia of Philosophy*, and, with Sarah McGrath, "Is Reflective Equilibrium Enough?" (*Philosophical Perspectives* 2010).

**Peter Klein** received his PhD from Yale University and has taught at Rutgers University since 1970. His work focuses on three issues in epistemology: the defeasibility theory of knowledge, skepticism, and, more recently, infinitism. Recent publications in each of the areas include, "What Makes Knowledge the Most Highly Prized Type of Belief?" in *Sensitivity Principle in Epistemology*, ed. T. Black and K. Becker (Cambridge University Press, 2012), "Epistemic Justification and the Limits of Pyrrhonism," in *Pyrrhonism in*

*Ancient, Modern, and Contemporary Philosophy*, ed. Diego Machuca (Springer, 2011), and “Infinetism,” in *The Continuum Companion to Epistemology*, ed. A. Cullison (Continuum, 2012).

**Jonathan L. Kvanvig** is Distinguished Professor of Philosophy at Baylor University. His primary work is in metaphysics and epistemology, philosophy of religion, and philosophy of logic, and his publications include *The Intellectual Virtues and the Life of the Mind* (1992), *The Value of Knowledge and the Pursuit of Understanding* (2003), *The Knowability Paradox* (2007), and *Destiny and Deliberation* (2011).

**Matthew McGrath** is Professor of Philosophy at the University of Missouri. He has published articles in epistemology (on pragmatic encroachment, memory, and perception) and in metaphysics (on material composition, identity over time, and truth). He and Jeremy Fantl are the authors of *Knowledge in an Uncertain World* (Oxford University Press, 2009).

**Duncan Pritchard** is Professor of Philosophy at the University of Edinburgh. His books include *Epistemic Luck* (Oxford University Press, 2005), *The Nature and Value of Knowledge* (with A. Millar and A. Haddock, Oxford University Press, 2010), and *Epistemological Disjunctivism* (Oxford University Press, 2012).

**James Pryor** is Associate Professor of Philosophy at New York University. He works in epistemology and philosophy of language and neighboring parts of the field.

**Baron Reed** is Associate Professor of Philosophy at Northwestern University. He has published articles on skepticism, fallibilism, the nature of knowledge, and various other problems in epistemology.

**Patrick Rysiew** is Associate Professor of Philosophy at the University of Victoria, Canada. His primary research interest is in epistemology, including its points of intersection with certain issues in philosophy of language and psychology.

**James Van Cleve** formerly taught at Brown and is now Professor of Philosophy at the University of Southern California. He works in epistemology, metaphysics, and the history of early modern philosophy. He is currently working on a book entitled *Problems from Reid* (Oxford University Press, forthcoming), a sequel to his *Problems from Kant* (Oxford University Press, 1999).

**Jonathan Vogel** is Professor of Philosophy at Amherst College. His research is primarily in epistemology, but he also works in the history of philosophy and the philosophy of mind. He is the author of *Skepticism and Knowledge of the External World* (Oxford University Press, forthcoming).

**Roger White** is Associate Professor of Philosophy at Massachusetts Institute of Technology. His research interests span traditional and formal epistemology and philosophy of science. He has published articles on perceptual justification, skepticism, induction, Bayesianism, disagreement, and debunking arguments.

Timothy Williamson is the Wykeham Professor of Logic at the University of Oxford. He is the author of *Identity and Discrimination* (Blackwell, 1990), *Vagueness* (Routledge, 1994), *Knowledge and Its Limits* (Oxford University Press, 2000), *The Philosophy of Philosophy* (Blackwell, 2007), *Modal Logic as Metaphysics* (Oxford University Press, 2013), and numerous articles. *Williamson on Knowledge* (Oxford University Press, 2009), edited by Patrick Greenough and Duncan Pritchard, contains 15 essays on his epistemology with his replies.

Linda Zagzebski is George Lynn Cross Research Professor and Kingfisher College Chair of the Philosophy of Religion and Ethics at the University of Oklahoma. She has published extensively in epistemology, philosophy of religion, and a style of virtue ethics she calls “exemplarist virtue theory.” Her most recent book is *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief* (Oxford University Press, 2012). This book was written with the support of a Guggenheim Fellowship (2010–2011). A book in progress on exemplarist virtue theory is supported by a grant from the Templeton Foundation.

# Preface to the Second Edition

The second edition of *Contemporary Debates in Epistemology* contains opposing essays on five new topics: the analyzability of knowledge and how to do epistemology, pragmatic encroachment, the relation between knowledge and intellectually virtuous motives, the relation between knowledge and luck, and evidential slack. These essays can be found in Chapters 1, 4, 6, 7, and 12. The inclusion of these new chapters made it necessary to drop two debates from the first edition: one on conceptual content in perceptual experience, the other on epistemic responsibility. We have also added three new essays to the debate on immediate justification (Chapter 9) and two additional essays to the debate on justification and coherence (Chapter 10). Nine of the first edition's topics have been retained, so the second edition contains debates on altogether 14 chapters.

Significantly updated and enlarged, we believe that the second edition will, even more so than the first, be essential and fascinating reading to fellow epistemologists, graduate students, and advanced undergraduates. We wish to thank the contributors for debating each other vigorously and with sophistication, Travis Gilmore for proof-reading the manuscript, and Jeff Dean, Lindsay Bourgeois, and Jennifer Bray at Wiley-Blackwell for their invaluable assistance in putting this volume together.

Matthias Steup  
John Turri<sup>1</sup>  
Ernest Sosa

---

<sup>1</sup>JT's work on this volume was supported by the Social Sciences and Humanities Research Council of Canada, the British Academy, the Association of Commonwealth Universities, the National Endowment for the Humanities, The Character Project at Wake Forest University and the John Templeton Foundation (neither of which necessarily endorses any opinion expressed here), and an Ontario Early Researcher Award.

# Preface to the First Edition

This volume, part of Blackwell's Contemporary Debates series, is a collection of 22 essays on 11 central questions in epistemology. Each question is addressed by a pair of authors, acknowledged experts in their respective fields, who argue for clearly opposite points of view.

When reading the epistemological literature of the previous two or three decades, noting the bewildering plethora of views on "S knows that p" and seemingly endless supply of counterexamples, modifications, and refinements, it is easy to lose sight of the underlying, fundamental issues. The debate format, upon which this volume is based, is ideally suited to bring these issues into clear focus, and to cast the spotlight on what the arguments are that motivate opposing points of view. Thus, we believe that this collection will enliven epistemology and make exciting reading for scholars working in the field, graduate students, and advanced undergraduates alike. If it achieves this aim, it will be due to the hard work of the contributors. We wish to thank them for their efforts, and Jeff Dean and Nirit Simon at Blackwell Publishing for guiding this project towards completion.

Matthias Steup  
Ernest Sosa



## Chapter One

# Should Knowledge Come First?

What is the place of knowledge within epistemology? This is a methodological question of first importance. Timothy Williamson argues that knowledge should come first. Methodologically, this means that we shouldn't expect an informative analysis of knowledge in terms of belief, truth, plus some further set of necessary, non-circular conditions. Nor should we accept an impoverished conception of evidence or epistemic normativity that would be acceptable to a skeptical interlocutor. Rather, we should begin doing epistemology by focusing on epistemic access itself, and knowledge itself is the most natural candidate for this access. We can then understand other notions such as evidence and justification in terms of knowledge, which is a factive, world-involving state. In opposition to Williamson's view, Trent Dougherty and Patrick Rysiew argue for an *experience-first* approach to epistemology. Experience is where it all begins – indeed, where it *must* begin. Understanding experience to include perceptual awareness, introspective awareness, and rational insights, Dougherty and Rysiew argue that it is our basic evidence – that in virtue of which all else is made evident – that must come first in any intellectual project. On this approach, it is experience that ultimately justifies belief, guides rational thinkers, signifies truth, and settles disputes.

## Knowledge First

*Timothy Williamson*

Epistemology matters.<sup>1</sup> It is not just fascinating in itself; its concerns arise in every serious form of human inquiry. How much does this evidence support that claim? Is this alleged evidence just another claim, itself in need of evidence? Is the category of evidence even relevant here? Such disputes are often resolved by the internal standards of the inquiry. But sometimes they go deeper, raising issues adequately addressed only at a level of abstract reflection characteristic of philosophy. It is all the more striking

that in the second half of the twentieth century epistemology acquired a reputation (of which many practitioners still seem unaware) as sterile and inward-looking. Some of that was just the usual complaint about analytic philosophy, that it is boring and inaccessible to the untrained. But much of it came from other analytic philosophers, well acquainted with analytic epistemology, who still found it small-minded and old-fashioned in relation to its proper task.

One aspect was the post-Gettier industry of trying to analyze knowledge in terms of belief, truth, and non-circular further conditions. The increasingly gerrymandered definitions were obvious signs of a degenerating research program. Most of them, if correct, seemed to make knowledge too *grue*-like to be worth analyzing. But in any case they succumbed one after another to counterexamples. Moreover, no prior reason to expect knowledge to have such an analysis withstood scrutiny. Evidence accumulated that few if any words of natural language are understood by means of complex definitions. Nor does the nature of knowledge provide any clear evidence that it has such an analysis. That knowing entails believing truly does not show that for some non-circular condition C, knowing is equivalent to believing truly and meeting C. In consequence, the project of analyzing knowledge has lost its importance to analytic epistemology.

Another common charge is that epistemology is obsessed with the problem of skepticism, wasting its time on an imaginary opponent. The charge is partly unfair. Not only is analytic epistemology not mainly concerned to answer skepticism, the issue usually arises when apparently legitimate ways of criticizing ill-founded beliefs, applied more systematically, lead to skeptical conclusions. Epistemologists did not willfully introduce the skeptical tendency from outside; it was already in us. That does not mean that the skeptic is right, just that we should consider the issue.

However, skepticism plays a further role in defining the framework of much analytic epistemology. Many practitioners take the key epistemological notion to be not knowledge but *justification*, in a specifically epistemic sense (a qualification henceforth understood). They typically explain the difference between knowledge and justification by contrasting an everyday situation with a skeptical scenario in which everything appears the same to one but one is a brain in a vat. In the good case, one knows that one has hands. In the bad case, since one lacks hands, one does not know that one has them; one merely appears to oneself to know that one has hands. By contrast, such epistemologists claim, in the two cases one is justified to exactly the same degree in believing that one has hands. They intend this notion of justification for general epistemological purposes, not only for handling skepticism. In particular, they treat it as the appropriate normative standard for criticizing beliefs.

That view casts appearances in a leading role: justification supervenes on them. But what is epistemologically so special about appearances? One answer is that if cases appear the same to one, then one cannot discriminate between them, and cannot fairly be criticized in one case with respect to a feature it does not share with the other, since one cannot discern it. The underlying principle is that justification is exactly the same in cases indiscriminable to the subject. But that principle falls to an objection from the non-transitivity of indiscriminability. Consider a long sorites series of cases  $\alpha_0, \dots, \alpha_n$ , where the subject  $\alpha_i$  and  $\alpha_{i+1}$  are so similar in appearance that they are indiscriminable ( $i=0, \dots, n-1$ ), but  $\alpha_0$  and  $\alpha_n$  are so different in appearance that they are easily discriminable. By the principle, justification is exactly the same in  $\alpha_i$  and  $\alpha_{i+1}$  ( $i=0, \dots, n-1$ ). Therefore, by the transitivity of exact sameness in a given respect,

justification is exactly the same in  $\alpha_0$  and  $\alpha_n$ . But that is absurd, for justification differs between easily discriminable cases for some proposition. To avoid such objections, one must stick to the weaker principle that when appearances are exactly the same, so is justification. But why predict so tight a link between justification and appearances, if it is not mediated by indiscriminability?

Another idea is that in some sense one is always fully *acquainted* with present appearances to one, even if one cannot discriminate slight differences between them. This may help to explain the epistemological privileging of appearances (if one is never fully acquainted with anything else). The upshot is some sort of phenomenal conception of evidence. But by now, alarm bells should be ringing. The idea of full acquaintance with appearances has no basis in contemporary psychology or cognitive science. Nor has it much phenomenological plausibility. When I ask myself what I am acquainted with, the physical objects in front of me are far more natural candidates than their appearances. If I try to introspect or otherwise catch how things appear to me, I experience confusion, characteristic of embarking on an ill-defined task. Rather, full acquaintance with appearances is a wild postulate of a specific type of epistemological theory, one that requires something to be fully and unproblematically given to the subject to serve as the basis for justification. This is a barely updated Cartesianism: however vulnerable I am to doubt, ignorance, and error, something in me is clear and distinct.

To a depressing extent, epistemology has served as the refuge of an otherwise discredited philosophy of mind, supporting and supported by the definition of epistemic normativity in terms of a skeptical challenge. For justification was explained by the contrast between the ordinary case and the skeptical scenario. There may even be a distorting selection effect, by which those inclined to think along such lines are disproportionately drawn to, and rewarded in, epistemology rather than other branches of philosophy. It is Cartesianism that makes epistemology the starting point.

Suspicious of full acquaintance with sensory appearances, we might strip them out of the picture. What that leaves of the inner is a formal structure of beliefs, for which the norm of justification above reduces to mere internal coherence. Subjective Bayesianism is the best developed such view. Despite its mathematical virtues, it fails to make most distinctions of epistemological significance. It treats alike you and someone with the same credences but a radically different perceptual experience of the world.

The starting point of Cartesian epistemology is the comparison between the good case and the corresponding bad case. From a contemporary perspective, what they most obviously share is an internal microphysical state *S*. Consequently, they also share any mental state that supervenes on *S*. But *S* has no privileged *epistemological* status. Our internal microphysical states are typically unknown to us, and can only become known through arduous scientific investigation. Similarly, any mental state that supervenes on *S* has as such no epistemological privilege. It may be a state of depression that we cannot introspect ourselves to be in. What matters is whether we are aware of being in the given state, whether we *know* that we are in it.

Epistemologies that explicitly make knowledge a secondary phenomenon may nevertheless implicitly put it first, because they select the mental states to which they officially assign a privileged epistemological status for their supposed amenability to being known. They typically take that amenability for granted, rather than subjecting their crucial choice of starting point to open reflection. The less amenable the selected mental states turn out to being known, the more arbitrary becomes their promotion to

a privileged epistemological status. Once we give up Cartesian fantasies about the mind, we can recognize that no special sort of fact is automatically amenable to being known, although many sorts of fact often are known. Rather than seeking a domain to which we have privileged epistemic access, we should concentrate on epistemic access itself. By far the clearest explication of “epistemic access” is simply knowledge. Thus attempts to start epistemology with something much more internal than knowledge nicely illustrate the naturalness of starting with knowledge itself.

On one knowledge-first view, our total evidence consists of facts we know, irrespective of whether they are facts about our mental states. We are in no position to use facts we don't know as evidence. When we acquire new evidence in perception, we do not first acquire unknown evidence and then somehow base knowledge on it later. Rather, acquiring new evidence is acquiring new knowledge. That knowledge need not itself be based on further evidence, nor is it evidence for itself in some non-trivial way. But it is evidence for or against potential answers to questions to which we do not yet know the answer.

Equating evidence with knowledge helps reconnect epistemology with other fields. For one of the ways in which it marginalized itself was by depicting evidence as utterly unsuited to its role in science. The evidence for a well-confirmed scientific theory is typically a matter of public record. At least to a first approximation, it consists of facts intelligibly related to the theory and available to be known by anyone of suitable intelligence and training who takes the trouble to find out. It does not consist of facts about the present mental states of scientists or anyone else, facts that are no matter of public record and whose evidential relation to the theory itself has never been properly explained. Although the fact that a physicist believes a physical theory may raise its probability, the link requires an auxiliary sociological hypothesis and is hardly an evidential relation of the sort with which physics typically deals.

If our evidence is what we know, the evidence differs between the good and bad cases, contrary to what skeptics and many other epistemologists assume. For in the good case but not the bad, the subject's evidence includes the fact that she has hands. Although one can stipulate an alternative sense for the word “evidence” in which the evidence is the same, the challenge is to give epistemological significance to the new sense. The preceding reflections suggest that any idea that the two cases are evidentially equal is no basic insight but a product of misconceived epistemological theorizing. In particular, the fact that for all one knows in the bad case one is in the good case does not entail that one has the same evidence; it just means that for all one knows in the bad case one has the same evidence as in the good case. One is not always in a position to know whether one's evidence includes a given proposition. Although we might prefer a notion of evidence that does not work like that, we have no right to expect one.

A knowledge-first approach discourages trying to explain knowledge in terms of belief. We may even try the reverse, explaining belief in terms of knowledge. Here is a simple picture. Beliefs are the products of cognitive faculties whose function is to produce knowledge. When and only when all goes well, beliefs constitute knowledge. Even if something goes wrong, the belief may still be true, just as someone's scheme for getting rich may fail while they become rich by an unintended chain of events. Thus knowing is the successful state, believing the more general state neutral between success and failure. Knowing corresponds to doing, believing to trying. Just as trying is naturally understood in relation to doing, so believing is naturally understood in relation to knowing.

If justification is the fundamental epistemic norm of belief, and a belief ought to constitute knowledge, then justification should be understood in terms of knowledge too. Indeed, a belief is fully justified if and only if it constitutes knowledge. Although your belief that you have hands is fully justified, the corresponding brain in a vat's belief is not. But the brain in a vat has a good excuse for believing that it has hands, because, for all it knows, its belief that it has hands is justified, since, for all it knows, it knows that it has hands. Confusion between justifications and excuses undermines much talk of epistemic justification.

Some beliefs fall shorter of justification than others. In that respect we can grade beliefs by their probability on the subject's evidence, that is, on the subject's knowledge. A theory of evidential probability can be developed along such lines. It fills a gap between purely subjective probabilities, Bayesian credences ("degrees of belief"), and purely objective chances in indeterministic physics. When we ask how probable a theory is on our evidence, we want something less dependent on our doxastic state than a credence but more dependent on our epistemic state than a chance.

The most salient feature of knowing as the focus of epistemology is that it is a world-involving state. For it is factive: knowing that *P*, unlike believing that *P*, entails that *P*. Thus the state of the external environment constitutively constrains one's epistemic state. More specific factive epistemic states include perceiving that *P* and remembering that *P*. But even believing involves the world in another way. For the external environment constitutively constrains the contents of most intentional states; belief and knowledge are no exceptions. Croesus knew and believed that he was rich in gold. Despite being in the same internal microphysical states, Twin-Croesus on Twin-Earth neither knew nor believed that he [Croesus] was rich in gold. He had no knowledge or beliefs about Croesus, since he never had any suitable contact with him, however indirect. Nor did he know or believe that he [Twin-Croesus] was rich in gold. He wasn't. He was rich in another material, superficially like gold. Although many attempts have been made to define some sort of "narrow content" for Croesus and Twin-Croesus to share, all rely on deeply problematic assumptions. In any case, the intentional states that normally matter to us are broad states like believing that Croesus was rich in gold and wanting to be rich in gold oneself. Factive states involve the world twice over, in both their contents and their attitudes to those contents.

The ways in which an intentional state involves the world are not impurities. They are its point. The function of intentional states is to enable us to engage intelligently with the world. Take another example. Only those suitably related to Heloise can be in the state of loving her. The idea that the real core of loving Heloise is a mental state one can be in even if there is no Heloise looks more like a symptom of pathological self-absorption than a serious philosophy of mind. Without something loved or hated, there is at most an illusion of love or hate. Similarly, the connection to water is not accidental to desiring water or believing that there is water over there. The same goes for factiveness. A connection to the external environment is not accidental to the mental nature of seeing that it is raining, nor is a connection to past events accidental to the mental nature of remembering that it was raining. Misperceiving must be understood as a deviation from perceiving, and misremembering as a deviation from remembering. A neutral state that covers both perceiving and misperceiving is not somehow more basic than perceiving, nor is a neutral state that covers both remembering and misremembering more basic than remembering, for what unifies the various cases

of each neutral state is their relation to the successful state. Likewise for the generic factive state of knowing: a neutral state that covers both knowing and seeming to oneself to know is not somehow more basic than knowing, for what unifies the various cases of seeming to oneself to know is their relation to knowing. That whenever one is in such a broad state one is in some specific internal microphysical state too is no reason to postulate a corresponding narrow mental state.

An internal starting point for epistemology is false to the nature of mental life. In bracketing the differences between the good case and the skeptical scenario, the internalist approach to justification does not isolate a purely mental dimension; it merely ignores those aspects of the subject's awareness present in the good case but absent in the bad one. Much epistemology has been in denial about the depth of externalist developments in the philosophy of mind over recent decades, as though broad mental states could be analyzed into narrow mental states and their causal relations to the external world. But the postulated underlying layer of narrow mental states is a myth, whose plausibility derives from a comfortingly familiar but obsolescent philosophy of mind. Knowledge-first epistemology is a further step in the development of externalism.

A closely related contrast occurs in the philosophy of language. Truth-conditional, referential semantics is an externalist program. On such a theory, in a context of utterance the atomic expressions of a language refer to worldly items, from which the truth conditions of sentences are compositionally determined. Just like the contents of someone's intentional states, the truth-conditional semantics even of their idiolect is far from supervening on their internal microphysical states; it also depends on what external objects they are in causal contact with. Reference and truth, like knowledge, are matters of success, not of something neutral between success and failure. Insofar as there is a competing internalist program, it is conceptual role or inferential semantics, on which the meaning of an expression is something like its place in a web of inferential relations. Although inferential semantics can be given an externalist twist, in practice many of its proponents are motivated by internalist sympathies. On internalist inferential semantics, the inferential relations of an expression do not depend on what, if anything, it refers to, although there may be dependence in the opposite direction if reference is determined by a combination of inferential and external causal relations. Inferentialism faces grave problems of principle, for instance in separating patterns of inference that are to count as essential to the meaning of an expression from those that will count as accidental (a form of the analytic/synthetic distinction). Moreover, the internalist version has particular difficulty in establishing an adequate relation between meaning and reference. Even more striking is the disparity in practice between the success of referentialist truth-conditional semantics as a flourishing research program, pursued by both philosophers of language and linguists, which has provided invaluable insight into semantic phenomena in natural languages, concerning both overall structure and the behavior of specific expressions, and the lack of progress of inferentialism, which remains in a largely programmatic state. By that pragmatic criterion, referentialism beats inferentialism hands down. Internalism has proved to be an obstacle to new insights in the philosophy of language.

The fruitfulness of referential semantics is an encouraging precedent for knowledge-first epistemology, since both take as basic the central forms of success distinctive of their field: truth and reference in semantics, knowledge in epistemology. In the long run, knowledge-first epistemology too should be judged by its fruitfulness

as a research program, compared to its competitors. I have shown elsewhere how a knowledge-first methodology casts light on such matters as the nature of indiscriminability and the norm of assertion. Here, for reasons of space, I will focus on another issue: applications of epistemic logic in epistemology.

In recent years it has become clear that formal models of epistemic logic enable us to analyze some epistemic phenomena in a more disciplined, systematic, and rigorous way than we can achieve through qualitative description in ordinary prose, even eked out with epistemological jargon. We can sometimes learn far more about the target phenomena by mathematically exploring the consequences of a model than by the shaky reasoning and appeals to the “obvious” characteristic of so much traditional epistemologizing. Of course, the models usually involve simplifications and idealizations, just like mathematical models in natural science. Humans are no more logically omniscient than planets are point masses. As always, formal methods will not give good results when applied with bad judgment. It takes experience and skill to know which simplifications and idealizations are appropriate for a given problem. But epistemologists are gradually acquiring the relevant experience and skills.

What is the connection between epistemic logic and knowledge-first epistemology? In standard epistemic logic, the basic epistemic operator is  $K$ , read “The agent knows that ...,” and interpreted in each model by means of an accessibility relation of epistemic possibility between worlds. A world  $x$  is epistemically accessible from a world  $w$  just if for all one knows in  $w$  one is in  $x$ , that is, everything one knows in  $w$  is true in  $x$ . A formula  $KP$  is true in a world  $w$  just if the formula  $P$  is true in every world epistemically accessible from  $w$ . Since epistemic accessibility is required to be reflexive, every instance of the factiveness schema  $KP \rightarrow P$  is true in every world in every model. Often the framework is multi-agent:  $K$  has different subscripts for different agents, and each agent has their own accessibility relation. The truth condition for  $KP$  is not intended as an analysis of knowledge in independent terms, for epistemic accessibility is itself explained in terms of knowledge. Rather, the semantics simply decodes the information about knowledge conveniently encoded in the accessibility relation. This framework accords with a knowledge-first methodology.

However, the accord so far is rather superficial. We can equally well introduce a basic doxastic operator  $B$ , read “The agent believes that ...,” and interpreted in each model by means of an accessibility relation of doxastic possibility between worlds. A world  $x$  is doxastically accessible from a world  $w$  just if everything one believes in  $w$  is true in  $x$ . A formula  $BP$  is true in a world  $w$  just if  $P$  is true in every world doxastically accessible from  $w$ . Since doxastic accessibility is not required to be reflexive, not every instance of the schema  $BP \rightarrow P$  is true in every world in every model. This framework accords with a belief-first methodology.

A clue that the accord with a knowledge-first methodology goes deeper is that in by far the most widely applied class of models of epistemic logic accessibility is an equivalence relation (reflexive, symmetric, and transitive), which automatically validates factiveness, and thereby favors interpreting the operator in terms of knowledge rather than belief. Such models are routinely used in most applications of epistemic logic in computer science and economics, for example to model common knowledge. The users are not motivated by any philosophical prejudice in favor of knowledge-first epistemology. They prefer models in which accessibility is an equivalence relation for their simplicity, tractability, and naturalness.

Once we go into finer-grained epistemology, we can no longer require epistemic accessibility to be an equivalence relation, for doing so validates the positive and negative “introspection” schemas  $KP \rightarrow KKP$  (if one knows that  $P$ , one knows that one knows that  $P$ ) and  $\neg KP \rightarrow K\neg KP$  (if one does not know that  $P$ , one knows that one does not know that  $P$ ), which fail for important and interesting reasons. Negative introspection fails because in the skeptical scenario one does not know that one has hands, but one is in no position to know that one does not know that one has hands. Positive introspection fails for reasons closely connected with the non-transitivity of indiscriminability discussed above, since it is equivalent to the transitivity of epistemic accessibility. They are comparatively harmless in most applications in computer science and economics because there the emphasis is on understanding essentially multi-agent epistemic phenomena, such as problems about the transmission of information. For those purposes, failures of introspection within a single agent constitute noise, best eliminated from the model to enable us to discern more clearly the phenomena distinctive of the multi-agent case. In epistemology, by contrast, our purpose is usually to understand phenomena that already occur in the single-agent case, so idealizing away failures of introspection is not harmless. This exemplifies the sort of consideration relevant to selecting an appropriate model. In any case, the fruitfulness of what is in effect a knowledge-first methodology when epistemic logic is applied to other disciplines is at least some indication that knowledge-first epistemology is on the right track.

A more direct point is that the knowledge-first equation of one’s total evidence with one’s total knowledge facilitates a natural way of introducing evidential probabilities into models of epistemic logic. At a coarse-grained level of individuation adequate for present purposes, a proposition in a model is simply a set of worlds, those in which it is true. One proposition entails another just if the former is a subset of the latter. For each world  $w$ , the set  $R(w)$  of worlds accessible from  $w$  is the strongest thing known in  $w$ , in the sense that the propositions known in  $w$  are just those  $R(w)$  entails. Since  $R(w)$  is what one knows in  $w$ , on the equation of evidence with knowledge the worlds consistent with one’s evidence in  $w$  are exactly those in  $R(w)$ . Now suppose that we have a prior probability distribution on the space of worlds. In the simplest case, there is a finite number  $n$  of worlds, each of prior probability  $1/n$ . Then we define the probability of a hypothesis  $h$  on the evidence in a world  $w$  as the conditional probability of  $h$  on  $R(w)$ . That is, we eliminate the worlds inconsistent with the evidence and scale up the probabilities of those that remain to total 1. Since accessibility is reflexive,  $R(w)$  always contains  $w$ , so we never conditionalize on the empty set. Such models can be used to explore the evidential probabilities of states of knowledge, ignorance, and evidential probability in far more systematic and controlled ways than were previously available. For example, it can be argued that sometimes our evidence eliminates a hypothesis  $h$ , even though it is almost certain on our evidence that our evidence does not eliminate  $h$ .

Such models with belief or excusable belief in place of knowledge are implausible. They make evidence non-factive, so a true proposition is sometimes inconsistent with one’s evidence. But that surely misdescribes a case of a true proposition inconsistent with what one mistakenly takes to be one’s evidence.

The obvious place to look for an alternative formal account of evidential probability is the more familiar Bayesian tradition. But it fails to integrate probabilities with an epistemology of evidence. The standard way of updating probabilities is by conditionalizing them on what is called “new evidence,” but rarely with any account



of what constitutes evidence. Although ultra-subjective Bayesians may count whatever the agent updates on as evidence, the claim that neo-Nazis have evidence inconsistent with the occurrence of the Holocaust is disreputable. A more flexible form of updating is Jeffrey conditionalization, on which there need be no proposition that is conditionalized on, but for the same reason it is even harder to integrate with anything that deserves to be called epistemology. Whether a particular change of probabilities is justified is not a purely formal matter; the usual Bayesian tradition provides only formal criteria.

Some rivals to knowledge-first epistemology seem to make evidence non-propositional (for example, if it comprises perceptual appearances). However, it is hard to explain how non-propositional evidence contributes to updating probabilities. We might therefore try to weld such an account of evidence onto the Bayesian framework by equating the new evidence (whatever it is) with the set of worlds in which the agent acquires just that new evidence, and requiring updating by conditionalization on that new evidence proposition. That is in effect to treat exact sameness of new evidence as the relevant accessibility relation. But, as already noted, exact sameness in a given respect is an equivalence relation. Thus such models are isomorphic to the simplest models of epistemic logic above, which accord with a knowledge-first methodology. Even though the accessibility relation is no longer characterized in terms of knowledge, we are back with something suspiciously like a knowledge-first framework. However, for the same reasons as before, we still have the positive and negative introspection principles that if one's evidence entails that P, then one's evidence entails that one's evidence entails that P, and that if one's evidence does not entail that P, then one's evidence entails that one's evidence does not entail that P. Those principles exclude a sort of higher-order uncertainty about the boundaries of one's evidence. But if evidence provides an epistemologically serious way of discriminating propositions, that positive introspection principle will be vulnerable to a version of the argument from the non-transitivity of indiscriminability. Thus the envisaged reduction of apparently non-propositional evidence to propositions yields the analogue of the least sophisticated, most elementary and objection-prone version of the knowledge-first account. A proper knowledge-first epistemology integrates epistemic logic and evidential probability in a way that makes more epistemological sense than any available alternative.

The rigorous development of knowledge-first epistemology began quite recently. We might therefore expect its longer-established rivals to be far ahead, judged by their capacity to generate epistemologically illuminating formal models. Instead, they are already in danger of being left behind. That is one of many signs that they belong to a stagnating research program. It is time to give knowledge-first epistemology its chance to do better.

## Note

- 1 The ideas sketched here are explained and argued for in more detail in my *Identity and Discrimination* (Oxford: Wiley-Blackwell, 1990; revised and updated, 2013), pp. 4–47; *Knowledge and its Limits* (Oxford: Oxford University Press, 2000); “On Being Justified in One’s Head,” in M. Timmons, J. Greco, and A. Mele, eds., *Rationality and the Good* (Oxford: Oxford University Press, 2007); *The Philosophy of Philosophy* (Oxford: Blackwell, 2007), pp. 208–277, and “Very Improbable Knowing,” *Erkenntnis* (forthcoming). Those works cite earlier contributors to the development of knowledge-first epistemology.

# What Is Knowledge-first Epistemology?

*Trent Dougherty and Patrick Rysiew*

What is knowledge-first epistemology? It is not yet clear to us how to answer that question. One key assertion seems to be that knowledge is “unanalyzable” – that is, not neatly factorizable into component parts. As evidence for this claim, Williamson cites the facts that (i) the Gettier problem hasn’t been solved in four decades, and (ii) attempts to solve it have led to clumsy analyses of knowledge. He also mentions there is no reason to think it is factorizable in the first place. He notes that its entailing belief and entailing justification does not entail that those things are constituents of knowledge. We agree, but note that there being such provides a pretty natural explanation for the relevant entailments. It is unclear what Williamson’s alternative explanation of the latter are, but, before we comment briefly on (i) and (ii), his explanation seems to go like this. Knowledge is a kind of success, and mere true beliefs are failures to achieve this success, even when justified. So we can think of (mere) belief, (mere) true belief, and (mere) justified belief as “botched knowledge” (Williamson, 2000, p. 47). But of course Williamson doesn’t think they are really a kind of knowledge. “Former president” doesn’t pick out a special kind of president. It is not as though there is the genus *Presidents*, one species of which is *Former*. The same goes for knowledge and “botched knowledge.” It is not as though there is the genus *knowledge*, one species of which is *botched*. Clearly, this is not at all what Williamson has in mind. But then it is hard to see how thinking of the relation between knowing and certain other states and goods on the model of the relation between doing and trying explains the data in question – namely, that knowledge entails belief and justification. Traditional epistemologists (and they might be knowledge-first in some sense; see Conee, “Truth Connection,” chapter 10 of Conee and Feldman 2004) have a simple explanation for this. For entailment is modeled in formal semantics as set inclusion. *Cat* entails *mammal* because the cats are a subset of the mammals. On the traditional view, *knowledge* entails *belief* because knowledge is a kind of belief, the kind that meets the conditions required to be knowledge. *Knowledge* entails *justified belief* for the same reason: it is a kind of justified belief, the kind that meets further conditions for being knowledge. Truth is one such further condition. Being based on one’s evidence in such a way to avoid a deviant causal chain is another. The latter kind of condition is hard to spell out, since there are so many ways a causal chain can go wrong between takeoff and landing, even if one gets to the right destination in the end.

There is another analogy available, akin to the one Williamson himself favors, which also incorporates teleological thinking into our conception of knowledge. Perhaps belief (for example) is akin to *intention* (rather than to Williamson’s *trying*) and knowing to action. Just as intentions are intentions to perform some action, believings “aim at” knowledge; and just as there are failed intentions, there are failed attempts at knowing (merely true beliefs, say). *This* allows that there is some good sense in which it might be proper to think of belief in terms of its relation to knowledge; so too, it might explain why knowing entails believing. But it also undercuts the motivation for an approach that puts “knowledge first” in some more interesting sense.

For it is very natural to think of actions as individuated in part by the intentions involved – to think, that is, that actions (vs. mere movements, etc.) “include” intentions. Likewise, it is natural to think of knowings as including believings – indeed, to think that what’s known (when it is) is identified in part by the belief(s) involved. But all this merely rehearses the traditional idea that belief enters into our understanding of knowledge, and that knowing is a species of belief.

## **(Un)analyzability**

Williamson thinks considerations pertaining to the Gettier literature show that the justified true belief (JTB) approach to knowledge is moribund at best. And he seems to think that this spells trouble for the value of justification relative to knowledge. But this is far from clear. On the contrary, in the olden JTB days, one might have thought that justification was important only because it was part of an analysis of knowledge. But the unanalyzability of knowledge – if unanalyzable it is – could in fact be seen as a liberation of justification to assume importance in its own right. Kvanvig (1992; 2003, p. 192) and Greco (2010, p. 9ff.), in defending an externalist notion of knowledge, theorize that the intuitions of epistemic justification internalists might be about not knowledge, but understanding, where understanding stands between knowledge and wisdom in value. Suppose knowledge is unanalyzable. Either understanding entails knowledge or not. If it doesn’t (Kvanvig, 2003), then knowledge doesn’t help us understand understanding, whereas justification might. If it does (Grimm 2006), then it appears to be a special subset of knowledge where (at a minimum) certain internalist goods are added (seeing connections, etc.), which might make it a particularly prized kind of knowledge. It then remains an open question whether these goods apart from knowledge are more valuable than knowledge apart from these goods.

Other examples in the same vein are available. There is, for instance, Sosa’s well-known distinction between “animal” and “reflective” knowledge (1991) – though it is perhaps not clear in what sense these are different *kinds* of knowledge, as opposed to a single kind with some “extra goods” sometimes being added, it being an open question what, if anything, accounts for the value of the latter (Kornblith, 2004). (The same goes for Lehrer’s distinction between “discursive” and “primitive” knowledge (2000).) And there is Foley’s argument (2004) that epistemologists have for too long thought that egocentrically rational and reliable belief – roughly, internalistically and externalistically justified belief, respectively – must converge, as opposed to each calling for its own theory and having its own distinctive value.

Much of Williamson’s discussion of “traditional epistemology,” it seems to us, equates it with a particular strain of internalistic theorizing; and much of Tim’s dissatisfaction, we think, is with the presumption that a certain form of internalistic justification *must* be a component of knowledge. This strikes us as a good worry to have. But the examples just mentioned serve to illustrate that abandoning that presumption does not commit one to any specifically “knowledge first” ideas.

So too, all the relevant parties can reject the idea that knowledge admits of any neat analysis. “Maybe,” as Plantinga says, “there isn’t any neat formula, any short and snappy list of conditions (at once informative and precise) that are severally necessary and jointly sufficient for warrant; if so, we won’t make much progress by grimly pursuing

them” (1993, p. 20). Perhaps “the program of analysis,” as Williamson calls it (Williamson, 2000, p. 31), is a hangover from the heyday of logical atomism (Williamson, 2000); perhaps it is rooted in a faulty theory of concepts<sup>1</sup>; perhaps the best we can reasonably hope for is “reflective understanding” (Williamson, 2000, p. 33). Still, it’s an open question whether, within such an understanding, knowledge will have to be taken as unanalyzable, and justification (for example) understood as an entirely derivative notion.

## **Justification and Excuses**

Suppose you unwittingly receive a perfect forgery of an authorization granting you permission to explore a protected piece of land. Note in hand, you proceed past the many “No Trespassing” signs. After about half an hour, you encounter a patrol officer who inspects your alleged permit, detects the forgery, and escorts you off the premises but does not prosecute you. Here are two competing descriptions of the event. First comes the one we take to be the natural interpretation: You were justified in crossing the “No Trespassing” signs *because* you had the misleading signification of the note. Not only are you not to blame, your behavior is not subject to any legitimate criticism. The bare fact of being at odds with a law is irrelevant from a normative point of view, and knowledge is a normative notion. Next comes Williamson’s interpretation. He seems to imply that your behavior was simply unjustified. Your ignorance excuses you from any punishment for this unjustified behavior, but the behavior was unjustified from start to finish. Saying that one was justified in believing they were justified does not change this result.

Suppose the latter interpretation is right. Suppose, that is, that your behavior was unjustified *full stop*, and that you are merely blameless. Taking the epistemic case: suppose, as Williamson holds, that in “the bad case” one is not justified but merely blameless. Doesn’t this go along with thinking that justification, understood apart from knowledge, just doesn’t have much real work to do? Not at all. Kent Bach (1985), for example, has argued that much theorizing about justification conflates issues of justified belief with issues of justified/blameless believers. The result, Bach thinks, isn’t that justification is of no great theoretical interest, but that theorists are freed up to pursue externalist theories thereof, including ones whereby those in the bad case don’t have any justified beliefs. Similarly, Rysiew (2011) considers as a live option the view that those in the bad case lack any real (as opposed to apparent) evidence – “contrary,” as Williamson says, “to what sceptics and many other epistemologists assume” (p. 4). He does so, however, without endorsing “E=K” (and, in fact, while allowing for the sort of view of evidence we outlined in our opening statement). Once again, then, there are various extant theories that preserve the result Williamson ultimately recommends, but without going the knowledge-first route.

## **The Good and Bad Cases**

Williamson insists that what is common to perceiving and misperceiving is not more basic than perceiving. Fortunately, we do not need any basicity claim. All we need is that there *is* something in common. What the veridical and illusory cases have in

common, clearly enough, is what it is like to be in those states. In addition to shared qualitative character, they have the same representational contents. Both experiences assert, as it were, the same thing. And, of course, this phenomenology is *in a way* more direct to us than the external world, since it is in virtue of our experience that we are aware of the world. (Note that this only makes experience the medium, and not the object, of awareness.) Williamson has strong words indeed for the view that narrow content is the “real core” of experience. We are quite glad, then, that “experience first” epistemology is committed to no such thesis. Again, all we need is that there is something *shared* between them. And this shared thing – what it’s like to have the world look that way – is where we need to begin in finding out how the world is. Whether or not the terminology or theoretical machinery of “narrow content” is the best way of getting at it, *what it’s like* to see a red mug is exactly *what it’s like* to see a perfect hologram of a red mug; no cognitive science can reveal *that* to be false. In the good case, *as a matter of fact* what it’s like is causally hooked up to the world in the right kind of way, such that the experiential signs are not misleading and so we have not only justified belief but knowledge. In the bad case, we lack knowledge, but there is some credit in heeding misleading evidence.

Of course, once again, maybe the latter such credit should not be identified with justification of the sort required for knowing. Even so, it should be clear that resisting putting knowledge first does not require – and in our case, does not involve – throwing ourselves behind “the veil of ideas,” seeing subjects as being acquainted only with “appearances,” and suchlike. We take our view to be perfectly compatible with direct realism about the external world and not to be a version of sense-datum theory (Chisholm adopts the adverbial view to defend direct realism, but is an “experience first” epistemologist; Chisholm, 1989, esp. pp. 66ff.). See also Huemer, 2001, chapter IV, esp. sect. 5). So too, it should be clear that ours is not a view that is born out of an obsession with the problem of skepticism. That experience comes first in the indicated sense is simply a fact of our everyday epistemic lives.

## Regarding Indiscriminability

Williamson offers a sorites argument against the principle that “justification is exactly the same in cases indiscriminable to the subject.” Indiscriminability, he hypothesizes, is the link between appearances and justification. (We are a bit worried about the shift in that discussion from talk of *beliefs* being justified to *persons* being justified, for it suggests a possible running together of the idea of epistemic justification and epistemic responsibility.) What he may have in mind is this: It seems that our unfortunate envatted counterparts are just as justified as we are in believing in an external world, for that is *all we have to go on*. We have already indicated how our own central claims do not require that result. But suppose it is right. What is the problem? Williamson claims that this picture falls prey to a sorites argument. One problem with his argument is that it assumes experiences are fully determinate rather than vague. Yet fairly early on in the development of contemporary empiricism – in response to the Problem of the Speckled Hen – both Ayer (1940) and Chisholm (1942)<sup>2</sup> noted that experiences or the characters of sensing are in fact not fully determinate (see Tye, 2009, for an excellent discussion of a representationalist perspective on this problem and Dougherty, 2011,

for more context). We do not have  $RED_{\lambda=.72744}$  experiences and  $RED_{\lambda=.72745}$  experiences. Rather, we have “reddish” or even “somewhat reddish” experiences. All basic evidence is necessarily vague. But even if higher-order anti-luminosity prevents this, nothing important follows, for, as Jeffrey showed, we can get along perfectly well with uncertain evidence.

## Regarding Formal Methods

We agree that formal models can sometimes be illuminating of epistemic phenomena. And it is hard to disagree that “shaky reasoning” isn’t the way to go. However, we suggest that traditional epistemologists sometimes exercise the same kind of careful, skilled inquiry as do formal epistemologists, even when they are not working with numbers or special symbols. Williamson mentions several of the many risks of using formal models, and it seems to us that the cost-to-benefit ratio is about the same in formal and non-formal modes of inquiry, when pursued conscientiously by capable parties.

It is unclear how any kind of Bayesianism is an *alternative* to (almost) any epistemic logic, for it is unclear whether they are designed to do the same thing. For example, if we let the weak epistemic modality be “it is permissible for the agent to believe that,” there are forms of Bayesianism perfectly compatible with the theorems such an epistemic logic would include. Likely this is so also if the strong operator reads “It is certain that.” Also, we don’t know in advance whether the simplifications and idealizations of the respective theories would “hook up” in a direct enough way to constitute a rivalry. And if there were rivalry, we would perceive no threat to experience-first epistemology, because it is not at all clear that the success of epistemic logic in being illuminating furnishes any reason to adopt the knowledge-first approach.

Williamson says Bayesianism fails to integrate probabilities with an epistemology of evidence. Like all philosophers, Bayesians start out with a set of problems to solve. Most Bayesians are concerned with ways of characterizing coherence properties, which are generally agreed to be good-making features of one’s noetic structure. We doubt Williamson disagrees. Other Bayesians are interested in formal learning theory and are not concerned with where the evidence comes from, but rather with what one ought to do with it when one gets it. So a theory of evidence is simply outside the scope of standard Bayesian pursuits. But of course many Bayesians have a favored view of evidence (Swinburne, 2001, for example). Many probabilists don’t commit to any formal learning theory and simply see probability logic as exactly parallel to first-order logic. Its job is to tell you what choices you face given your current commitments (Howson and Urbach, 1996). In Jeffrey’s case, the assigning of probabilities to basic propositions is a *techne* – it requires an art of judgment that is generally acquired by practicing in the appropriate community (Jeffrey, 1992). Williamson says, “When we ask how probable a theory is on our evidence, we want something less dependent on our doxastic state than a credence but more dependent on our epistemic state than a chance” (p. 5). This is just what we get on an experiential theory of evidence wedded to a Chisholm-like theory of evidence that states objective, material epistemic principles. We end up with a form of epistemic probability that strikes the appropriate balance between subjectivity and objectivity.

Williamson says that traditional epistemology, since it has been going on much longer than knowledge-first epistemology, should be ahead when “judged by their capacity to generate epistemologically illuminating formal models” (p. 9). It is not at all clear why it should have been expected to generate *any* such models. Nor is it clear that what he calls “traditional” epistemology *has* been around longer. Indeed, it is sometimes criticized as having started essentially with Descartes, whereas some kind of “knowledge first” view seems to go back to Aristotle, and to have been expounded by Scholastic philosophers.<sup>3</sup> Mathematical philosophy germinated in the nineteenth century, budded in the 1920s, and began to blossom in the mid to late 1970s. So we doubt we should expect any kind of theory to bear heavy formal fruit yet. (And it is worth noting that it has been only a little more than four decades since Gettier’s paper, and that some people are satisfied that they have solved the problem, and in a very non-grue-like fashion (Feldman, 2003, p. 125), and interesting work continues to be done on the topic (Bernecker, forthcoming). Furthermore, it has now been a fourth of that time since publication of *Knowledge and Its Limits*, and there remain many, many details to work out. It would be unreasonable to expect of any view, however, that a project worth doing could be done in the lifetime of one philosopher.) However, Bayesianisms and more general probabilisms are alive and well, indeed flourishing across the globe today, and most of them are compatible with most theories of evidence, including the commonsense empiricist theory that our fundamental evidence – at the base or frame of our noetic structure – consists in the experiences by which the world is revealed to us but which sometimes lead us astray.

Finally, Williamson notes that some rivals to knowledge-first epistemology take evidence to be non-propositional. For example, my warm feeling can be evidence that it is hot in here. But then he says that it is hard to explain how non-propositional evidence contributes to updating probabilities. This issue has been discussed quite a bit elsewhere,<sup>4</sup> including in our opening remarks, so there is no need to add to them here.

## Notes

- 1 Cf. Kornblith (2007), who notes trouble for the traditional philosophical project of conceptual analysis: “Knowledge, for example, may be analyzed, on certain views, as justified, true belief meeting some additional, and difficult to specify, condition. It is taken for granted that the form of a proper analysis is just some such set of individually necessary and jointly sufficient conditions. The idea that our concepts are mentally represented in this form is what psychologists refer to as the Classical View of concepts. Since the early to mid-1970s, it has become increasingly clear that the Classical View is not correct.” The moral Kornblith draws from this, however, isn’t Tim’s. Rather, it is that the standard philosophical method, which includes liberal appeal to intuitions, is not a reliable method of understanding our concepts – and, he thinks, our concepts are not plausibly viewed as the target of philosophical understanding anyway.
- 2 In fact, Chisholm explicitly considers the non-transitivity of indistinguishability (1942, p. 371). He notes that Russell was dealing with this problem as early as 1921. He claims that this allows basic evidence to be certain, though we are not committed to that.
- 3 “Whereas Descartes seeks to place philosophy and science on firm foundations by subjecting all knowledge claims to a searing methodological doubt, Aristotle begins with the conviction

that our perceptual and cognitive faculties are basically dependable, that they for the most part put us into direct contact with the features and divisions of our world, and that we need not dally with sceptical postures before engaging in substantive philosophy” (<http://plato.stanford.edu/entries/aristotle/>). Yet Aristotle held that knowledge of the external world was by means of “sensible species.” Reid has an interesting discussion of Aristotle’s epistemology (*IP* IV 2, *W* 372a–b).

- 4 See note 9 from our opening statement “Experience First.”

## References

- Ayer, A.J. (1940) *The Foundations of Empirical Knowledge*. London: Macmillan.
- Bach, K. (1985) A rationale for reliabilism. *The Monist* 68 (2): 246–263.
- Bernecker, S. (forthcoming) Keeping track of the Gettier problem. *Pacific Philosophical Quarterly*.
- Chisholm, R. (1942) The problem of the speckled hen. *Mind* 51 (204): 368–373.
- Chisholm, R. (1989) *Theory of Knowledge*, 3rd edn. Upper Saddle River, NJ: Prentice Hall.
- Conee, E. and Feldman, R. 2004. *Evidentialism: Essays in Epistemology*. Oxford: Oxford University Press.
- Dougherty, T. (2011) Evidence. In *Oxford Bibliographies Online: Philosophy*. <http://oxfordbibliographiesonline.com/view/document/obo-9780195396577/obo-9780195396577-0132.xml> (last accessed February 27, 2013).
- Feldman, R. (2003) *Epistemology*. Englewood Cliffs, NJ: Prentice Hall.
- Foley, R. (2004) A trial separation between the theory of knowledge and the theory of justified belief. In *Ernest Sosa and His Critics*, ed. J. Greco. Oxford: Blackwell.
- Greco, J. 2010. *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cambridge University Press.
- Grimm, S. (2006) Is understanding a species of knowledge? *British Journal for the Philosophy of Science* 57 (3).
- Howson, C. and Urbach, P. (1996) *Scientific Reasoning: The Bayesian Approach*. Open Court.
- Huemer, M. (2001) *Skepticism and the Veil of Perception*. Lanham, MD: Rowman & Littlefield.
- Jeffrey, R. (1992) *Probability and the Art of Judgment*. Cambridge: Cambridge University Press.
- Kornblith, H. (2004) Sosa on human and animal knowledge. In *Ernest Sosa and His Critics*, ed. J. Greco, 2004. Oxford: Blackwell.
- Kornblith, H. (2007) Naturalism and intuitions. *Grazer Philosophische Studien* 72: 27–49.
- Kvanvig, J. (1992) *The Intellectual Virtues and the Life of the Mind: On the Place of the Virtues in Contemporary Epistemology*. Lanham, MD: Rowman & Littlefield.
- Kvanvig, J. (2003) *The Value of Knowledge and the Pursuit of Understanding*. New York: Cambridge University Press.
- Lehrer, K. (2000) Discursive knowledge. *Philosophy and Phenomenological Research* 60 (3): 637–653.
- Plantinga, A. (1993) *Warrant and Proper Function*. Oxford: Oxford University Press.
- Reid, T. (1895/1983) *Thomas Reid: Philosophical Works*, 8th edn. Edited by William Hamilton. New York: Georg Olms Verlag.
- Rysiew, P. (2011) Making it evident: evidence and evidentness, justification and belief. In *Evidentialism and Its Discontents*, ed. T. Dougherty. Oxford: Oxford University Press.
- Sosa, E. (1991) *Knowledge in Perspective: Selected Essays in Epistemology*. New York: Cambridge University Press.
- Swinburne, R. (2001) *Epistemic Justification*. Oxford: Oxford University Press.
- Tye, M. (2009) A new look at the speckled hen. *Analysis* 69 (2): 258–263.
- Williamson, T. (2000) *Knowledge and its Limits*. Oxford: Oxford University Press.



# Experience First

*Trent Dougherty and Patrick Rysiew*

In what way is experience first, epistemically speaking? It is first, at least, in the order of immediacy: in short, experience is where we begin (which is not itself to say that it is what we know best). This has important implications that we will discuss shortly, but first we are happy to acknowledge that knowledge might indeed be first in some ways as well. For example, we appear to acquire the concept very early, before that of belief (much less justification) (Perner, 1993; Bartsch and Wellman, 1995; Williamson, 2000, p. 33, n.7). From this, though, nothing follows about the correct relative priority of the concepts in epistemological theory. (Still less does it tell us about the metaphysics of knowledge itself.) One might also think knowledge first in one or another teleological way. For instance, it may be that belief “aims at” knowledge, or that knowledge is “the norm of belief.” But, if true, this doesn’t entail any more substantive sense in which belief must be understood in terms of knowledge, not the other way around. Alternatively, one might think the main purpose of our cognitive system as a whole – or, at least, the portion(s) thereof in which epistemologists are traditionally interested – might be to acquire knowledge. This thesis faces opposition from either end. From a broadly naturalistic perspective, it might seem that mere true belief, or indeed just “getting by,” is the purpose of our cognitive architecture.<sup>1</sup> At the same time, there are familiar but understudied epistemic goods that, according to some (Kvanvig, 1992, 2003; Zagzebski, 1996; Greco, 2010), are clearly more valuable than knowledge: understanding and wisdom. Indeed, one of the big shifts in epistemology toward the end of the twentieth century was a return to interest in epistemic virtues. So knowledge may turn out to be just a middling epistemic desideratum. More radically, some Bayesians find no need for the notion at all.

The view just rejected as insufficient was a particular teleological interpretation of the “knowledge first” slogan, one which makes knowledge the first on the list of epistemological ends. Most of our lives, however, are occupied with securing means. An end having been set, the question is, What do I do now? This is where experience is first: in the quest for true belief, justification, knowledge, understanding, and wisdom we have no other starting point than experience. Our experiences (broadly construed to include what it’s like to have intuitions and rational insights, etc.) are our basic evidence, in the light of which all else that is evident is made evident. Experiences play well the roles that characterize evidence. We will consider four such roles (see Kelly, 2011, for more on the functional approach) and show how experiences are well suited to the task.

Let us be clear from the outset: it might be that no single thing can play all the roles typically ascribed to evidence (Kelly, 2011). Nevertheless, a good argument can be made that experiences can satisfy all adequately.<sup>2</sup> Furthermore, we do not here argue that experiences are the *only* thing that can play these evidential roles. The thesis, rather, is that experiences *can and do* play these roles, and that they do so, moreover, in a way that can reasonably be described as “basic” or “ultimate,” in the sense that they are evidential regress stoppers. Any chain of cited evidence must end with the

way the world appears to us to be. So on this view, experience is first in that it inhabits the ground floor of the intellectual edifice. The four roles we will consider as evidence are: (1) what justifies belief; (2) what rational thinkers judge by; (3) a guide to truth (evidence as sign, symptom, or mark); and (4) neutral arbiter (objectivity, publicity, intersubjectivity).

1. *Experience is what ultimately justifies belief*, for it is ultimately to your experience that your beliefs must be called to account. Its very ultimacy sometimes conceals this fact, for we rarely need to dig that deep. Our “derived evidence,” evidence that is based on immediate experience,<sup>3</sup> is rarely called into question in ordinary life. So, for example, you say you saw a bear and are challenged to provide evidence. You reply “Well, I saw a very large mammal with thick black fur foraging for berries.” That will usually suffice, for it is rare that we doubt that someone is able to accurately describe their experiences. But however socially awkward it might be, asking for further evidence for that claim is perfectly coherent and, in special cases, appropriate. When pushed to this deeper level, one has nothing else to appeal to besides one’s experiences. And note this as well. Once one *has* appealed to one’s experiences – construed here as the way things appear to them to be – there is no question of further evidence. Once you have said, “I had an experience with these properties (blackness, certain geometric patterns)” – or, if you prefer, “there seemed to be something large-ish and black-ish (etc.)” – calls for further evidence defending the claim that you had such an experience are wrong-headed (this feature makes them attractive to foundationalists as regress-stoppers). We have reached epistemic rock bottom. And of course, sometimes citing one’s experiences as evidence is the most natural thing in the world. I assert that the temperature has fallen. You ask me why I think this. I say that I feel cold. I do not have in mind the *fact that* I feel cold, but, rather, *my feeling* cold.
2. *Rational thinkers judge by their experience*. Given that our ultimate evidence consists in experiences, it is platitudinous that rational thinkers judge by their experiences. To continue the example introduced just above: given that you feel cold but didn’t before, it will, other things being equal, be rational for you to judge that the temperature has dropped. Or, consider another example: in your large backyard, you see a bird and can’t tell whether it is a female cardinal or a juvenile male. The female will be slightly more gray with slightly more orange beak and a slightly rounder body. You strain your eyes to get a better view and attend more closely to the bird and(!) to the experiences you are having (and note that the experience of the selfsame bird will change as you squint or put on your glasses, etc.). To come to a judgment concerning whether it is a female cardinal or a juvenile male, you will, if you are rational, judge by your experience. Note that the claim that one is attending to and being guided by features of one’s experience does *not* imply that experiences are the primary objects of knowledge (see Crane, 2011, for more on this) or that one is not *at the same time*<sup>4</sup> aware of the objects in the world – assuming that one is not hallucinating – which give rise to those experiences. Also note that being aware of a feature is not obviously itself a form of knowledge.

3. *Experience is a guide to truth: Evidence as sign, symptom, or mark.* The idea that, say, smoke is a sign of fire, is common enough. So, seeing smoke gives one evidence that there is fire. It is tempting to say that this is because smoke is a reliable indicator of fire. But what if it turned out that your experience was very atypical and that, in fact, most of the time smoke was not correlated with fire? You might say, "Well, it was a reliable indicator *in my experience*." Thus it seems that it is not *mere* objective correlation which makes something evidence. With this we are close to the notion of experience as evidence, but not all the way there. There is a notion that we might call "scientific evidence," where we are adopting an idealized third-person perspective; here, we say that something is evidence when we are already aware that Fs are positively correlated with Gs (see §1.1 of Conee and Feldman, 2008). But the notion epistemologists are interested in is such that all evidence is *someone's* evidence or evidence *for someone*. If your background experiences are different from mine, the same observation can be evidence that *p* for you and that  $\sim p$  for me.

In its basic sense, evidence is what *makes evident* some proposition. It is that in the light of which a proposition seems true. But this light is cast, fundamentally, by experience. So experience is our ultimate evidence. For example, we have all had, when considering some theorem of logic, that "aha" experience, the moment at and in virtue of which the theorem is made evident. It is most in evidence in self-evident propositions. It could be that self-evidence is factive, but it seems that as in all other areas, evidence can be misleading. The obviousness of, say, the naive axiom of comprehension seemed just as clear as the obviousness of some true axioms.<sup>5</sup> Experience reveals the world to us in perception. The features of a certain experience *make evident* to me that there is an elm before me, for example. Chisholm (who combined experience-first epistemology with direct realism) puts it this way: "In the case of being appeared to, there is something, one's being appeared to in a certain way, that one interprets as being a *sign* of some external fact."<sup>6</sup> Experiences are not the *objects* of knowledge, but they are the *medium* for knowledge.<sup>7</sup> For every state of knowledge, there is some experience that makes the fact in question evident. This is a central way in which experience is prior to knowledge. One's *feeling* cold is different from one's *being* cold, in the sense that one's core body temperature has dropped. But one's experience of coldness typically makes evident that one's core body temperature has dropped. In this way experience is the sign of what the world is like.

4. *Experience is a neutral arbiter among disputants.* How could private experiences play such a seemingly public role? Aren't they too "subjective"? It's not hard, actually. The way in which experience plays an intersubjective role is familiar to all. You are sailing with friends off the Gulf Coast, and in your periphery you think you see a dolphin jump. You turn to your friends and say "Did you see that?! I think I just saw a dolphin jump!" The friend standing closest says, "I saw something in that direction too, but it seemed too small to be a dolphin." Another says, "I thought I saw a white cap on the top of that shape." Another, "Me too." Your experience is overruled by their experiences. When the collective experiences are taken together, the evidence suggests that, though dolphin sightings are not infrequent here, what you saw was a stray wave rather than a dolphin. This is intersubjective in a way sufficient to satisfy all the constraints of scientific, legal, and medical inquiry.<sup>8</sup>

## Conclusion

As we mentioned above, it could well be that there is such a thing as derived evidence and that beliefs or propositions can play the roles too. Basic knowledge might be near the evidential foundations, but we reject both that all knowledge is evidence and that only knowledge is evidence. In fact, maybe we should all, like Thomas Reid, be *pluralists* about evidence. Reid says that “[w]e give the name of evidence to whatever is the ground of belief” (*IP* II 20, *W* 328a); and, he thinks, there are different types or sources of evidence: there is the evidence of sense, of memory, of consciousness, of axioms, of reasoning, and so on (*IP* IV 20, *W* 328a). Various kinds of experiences (perceptual, memorial, introspective), arguments, testimony, the judgment of recognized authorities, the marks or signs by which we distinguish between kinds of things, a person’s past actions, various “signs” of another’s mind and/or character (gestures, facial expression, etc.), observed connections in the world – these are all things which Reid seems to count as evidence.<sup>9</sup> Perhaps all these usages can be reduced to a single experiential notion, such as in phenomenal conservatism (see Huemer, 2001), or perhaps not. At any rate, there is no reason to think all and only knowledge is evidence, and there is ample reason to consider experiences a legitimate, and even the most basic form of evidence.

Of course, we have left many questions open – for instance, just *why* experience confers justification. That it does can be combined with any number of epistemological theories, internalist and externalist. Even externalist Alvin Plantinga recognizes that:

My perceptual beliefs are not ordinarily formed on the basis of *propositions about* my experience; nonetheless they are formed on the basis of my experience. You look out of the window: you are appeared to in a certain characteristic way; you find yourself with the belief that what you see is an expanse of green grass. You have *evidence* for this belief: the evidence of your senses. Your evidence is just this way of being appeared to; and you form the belief in question *on the basis of* this phenomenal imager, in this way of being appeared to. (1993, p. 98)<sup>10</sup>

The limitation of evidence to propositions (of any kind),<sup>11</sup> however, seems to us to reflect an over-intellectualization of inference and epistemic support.<sup>12</sup> To some extent, the focus on propositions is perfectly natural. As noted above, our everyday epistemic practices seldom oblige us to descend to the level of what we’ve talked about here as “ultimate evidence” (non-propositional experiences). And, when we do, the fact that we’re thinking and talking about such matters itself renders them into propositional form. Still, commitment to principles like “what gives probability must also receive it” (Williamson, 2000, p. 196) seems to lay evidence upon a Procrustean bed (not to mention the circularity or regress worries they might raise). We will say a bit more about this in our rejoinder.

## Notes

- 1 For more deflationary views, see for example Churchland (1987) and Stich (1990).
- 2 For a treatment of evidence that unifies the roles without assuming that any one thing can fulfill them all, see Rysiew (2011).
- 3 There is some question whether this is in fact a kind of evidence. See Littlejohn (2011).

- 4 Aquinas did not think this possible. See Adler (1985, pp. 14ff.).
- 5 Saying such things is, of course, fully compatible with a broadly reliabilist, or otherwise externalist (e.g., proper functionalist, etiological functionalist, etc.), approach to evidence and/or other epistemic goods. See Rysiew (2011).
- 6 Chisholm (1989, p. 67).
- 7 In the language of the medieval Aristotelian epistemologists, ideas are the *qua* of knowledge, not the *quod*. See, for example, Aquinas's "Treatise on the Powers of Man" in the *Summa Theologica*.
- 8 Of course, when we talk about evidence the way we do in a court of law – where physical objects in the room are referred to as "evidence" – we are, as always, speaking in a way that reflects what is mutually obvious: here, the fact that no one has any doubt that everyone is having the same kind of experience. That we in this way take for granted ultimate evidence speaks to its very ultimacy.
- 9 The discussion is not always explicit and systematic. For example, in comparing "the evidence of sense" with that of reasoning and consciousness, Reid doesn't come right out and say just what the evidence of sense *is*. Some of the relevant passages include: *IP* II 20, *W* 328aff.; VII 3, *W* 481bff.; VI 5, *W* 441aff.
- 10 Based on the way Plantinga has defined "direct awareness" on p. 53, it follows that we are not directly aware of the grass. But on p. 189, he says that "in another and perfectly good sense of "directly aware of," I *am* directly aware of [the object]."
- 11 Tim defends this move in §9.5 of Williamson (2000). Conee and Feldman (2008) respond, then Dougherty (2011) pressed Williamson's case further, but reconvered after reading Conee and Feldman's (2011) reply.
- 12 It would be an instance of the "argumentational" view of evidence discussed by Rysiew (2011).

## References

- Adler, M.J. (1985) *Ten Philosophical Mistakes*. New York: Touchstone.
- Bartsch, K. and Wellman, H.M. (1995) *Children Talk about the Mind*. New York & Oxford: Oxford University Press.
- Chisholm, R. (1989) *Theory of Knowledge*, 3rd edn. Englewood Cliffs, NJ: Prentice Hall.
- Churchland, P. (1987) Epistemology in the age of neuroscience. *Journal of Philosophy* 84: 544–553.
- Conee, E. and Feldman, R. (2008) Evidence. In *Epistemology: New Essays*, ed. Q. Smith. Oxford: Oxford University Press.
- Conee, E. and Feldman, R. (2011) Reply to Dougherty. In *Evidentialism and Its Discontents*, ed. T. Dougherty. Oxford: Oxford University Press.
- Crane, T. (2011) The problem of perception. In *The Stanford Encyclopedia of Philosophy*, ed. E.N. Zalta. <http://plato.stanford.edu/entries/perception-problem/> (last accessed February 27, 2013).
- Dougherty, T. (2011) In defense of a propositional theory of evidence. In *Evidentialism and Its Discontents*, ed. T. Dougherty. Oxford: Oxford University Press.
- Dougherty, T., ed. (2011) *Evidentialism and Its Discontents*. Oxford: Oxford University Press.
- Greco, J. (2010) *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cambridge: Cambridge University Press.
- Huemer, M. (2001) *Skepticism and the Veil of Perception*. Lanham, MD: Rowman & Littlefield.
- Kelly, T. (2011) Evidence. In *The Stanford Encyclopedia of Philosophy*, ed. E.N. Zalta. <http://plato.stanford.edu/entries/evidence/> (last accessed February 27, 2013).
- Kvanvig, J. (1992) *The Intellectual Virtues and the Life of the Mind: On the Place of the Virtues in Contemporary Epistemology*. Lanham, MD: Rowman & Littlefield.
- Kvanvig, J. (2003) *The Value of Knowledge and the Pursuit of Understanding*. New York: Cambridge University Press.

- Littlejohn, C. (2011) Evidence and knowledge. *Erkenntnis* 74: 241–262.
- Perner, J. (1993) *Understanding the Representational Mind*. Cambridge, MA: MIT Press.
- Planginta, A. (1993) *Warrant and Proper Function*. Oxford: Oxford University Press
- Reid, T. (1895/1983) *Thomas Reid: Philosophical Works*, 8th edn. Edited by W. Hamilton. New York: Georg Olms Verlag.
- Rysiew, P. (2011) Making it evident: evidence and evidentness, justification and belief. In *Evidentialism and Its Discontents*, ed. T. Dougherty. Oxford: Oxford University Press.
- Stich, S. (1990) *The Fragmentation of Reason*. Cambridge, MA: MIT Press.
- Williamson, T. (2000) *Knowledge and its Limits*. Oxford: Oxford University Press.
- Zagzebski, L. (1996) *Virtues of the Mind*. Cambridge: Cambridge University Press.

# Knowledge Still First

*Timothy Williamson*

Trent Dougherty and Patrick Rysiew (D&R) advocate “experience-first epistemology.” They regard experience as our only starting point for inquiry. That may sound like common sense: surely we must start with what we have learnt over the course of our lives. But D&R don’t mean the slogan that way. Climbing Toubkal is one of my valued experiences, but it isn’t an experience at all in their sense. For D&R, the “experience” was *what it was like to climb Toubkal* or *how climbing Toubkal appeared to me*, ways of experiencing I share with someone merely in an appropriate reality simulator, unlike climbing Toubkal itself. Thus their appeals to everyday comments such as “It was a reliable indicator in my experience” are quite misleading. D&R mean that *appearances* are our only starting point for inquiry, which is not common sense but theory much in need of argument. One might have hoped to start with obvious facts about the external world, but they are not appearances: the fact that there are other people is not the fact that there appear to be other people.

D&R argue that all chains of justification for belief ultimately lead back to appearances: if you cite something other than appearances, you can be challenged for its justification, but once you cite appearances, any such further challenge is “wrong-headed.” Well, if the challengers are D&R, they can challenge whatever is not an appearance and stop when they are given appearances, but that shows something about them, not about the epistemic role of appearances. Several of their illustrative appeals to experience, such as “There seemed to be something large-ish and black-ish (etc.),” raise an immediate difficulty. Appeals to *past* experience can easily be challenged by other evidence: the police may discover that immediately after the sighting I wrote in my diary, “There did not seem to be anything large-ish or blackish,” and that my memory is notoriously unreliable. Perhaps D&R’s repeated use of the past tense is a (revealingly natural) slip; on reflection they may prefer to restrict one’s ultimate justifiers to one’s *present* experiences. The need for such a restriction dramatizes their distance from the commonsense conception of judging on the basis of one’s experience.

D&R’s inchoate account of intersubjective evidence is revealing too. They treat it as a matter of pooling appearances to different subjects, not mentioning that on their view each subject starts only with the appearances of the others’ reports of their

appearances, on a par with other appearances of the external world, so that nobody is in a position to do the pooling. Their glib comment that “It’s not hard, actually” suggests inexperience of the complex problems facing any attempt to develop a serious theory of shared scientific evidence within their framework (problems related to the failure of Carnap’s program in the *Aufbau*).

Even when we restrict discussion to present appearances to the subject, D&R’s theory is quite implausible. For example, suppose I see two complicated figures and report that the figure on the left appears to me in exactly the same way shapewise as the figure on the right. When someone points out a difference in shape between the two figures, I notice it immediately. A plausible hypothesis is that all along the figure on the left appeared to me in a slightly different way shapewise from the figure on the right, but I failed to notice the difference; even if the two figures appeared to me to be exactly the same in shape, it does not follow that the way they appeared shapewise was exactly the same. Since such errors about present appearances to one are possible, I may be appropriately challenged to justify my claim that the two figures appear to me in exactly the same way shapewise, and in other circumstances may aptly reply that my vision is normal and the two figures were cast from the same mold. The idea that justification ends with appearances is neither a datum nor a discovery, but wishful thinking in the service of an outdated epistemological theory.

Somehow attributing ultimate justification to the appearances themselves, rather than facts about them, does not help. To vary the example, suppose that although two figures really do appear to me in exactly the same way, my past record provides massive evidence that I have failed to notice a slight difference between two ways of appearing. Am I justified in claiming that the ways of appearing are identical? D&R’s account is too undeveloped to engage with such questions.

D&R later invoke Jeffrey conditionalization to handle uncertain evidence, but the appeal is notoriously empty without constraints on which Jeffrey conditionalizations are epistemically permissible in given circumstances, which they do not attempt to provide. Instead, as D&R interpret Jeffrey, “the assigning of probabilities to basic propositions is a *techné* – it requires an art of judgment that is generally acquired by practicing in the appropriate community.” Maybe so, but that does not answer the question, for it says nothing about the content of the norm good judges are better at satisfying than bad ones. Changing horses, D&R claim that “on an experiential theory of evidence wedded to a Chisholm-like theory of evidence” we “end up with a form of epistemic probability that strikes the appropriate balance between subjectivity and objectivity,” with no hint of the hard formal and interpretative work needed to turn such programmatic remarks into a serious option.

D&R write: “it seems to us that the cost-to-benefit ratio is about the same in formal and non-formal modes of inquiry, when pursued conscientiously by capable parties.” But we do not have to choose between exclusively formal and exclusively non-formal modes of inquiry. A better mode of inquiry than either combines formal and informal considerations, using informal reasoning to select appropriate formal models and formal models to discipline informal reasoning. The key is to integrate the two sides. Although various forms of Bayesianism are formally consistent with experience-first epistemology, D&R do not come to grips with the challenge of integrating them into a combined approach.

In “Knowledge First” I discussed the underlying motivation for experience-first epistemology. One source is the principle that justification is exactly the same in cases

indiscriminable to the subject. I used the non-transitivity of indiscriminability to refute that principle. D&R complain that my argument falsely “assumes experiences are fully determinate rather than vague.” Their complaint is mistaken. My argument makes no such assumption, as the reader can check. It works perfectly well if experiences are vague. Thus the refutation of the principle stands.

Although much more could be said, the upshot remains that D&R have failed to rehabilitate experience-first epistemology. I turn to their objections to knowledge-first epistemology. For all their puzzlement about its nature, it is an easily recognizable way of doing epistemology. Naturally there are borderline cases, just as there are for the rival ways of doing epistemology, such as D&R’s. For reasons of space, I focus on just four issues.

1. D&R wonder how knowledge-first epistemology can explain the presumed entailment from knowledge to belief. They make heavy weather of my suggestion that knowing stands to believing roughly as doing stands to trying. The picture is that when one (intentionally) does something, one tries to do it and succeeds, although for Gricean reasons it is conversationally misleading for someone to say just that one tried to do it. Thus doing entails trying. The converse fails; unsuccessful trying is mere trying, botched doing, trying without doing. Similarly, when one knows something, one believes it; one as it were aspires to know it and succeeds, although for Gricean reasons it is conversationally misleading for someone to say just that one believes it. Thus knowing entails believing. The converse fails; believing that falls short of its aspiration to knowledge is mere believing, botched knowing, believing without knowing. Of course, such analogies never go all the way, otherwise they would be identities, but D&R’s comments raise no difficulty for the picture just sketched.

Equally puzzling is D&R’s alternative explanation on behalf of traditional epistemologists. The key is supposed to be that “entailment is modeled in formal semantics as set inclusion.” But epistemologists of any persuasion can use that commonplace to articulate the entailment from knowing to believing: the set of cases in which one knows that P is a subset of the set of cases in which one believes that P. D&R have identified no substantive issue here.

2. D&R suggest that *understanding* may trump knowledge as an epistemic value. The presupposition that understanding does not involve knowledge is widespread (for example, in discussions of what philosophy aims at), but hardly withstands scrutiny. If you do not know how a jet engine works, you do not understand how it works. If you do not know why the plane crashed, you do not understand why it crashed. Fortunately, D&R allow understanding to involve knowledge. In that case, they suggest, understanding is “a special subset of knowledge where (at a minimum) certain internalist goods are added (seeing connections, etc.),” and these goods may have a value independent of knowledge. They supply no warrant for the gloss “internalist” here. Understanding why the plane crashed may involve seeing the connection between the pilot’s tiredness and his poor decision making. What’s internalist about that? What matters is whether the investigators come to know that the pilot’s tiredness was connected in the relevant way with his poor decision making, and then use that knowledge to acquire other relevant knowledge of the crash. There is no comfort here for attempts to displace knowledge as the central epistemic value.



An essential prerequisite for useful discussion of the relation between knowledge and understanding is systematic explicitness about *what* is to be known or understood. The better that discipline is maintained, the less understanding looks like a rival value to knowledge. Rather, some things are more worth knowing than others. Knowing why or how something happened is often much more valuable than just knowing that it happened. We should not imagine that knowing why or how something happened is non-propositional knowledge, just because it goes beyond knowing that it happened. Knowing why it happened is knowing that it happened for a certain reason; knowing how it happened is knowing that it happened in a certain way. In knowing the answers to some questions, we know far more than we do in knowing the answers to other questions. A good challenge for any epistemologist is to develop a non-trivial theory about that dimension of comparison. But such an investigation need have no tendency to downgrade the central value of knowledge.

3. "The rigorous development of knowledge-first epistemology began quite recently," I noted, compared to the corresponding development of experience-first epistemology; I gave as one sign that the latter is a stagnating research program its relative lack of progress in generating epistemologically illuminating formal models. In response, D&R make various historical observations. As they indicate, knowledge-first epistemology itself is not new; it has deep roots in the Aristotelian tradition (the introduction to *Knowledge and its Limits* starts with an epigraph from Aristotle's *Metaphysics*). My point concerned its recent integration with formal epistemology, in the shape of both epistemic logic and objective Bayesianism. Although formal epistemology in an experience-first spirit goes back much longer, attempts to integrate the formal and informal sides remain epistemologically naive or empty, in particular with respect to the general theory of evidence. As already explained, D&R do not come to terms with this challenge.
4. D&R repeatedly emphasize the variety of rivals to knowledge-first epistemology, including approaches that accept some knowledge-first theses while rejecting others. As in natural science, you don't get far in criticizing a research program just by listing its rivals, without showing that one of them is doing better, or at least equally well. In the long run, we can adequately judge a serious program only by letting its proponents develop its potential. Experience-first epistemology has had a long run, with poor results. Time for alternatives, of which knowledge-first epistemology has made most recent progress.

## Still Nowhere Else to Start

*Trent Dougherty and Patrick Rysiew*

Unlike climbing Toubkal, something's looking tall or seeming upon reflection to be true aren't typically things to write home about. Often, such experiences aren't noticed (as such) at all. For well-functioning humans in normal environments, however, they

too are ways of being engaged with the world. That some facts about that world are obvious (evident), and that we sometimes start there, is perfectly compatible with our view, as is the fact that such beliefs as we have about our experiences can be mistaken.

Williamson mentions that (in our framework) pooling evidence requires the use of testimony as if that's a problem. But "a serious theory of shared scientific evidence" will surely be an endeavor in social epistemology involving trust and an acknowledgment of the epistemic significance of (seeming) comprehension of others' presentations-as-true.

Williamson tends to associate experience-first epistemology with some pretty radical forms of internalism (e.g., Descartes's, Carnap's). But, as it was for Reid (hardly a Cartesian!), it is for us the "*belief-evoking experiences* characteristic of [our] faculties" (Wolterstorff, 2004, p. 80, emphasis added), and of evidence generally, that is epistemically central. "The evidence of sense," for instance, "is neither the proposition itself nor some other proposition but the sensory experience one is having" (Wolterstorff, 2004, p. 83; see Plantinga, 1993, p. 98 for a contemporary externalist avowal of experience as evidence).

Regarding the phenomenal sorites argument, as we've stressed, the "experience first" approach doesn't stand or fall with the idea that argument targets. Even so, we ask that the reader consult the footnotes and decide for themselves whether the argument succeeds, and to consider as well whether it doesn't require the dubious thesis that there can be indiscernible non-identical qualia.

We remain confident that we have described a viable and vital research project, formal and informal.

## Acknowledgments

Thanks for comments are due to Robert Audi, Earl Conee, John Hawthorne, Scott Sturgeon, Mike Huemer, Kevin McCain, John Matheson, and Evan Fales.

## References

- Plantinga, A. (1993) *Warrant and Proper Function*. Oxford: Oxford University Press.  
Wolterstorff, N. (2004) Reid on common sense. In *The Cambridge Companion to Thomas Reid*, ed. T. Cuneo and R. van Woudenberg. Cambridge: Cambridge University Press.

## Chapter Two

# Is Knowledge Closed under Known Entailment?

If you know that P is true, and you know that the truth of P guarantees the truth of Q, then you are thereby automatically in a position to know that Q is true. All you need to do is competently deduce Q. In short, knowledge is closed under known entailment. That is, roughly, what the epistemic closure principle says. Fred Dretske rejects the closure principle. Dretske points out that closure principles for mental states other than knowledge, such as regret, are false. Dretske also points out that there are intuitive counterexamples to the epistemic closure principle. For example, if you're at the zoo and see black-and-white striped equine animals in the pen marked "Zebra," you know that the animals are zebras; and their being zebras guarantees that they're not cleverly painted mules; but many think you don't know that they're not cleverly painted mules. In his response to Dretske, John Hawthorne usefully distinguishes several versions of the closure principle, and defends his preferred version. Hawthorne argues that if we give up his preferred version of the closure principle, we'll have to also give up the "Distribution" principle, which says that if you know the conjunction P and Q, and you're able to deduce P therefrom, then you're in a position to know P. Hawthorne argues furthermore that Dretske's view runs into trouble when we combine it with the very plausible thesis that knowledge is the norm of assertion. Hawthorne also raises several problems for Dretske's positive theory of knowledge, which features the notion of conclusive reasons.

## The Case against Closure

*Fred Dretske*

Closure is the epistemological principle that if S knows that P is true and knows that P implies Q, then, evidentially speaking, this is enough for S to know that Q is true. Nothing more is needed. If S believes Q on this secure basis – on the basis of two things

he knows to be true – then S knows that Q is true. One knows everything that one knows to be implied by what one knows.

It is important to distinguish closure from *modus ponens*, a principle of logic that it superficially resembles. Closure is stronger. *Modus ponens* says that if P is true, and if P implies Q, then Q must be true. Closure tells us that when S knows that P is true and also knows that P implies Q, then not only must Q be true (*modus ponens* gets you this much), S must *know* it is true.<sup>1</sup>

Why must S know Q to be true? Why does one have to know everything one knows to be implied by what one knows? One doesn't, after all, have to regret everything one knows to be implied by what one regrets. Tom regrets drinking three martinis last night, but he doesn't regret what he knows to be implied by this – drinking *something* last night. Nor does he regret the reality of the past even though he knows that drinking three martinis last night implies that the past is real. If

- (R) S regrets P  
S knows that P implies Q  
Therefore, S regrets Q

is not a valid argument, why should

- (K) S knows P  
S knows P implies Q  
Therefore, S knows Q

be accepted as valid?

This question defines the issue between John Hawthorne and me. I will be taking the negative side of the debate: knowledge is *not* closed under known logical implication. (K) is *not* a valid argument. There are things we know to be implied by what we know that we do not know to be true. Just as a great many conditions I do not regret have to exist (and I know they have to exist) in order for the condition I regret to exist, a great many conditions I do not know to exist have to exist (and I know they have to exist) in order for the conditions I know to exist.

## 1 Transmissibility

Let me begin this examination by talking about a problem that is closely related to closure: transmission of evidential warrant. I shall come back to closure later (in sections 2 and 3).

Our ways of discovering P are not necessarily ways of discovering what we know to be implied by P. From the fact that you know that P implies Q, it does not follow that you can see (smell, feel, etc.) that Q just because you can see (smell, feel, etc.) that P. Despite knowing that cookies are objective (mind-independent) objects, I can see (roughly: tell by looking) that there are cookies in the jar without being able to see, without being able to tell by looking, that there are mind-independent objects. A claim to have found out, by looking, that there are cookies in the jar is not a claim to have found out, by looking, that there is a material world. Maybe one has to know there are

physical objects in order to see that there are cookies in the jar (we will come back to that), but one surely isn't claiming to see that there are physical objects in claiming to see there are cookies in the jar. After all, hallucinatory cookies "in" hallucinatory jars can look exactly like real cookies in real cookie jars. So one cannot, not by vision alone, distinguish real cookies from mental figments. One cannot see that the world really is the way it visually appears to be. A way of knowing there are cookies in a jar – visual perception – is *not* a way of knowing what one knows to be implied by this – that visual appearances are not misleading.

The claim to have found out, by visual means, that there is still wine left in the bottle ("Just look; you can see that there is") is not a claim to have found out, at least not by visual means, what you know to be implied by this – that it, the liquid in the bottle, is not merely colored water. You know it is wine because, let us suppose, you tasted it a few minutes ago. That you learned it is wine, and not merely colored water, by tasting does not prevent you from now (minutes later) seeing that there is still some wine left in the bottle. Seeing that P does not mean you can see that Q just because you know P implies Q.

In *Seeing and Knowing* (1969) I tried to describe this phenomenon by introducing a technical term: *protoknowledge*. Protoknowledge was a word I made up to describe things that had to be true for what you perceived to be true but which (even if you knew they had to be true) you couldn't perceive to be true. That there are material objects has to be true for there to be cookies in the jar (and, hence, for you to see that there are cookies in the jar), but it isn't something you see to be so. It is a piece of (what I was calling) protoknowledge. When you see, just by looking, that there is wine in the bottle, the fact that it is not colored water is also protoknowledge. It has to be true for there to be wine in the bottle (what you perceive to be so), but it is not a fact that (normally) you can see to be so. In describing how one knows there is still wine left in the bottle one is not (not normally) describing how one knows it is not merely colored water.

Now, nearly half a century later, as a result of work by Davies (1998, 2000) and Wright (2003), there is, I think, a more revealing way of making this point. The idea is that some reasons for believing P do not *transmit* to things, Q, known to be implied by P. In normal circumstances – a wine tasting party, say – one's reasons for thinking there is wine left in the bottle do not transmit to, they are not reasons for believing, what you know to be implied by this – that the liquid in the bottle is not merely colored water. Having already tasted it, you may know that it is wine and not just colored water, but the point is that your reasons for believing the one (visual) are not reasons for believing the other (gustatory). Colored water in the bottle would look exactly the same as the wine. The reasons you have for believing what you say you perceive (there is wine left in the bottle) are not *transmitted* to this known consequence (that it is not merely colored water) of what you perceive.

The non-transmissibility (to many of the known consequences) of most of our reasons for believing P is an absolutely pervasive phenomenon. I will, in this section, spend a little more time harping on about it since I think it critical for evaluating the plausibility of closure. Non-transmissibility does not itself imply the failure of closure since, as our wine example illustrates, even when S's reasons for believing P do not transmit to a known consequence Q, it may be that S must still know Q (perhaps on the basis of other reasons) in order to know P. Even though S cannot see that it is wine, not just colored water, it may turn out (this is what closure tells us) that S must know (perhaps by earlier tasting) that it

is not colored water if he is now to see (hence, know) that there is wine left in the bottle. Nonetheless, once one appreciates the wholesale failure of evidential transmission, the failure of closure is, if not mandatory, easier to swallow. Or so I will argue.

As we have already seen, perception, our chief (some would say our only) route to knowledge of the world around us, does not transmit its evidential backing to *all* the known consequences of what is perceived. We can see (hear, smell, feel) that P, but some of the Qs that (we know) P implies are just too remote, too distant, to inherit the positive warrant the sensory evidence confers upon P. When Jimmy peeks into the cookie jar and, to his delight, sees that there are cookies there, his visual experience of the cookies, the evidential basis for his knowledge that there are cookies there, is not evidence, not a reason to believe, that there is a physical reality independent of Jimmy's mind. Jimmy's experience of the cookies may be good reason to believe there are cookies in the jar, but it is not a good reason to believe that idealism is false. And it is not a good reason to believe that idealism is false even if Jimmy understands that cookies are mind-independent objects and that, therefore, what he sees to be the case (that there are cookies in the jar) implies that idealism is false. Looking in the cookie jar may be a way of finding out whether there are any cookies there, but it isn't – no more than kicking rocks – a way of refuting Bishop Berkeley.

So perceptual reasons – the sense experiences on which we base everyday perceptual judgments – do not transmit their evidential force to all the known consequences of the judgments they warrant.

- (T) R is a reason for S to believe P  
S knows that P implies Q  
R is a reason for S to believe Q

is not a valid argument. When we perceive that P, there are, so to speak, *heavyweight* implications of P that cannot be perceived to be so. There is transmission to *lightweight* implications, of course. If I can see that there are cookies in the jar, I can certainly see that the jar isn't empty, that there is something in the jar. For perception, though, there are *always* heavyweight implications, known implications to what one perceives (P) that one's perceptual reasons for P are powerless to reach. If there is any doubt about this, simply imagine Q to be a condition S knows to be incompatible with P but which, because of (perhaps extraordinary) circumstances, has the same sensory effects on S as P. Though incompatible with P, Q (as so specified) will look (feel, sound) exactly the same to S as P. Q is, as it were, a perceptual twin of P. With such a Q, S will not be able to perceive ~Q though he perceives P and knows that P implies ~Q.

Skeptics have used this formula to manufacture heavyweight implications for all our ordinary perceptual claims. No matter what you purport to know by perception, perception will not be the way you know that you are not being deceived by a Cartesian demon. It will not be the way you know you are not a disembodied brain in a vat being caused (by suitably placed electrodes) to experience the things an ordinary veridical perceiver experiences. One cannot see (hear, smell, feel, etc.) that one is not in this unfortunate position even though one knows one cannot be in this position and still be seeing (hearing, etc.) the things one believes oneself to be seeing (hearing, etc.). You can't see that you are not dreaming.

## 2 Closure

What does this have to do with closure? When you know that P implies Q, closure tells us you have to know Q to know P. It doesn't tell us that the reasons that promote your belief that P into knowledge that P must themselves promote your belief (assuming you have it) that Q into knowledge that Q. Maybe the reasons (see section 1) for P don't transmit to Q. Very well. Then closure tells us that if you know P you have to have *other* reasons (or, if reasons are not always required, whatever else it takes) to promote your belief that Q into knowledge that Q. But – or so closure says – you have to have *something* to make your belief that Q into knowledge in order to know P. If you don't, then – too bad – you don't know that P. Nothing I have said so far challenges this claim.

Is this true? Is it true that someone (who knows that cookies are material objects) cannot see that there are cookies in the jar unless he or she knows that there is a physical, mind-independent world? That they are not being misled by some clever deception? In Dretske (1970) I asked the same question about my ability to see that an animal is a zebra without knowing it was not a cleverly painted mule. It seemed to me then, as it still seems to me today, that if this is true, if knowledge is closed under known implication, if, in order to see (hence, know) that there are cookies in the jar, wine in the bottle, and a zebra in the pen, I have to know that I am not being fooled by a clever deception, that the “appearances” (the facts on which my judgments are based) are not misleading, then skepticism is true. I never see (to be so), and hence do not know (to be so), what I think and say I see (to be so).<sup>2</sup> This is not to deny that we know – and, perhaps, necessarily know – *most* of the implications of what we know. But not *all* of them. There are some things – *heavyweight* implications – we needn't know even though we know our knowledge depends on their truth.

Though I haven't performed a scientific study, my impression is that most philosophers (who bother thinking about it) believe that this is preposterous. Feldman (1999) thinks that abandoning closure is “one of the least plausible ideas to gain currency in epistemology in recent years.” DeRose (1995) finds it “intuitively bizarre” or “abominable.” Fumerton (1987) thinks the failure of closure is a “devastating objection” and Bonjour (1987) a *reductio ad absurdum* to any theory that implies or embraces it. Most philosophers, of course, are eager to reject skepticism. Rejecting closure as a way around skepticism, though, is quite another matter. As these reactions indicate, many philosophers are unwilling to do it. Skepticism is bad, yes, but for many philosophers, not *that* bad. Not bad *enough* to justify rejecting closure. That would be like rejecting logic because it gave us conclusions we didn't like.

Some of these reactions are, I think, a bit overdone. To deny closure is not to say that you can *never* know (find out, discover, learn) that Q is true by inferring it from a P that you know to be true. It is merely to deny that this can be done for *any* Q. Nor is denial of closure a way of embracing (without further explanation) DeRose's (1995) abominable conjunctions. Statements like “I know I have hands, but I do not know that I am not a handless brain in a vat” (a statement that denial of closure accepts as being possibly true) are clearly ridiculous. We can all agree about that. The question, however, is not whether anyone would ever say it (or, if they did, whether they would be greeted by stares of incomprehension) but whether it might be true if they said it. “The

refrigerator is empty, but has lots of things in it” is also an abominable conjunction. It might, nonetheless, be true. There may be no food or other items normally stored in refrigerators inside (thus making the refrigerator empty in the normal way of understanding what isn’t in empty refrigerators), but it may, nonetheless, be filled with lots of gas molecules (perfectly respectable things). The abomination in saying a refrigerator is empty but has lots of things in it comes not from any logical defect (there are lots of empty things that have gas molecules in them – pockets, classrooms, warehouses, etc.), but from a violation of normal expectations. In describing an object as a refrigerator (and not, say, a metal container) one is led to expect that the things that are in (or, in the case of its being empty, not in) it are the sorts of perishable items normally stored or preserved in refrigerators. To then include (second conjunct) gas molecules as things in refrigerators is to flout this entirely reasonable expectation about what sorts of things are to be counted as things for purposes of describing the contents of refrigerators. Why isn’t it ridiculous, for exactly the same reason, to say one knows one has hands but doesn’t know one isn’t a handless brain in a vat? The second conjunct introduces possibilities normally assumed to be irrelevant (not counted as possibilities) by someone who asserts the first conjunct.

This is only to say (Grice, 1967) that there are logical abominations (self-contradictory) and conversational abominations (perfectly consistent, and therefore possibly true statements, that violate conventional expectations). To say that S knows that there are cookies in the jar but doesn’t know he isn’t hallucinating them is certainly to say something absurd, but why suppose its absurdity is such (i.e. logical) as to render the statement incapable of being true? To demonstrate logical incoherence would require a theory about what it takes (and doesn’t take) to know something, and we are back to where we started: assessing the status of closure.

So I don’t think we (e.g., John Hawthorne and I) can just match brute intuitions on closure. Yes, closure sounds like an eminently plausible principle. Everything else being equal, then, we ought to keep it. But everything else *isn’t* equal. It is also plausible (some would say it was entirely obvious) that we know things about our material surroundings – that (when we see them for ourselves) there are cookies in a jar, zebras in the zoo, people in the room, and cars on the street. And it isn’t at all clear that we can have such knowledge while, at the same time, retaining closure. Something has to give. This reasoning won’t impress a skeptic, of course, but, if a case could be made for the claim that a rejection of closure was not just *a* way to avoid skepticism (most philosophers would agree with this) but the *only* way to avoid skepticism, it should carry weight with philosophers who find skepticism as “bizarre” or “abominable” as the denial of closure. This will be my strategy in the remainder of this paper. The only way to preserve knowledge of homely truths, the truths everyone takes themselves to know, is, I will argue, to abandon closure.

Philosophers, though, will insist that there are other, less costly ways of dealing with skepticism, ways that do not require the rejection of closure. One of the popular maneuvers is to make knowledge attributions sensitive to justificational context. In ordinary contexts in which knowledge claims are advanced, one knows (such things as) that there are cookies in the jar and (at the zoo) zebras in the pen. In such ordinary contexts one normally has all the evidence needed to know. One sees the cookies and the zebras and, given the absence of any special reason to doubt, this (for reasonably experienced adults) is good enough to know. So skepticism is false. Do these reasonably



experienced adults also know, then, that they are not being deceived by some kind of fake? That the sensory evidence on which they base their judgments is not misleading? That there really is a material world? That they are not a brain in a vat? Well, yes (so closure is preserved), but – and here is the kicker – only as long as they don't seriously consider these heavyweight implications or say that they know them to be true. For once they think about them or say they know them to be true, the context in which knowledge claims are evaluated changes and knowledge evaporates. In this altered context (no longer an ordinary context) one doesn't know that one is not being deceived because new alternatives (that one *is* being deceived), possibilities one cannot evidentially eliminate, have been introduced. Therefore (closure forces one to say) one no longer knows that there are cookies in the jar and zebras in the pen. One gets to know about cookies and zebras only as long as one doesn't think about or claim to know what, according to closure, one has to know in order to know such mundane things. According to this way of dealing with skepticism, philosophers who spend time worrying about heavyweight implications (How do I know I'm not dreaming? How do we know there is a material world?) are the most ignorant people in the world. Not only don't they know these heavyweight implications (maybe no one does), they don't (like everyone else) know the things (that there are cookies in the jar, zebras in the pen) that imply them. This, of course, includes almost everything they thought they knew.

This result, it seems to me, is pretty bizarre – *more* bizarre, in fact, than abandoning closure. It is a way of preserving closure for the heavyweight implications while abandoning its usefulness in acquiring knowledge of them. One knows (or is evidentially positioned to know) heavyweight implications only so long as one doesn't think about them or say that one knows them to be true. I think we can do better than this. By suitable restriction of closure (to lightweight and middleweight implications) we can have most of our cake (the only part worth eating) and eat it too.

Before launching this argument, though, let me mention a theoretical consideration that deserves attention in this regard. It is not, in the end, decisive on the issue of closure, but it is certainly relevant and with some readers it may carry weight. It does with me.

There is what I (and a great many others) regard as a plausible approach to the analysis of knowledge that leads quite naturally (not inevitably,<sup>3</sup> but naturally) to a failure of closure. If knowledge that P requires one (or one's evidence) to exclude not all, but only all relevant, alternatives to P, then, it seems, one is committed to a failure of closure. The evidence that (by excluding all relevant alternatives) enables me to know there are cookies in the jar does not enable me to know that they (what I see in the jar) are not papier mâché fakes since papier mâché fakes are not (usually) a relevant alternative. So the evidence that gives me knowledge of P (there are cookies in the jar) can exist without evidence for knowing Q (they are not fake) despite my knowing that P implies Q. So closure fails.

Although this is, as I say, a natural line of reasoning, it isn't very effective against the skeptic if the only reason for embracing a relevant alternative analysis of knowledge is that it captures commonsense (basically anti-skeptical) intuitions about when and what we know. If, that is, one's reasons for accepting a relevant alternatives analysis of knowledge is that it accords with our ordinary practice of claiming to know something (that there are cookies in the jar) without having specific evidence against possible mistakes (that, for example, they are not papier mâché fakes), then the

argument against closure becomes too tightly circular to be effective against a skeptic. One uses premises – basically what is ordinarily regarded as the kind of evidence good enough to know – that no self-respecting skeptic would concede.

A dialectically more effective strategy is to provide an independent (of skepticism) analysis of knowledge that yields the result that only certain alternatives (to what is known) are evidentially relevant. This is the strategy adopted in Dretske (1969, 1970, 1971) and Nozick (1981). If knowledge is belief based on the kind of conclusive reasons I describe in Dretske (1971), for instance, then closure fails. Things turn out this way because one can have conclusive reasons,  $R$ , for  $P$  ( $R$  is a conclusive reason for  $P \equiv_{df} R$  would not be true unless  $P$  were true) without having conclusive reasons for known consequences of  $P$ . For example,

- (C) It would not look to me as though there were cookies in the jar if there weren't cookies in the jar.

can be true – thus making the look of the jar (roughly: my experience of the jar) a conclusive reason for believing there are cookies in it – while

- (C\*) It would not look to me as though there were cookies in the jar if there were cleverly made papier mâché cookies in the jar.

is false – thus preventing the experience from being a conclusive reason for believing what I know to be implied by there being cookies in the jar – that they are not merely papier mâché fakes in the jar.

Unfortunately, though, the world has not rushed to embrace this theory of knowledge (or its information-theoretic successor in Dretske, 1981). Nonetheless, I think many of the arguments for something like a counterfactual condition on knowledge are good arguments, and I have yet to see effective counterarguments.<sup>4</sup> I won't, however, recapitulate the history here. One doesn't make reasons better by repeating them. Suffice it to say that if one is sympathetic to something like (C) as a condition on seeing (hence, knowing) that there are cookies in the jar, one has taken a critical step towards denying closure. One has accepted a condition on knowledge that need not be, and often isn't, satisfied by known consequences of what is known.

### 3 Heavyweight Implications

I argued in section 1 that some ways – the perceptual ways – of coming to know  $P$  do not transmit evidential grounding for  $P$  to all the known implications of  $P$ . Ordinary things we come to know by perception always have heavyweight implications that are out of range: we cannot see (hear, smell, or feel) that they are true. If closure is not to result in complete skepticism about the external world, then, it seems that there must always be other ways of knowing that these heavyweight implications are true. I can see that there are cookies in the jar, but I cannot see that there is an external world. So if we have to know there is an external world in order to see there are cookies in a jar, as closure tells us we must, then there must be other (than visual) ways we know that there is an external world. There must be a way other than perception in which I know

that I'm not a brain in a vat, that I'm not being massively deceived, that solipsism is false, that it is not all just a dream. What might these other ways of knowing be?

It is hard to see what other ways there could be since every way of knowing either fails to reach these heavyweight implications or generates its own heavyweight implications. *No* evidence<sup>5</sup> transmits to *all* the implications of what it is evidence for.

Testimony, for instance, is considered to be an important source of knowledge that is distinct from perception. I didn't see the tire for myself, but I know it is flat because she told me it was flat. That is how I know. Yet, although I know that tires are material objects, I can learn, by being told, that I have a flat tire without thereby learning that solipsism is false – that there is something else in the universe besides me and my own ideas. No one, as I recall, ever told me there was something else besides me in the universe, but, if someone (a philosophy professor?) did, it is hard to believe that *that* is the way I learned there are other minds and a material world. Yet being told is the way I come to know things – a great many things – that I know imply there is a material world. That there is a material world is something we know to be implied by what we are told (flat tires *are* material objects independent of me and my own thoughts and experiences) but it is not something we learn by being told. So, if closure holds, testimony – and, remember, this includes not only books, newspapers, and television but virtually our entire educational system – cannot be a source of knowledge unless we have some *other* way of finding out there is a material world. Some *other* way of discovering that solipsism is false.

Or consider memory. Although memory is not generally considered a way of *coming* to know (it is more like a way of *preserving* knowledge acquired by other means) it, too, has its heavyweight implications. Whatever cognitive mechanisms enable a person to preserve, by memory, the knowledge that P do not enable the person to preserve a knowledge of P's known consequences. If you remember having granola for breakfast this morning, then you know you had granola for breakfast this morning.<sup>6</sup> If you remember going to the bank before stopping at the bakery, then you know you went to the bank before stopping at the bakery. But, as we all know, we couldn't have had granola for breakfast this morning, or stopped at the bank before going to the bakery, unless time, the succession of events, and, in particular, the past is real. That the past is real, however, is not something one remembers. It doesn't sound like something one *can* remember. This implication is more like a presupposition – something one simply takes for granted – in remembering the nitty-gritty details of personal history – that one had granola for breakfast or went to the bank before going to the bakery. For most of the things we remember (*most* but perhaps not all: we sometimes remember what we are supposed *to do*) the reality of the past is a heavyweight implication. It is something we know to be implied by things we remember, but it is not itself something we remember. Our way of preserving knowledge is not a way of preserving what we know to be implied by what we preserve. If closure is true, though, we cannot remember what we had for breakfast this morning unless we know the past is real. Since memory isn't a way of finding out the past is real, what, then, is the method we have for finding this out? Perception? Testimony? Fossils? I don't think so. Neither did Bertrand Russell when he wondered how we could know the world wasn't created minutes ago complete with history books, fossils, and memory traces. If there is a past, fossils, memory, and history books tell us what happened in it – the historical details, as it were – but they cannot tell us what is implied by the existence of these details: that there actually was a past.

Instead of trying to catalogue all the different ways of coming to know – probably a hopeless task, anyway – let me try approaching this topic from a more general standpoint. Generally speaking, when we come to know that P is true, there is always some indicator, some source of information, that “tells us” P is true. In the case of direct perception, this is an experience of the very condition “P” describes. We come to know there are cookies in the jar by seeing (experiencing) the cookies themselves. In testimony the information arrives by more indirect means: friends, teachers, radio and television newscasters, witnesses, books, newspapers, and magazines act as intermediaries. The information arrives in verbal form: someone tells me – perhaps by scribbling a note – that so-and-so happened. We also rely on natural signs – tracks in the snow, rings in the stump, cloud formations, facial expressions, and spontaneous behavior (yawns, sneezes, etc.). Photographs and measuring instruments indicate, they carry information about, conditions we do not – sometimes cannot – directly perceive. In each case, our way of coming to know that P always seems to depend on there being something that, by its revealing properties, indicates that P is true. There is always something that brings this information to us. It is important, therefore, to understand that the relationships we describe by speaking of one thing *indicating* or *carrying information* about another condition, P, are relations that do not relate this thing to (all) the known implications of P. Though we rely on a measuring instrument to indicate the value of Q – that it is, say, 5 units – the measuring instrument does not indicate, does not carry information, that Q is not being misrepresented. The position of a pointer (pointing at “5”) on a well functioning meter indicates that the value of Q is 5 units, but it does not thereby indicate that the instrument is not broken, not malfunctioning, not misrepresenting a value of 2 units (say) as 5 units. Even when instruments (and this includes the human senses) are in perfect working order, they do not – they cannot – carry information that what they are providing is genuine information and not misinformation. That isn’t an instrument’s job. The job of a Q-meter is to tell us about Q, not about its reliability in providing information about Q. So even though we know that if Q really is 5 (as the meter says), then Q can’t be 2 being misrepresented (by the meter) as 5, the meter gives us the first piece of information (tells us that Q is 5) without giving us the second piece of information: that Q isn’t really 2 being misrepresented as 5. That isn’t information the Q-meter supplies. It is a heavyweight implication of the information a Q-meter supplies. Anyone who really wanted to know whether the instrument was misrepresenting Q (when it registered “5”) should not look more carefully at the pointer. A pointer pointing at “5” does not carry *that* information. The information it carries (when things are working right) is that Q is 5, not that Q doesn’t just appear to be 5.

This is true of all indicators, all sources of information. That is why there is nothing in the world – either mental or material – that indicates that there is a material world. Nothing in the present that indicates there is a past. Skeptics, of course, have known this for a long time. Indicators carry information about cookies in jars, wine in bottles, and zebras in pens, not information about the heavyweight implications of these facts. That is why fuel gauges indicate, and thereby tell you, that you have gas in your tank without telling you that there is a material world (something implied by there being gas and gas tanks). That is why rings in the stump can indicate the age of a tree without indicating that the past is real. That is why birth certificates can provide information about when someone was born without providing information that the

world (complete with birth certificates) was not created this morning. If there is an answer to skepticism about the past or about the material world, the answer will not be found by examining birth certificates or looking at tree stumps. If there is a past, if that heavyweight implication (of a tree being 200 years old or a person being 84 years old) is true, tree rings and birth certificates will not tell us it is true. That is something we take for granted *in*, not something we find out *by*, counting tree rings and reading birth certificates.

Insofar, then, as any accredited way of knowing P depends on information delivered about P, I conclude that *none* of our accredited ways of knowing about our material world are capable of telling us that there is a material world, none of the accredited ways of finding out about what people feel and think are ways of finding out that they are not mindless zombies, and none of the accepted ways of finding out what, specifically, happened yesterday are ways of finding out that there was a yesterday. If this is true, it leaves us with a choice. Either

- (1) *Skepticism*: we retain closure and accept skepticism. I cannot know P, for any P, unless I know all the (known) implications of P. But I can't know all these. There is nothing I am aware of that indicates (carries information) that the heavyweight implications are true. So (given closure) we don't know any of the things – and this includes most things – that imply these heavyweight truths.

or

- (2) *Denial of closure*: we preserve ordinary knowledge by denying closure. We don't have to know the heavyweight implications of P to know P. I don't have to know I am not a deluded brain in a vat to see (hence, know) that there are cookies in the jar. I can, and often do, get information about P sufficient unto knowing that P is true without getting information about the heavyweight implications of P.

For anyone who thinks they know something about the material world, the choice is easy. Closure must be denied. There are some (known) implications of what we know – the heavyweight implications – that we do not have to know to be true in order to know to be true what implies them.

There is, though, another possibility. The choice between (1) and (2) is so painful that some philosophers prefer to try wriggling between the horns of this dilemma. Perhaps we can keep closure *and* have knowledge if we are willing to extend – by an act of courtesy, as it were – knowledge to heavyweight implications. Though I typically have no evidence that I am not being systematically deceived by the facts on which I base my belief that P,<sup>7</sup> it might still be argued that our knowledge of heavyweight implications is special. Even though none of our accredited ways of knowing transmits evidence to them, we nonetheless get to know them anyway. These are things we get to know without having reasons to believe them, without being able to eliminate the possibilities (e.g. that I *am* being deceived) they directly compete with. Though I can't know they are zebras (not elephants or giraffes) in the pen without being able to distinguish (at this distance, in these circumstances) zebras from other sorts of animals, I can (according to the present hypothesis) know that they are not painted mules even though I can't (at this distance, in these circumstances) distinguish zebras from painted mules.

This strikes me as, at best, a bit of verbal hocus pocus designed to avoid skepticism while retaining closure. We have our cake and eat it too by simply stipulating that we know whatever we know to be implied by what we know even though we have no identifiable way of knowing it. In rejecting this option I assume, of course, that the visual appearance of a zebra – the sort of perceptual condition that normally leads one to say one can see that it is a zebra – is not evidence, is not itself a reason to think, that the zoo authorities did not replace the zebras with mules painted to look like zebras. I assume that the appearance of cookies in a jar, the sort of condition that prompts one to say one can see there are cookies in the jar, is not itself a reason to believe that the experience of the cookies is not misleading or delusory in some way.<sup>8</sup> Given these assumptions, it seems to me preposterous – abominable, if you will – to insist that we nonetheless know we are not being deceived in these ways because we know that *not* being deceived in these ways is implied by what our experience normally leads us to believe – that there are zebras in the pen and cookies in the jar. Despite its philosophical credentials (Moore, 1959), that sounds like chutzpah, not philosophy, to me.

## Notes

- 1 I hereafter drop the distracting reminder that S, if he believes Q, must base his belief in Q on the two facts he knows to be true: (1) P, and (2) If P then Q.
- 2 The parenthetical “to be so” is important. Skepticism is irrelevant to what *objects* I see, whether, for instance, I see cookies in the jar, wine in the bottle, and zebras at the zoo. What I don’t see if skepticism is true are *the facts* – the fact that there are cookies in the jar, wine in the bottle, and a zebra in the pen. One doesn’t have to see (hence, know) these facts to see the objects that figure in these facts. Perception of objects (events, states of affairs, conditions) is immune from skeptical challenge.
- 3 Stine (1976), Cohen (1988, 1999), and Lewis (1996), while adopting a relevant alternative account of knowledge, retain closure by embracing a fairly radical form of contextualism. DeRose (1995) gives an externalist form of contextualism and argues that it retains closure while avoiding skepticism. For a general survey and discussion of options see Brueckner (1985).
- 4 Except for Martin (1975), which forced me to adopt an information-theoretic formulation of “conclusive reasons.” See, also, Martin (1983), where the same example is used against Nozick (1981). An oft mentioned criticism is that the theory is subject to the sorts of counterexamples (unpublished) Saul Kripke gave of Robert Nozick’s (1981) theory. I do not think these examples work against my own theory even though there is a superficial similarity between Nozick’s and my accounts (we both use counterfactuals to express the required relations between knower and known). Since I am not relying on this theory to defend my rejection of closure, a detailed defense of the theory is out of place here. For the cognoscenti, though, I say this much. When there are, in the relevant neighborhood, 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), it turns out that on Nozick’s theory you can track (Nozick’s term) red barns (you wouldn’t believe it was a red barn unless it was a red barn) without tracking barns. Thus, you can know of a red barn you see that it is a red barn but not that it is a barn. This result is an embarrassment even for someone (like Nozick) who denies closure. The example, however, is not effective against a “conclusive reason” (or information-theoretic) style analysis since these theories are formulated in terms not of a *belief* tracking a condition, but of one’s reasons or evidence (the condition causing you to

believe) tracking the condition. 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 S has conclusive reasons to believe it is a barn: it would not look that way (like a red barn) unless it was a barn, and its looking like a red barn is what is causing S to believe it is 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 causing S to believe it is a barn (its color being irrelevant to the causing of the belief) then S does not know it is a barn. His experience carries the information that it is a barn (since it looks like a red barn to him and its looking like a red barn carries the information that it is a barn), but that isn't the aspect of experience that is causing him to believe. Knowledge is *information-caused belief* and in this second case the information (that it is a barn) isn't causing the belief. All this should be evident to anyone who has thought about an example Alvin Goldman (1975) introduced years ago. Even if I mistakenly take wolves to be dogs so that my belief that x is a dog does not "track" dogs in my environment I can nonetheless know of a dachshund, seen at close range in broad daylight, that it is a dog. What is crucial to knowing the dachshund is a dog is that it has a distinctive look (it is, in this respect, like a red barn), a look that only (dachshund) dogs have. If it is this distinctive look that causes me to believe it is a dog, I know it is a dog no matter how confused I am about wolves – no matter how much my beliefs about dogs fail to "track" dogs (no matter how many false dog beliefs I have).

- 5 Except, possibly, logically conclusive evidence. I ignore this possibility since no one, I assume, thinks we have logically conclusive reasons for beliefs about the external world, the past, and other minds – the sorts of beliefs to which skepticism is typically directed.
- 6 One can, of course, *think* one remembers P without knowing P. One can even do it when P is false. But this is best described as *ostensible*, not genuine, memory. If one truly remembers P, P is not only true, one knows it is true.
- 7 I think it is begging the question to object by saying that deception is most improbable. Yes, it is most improbable *given what we take ourselves to know about ourselves and the world we live in*, but given the hypothesis in question, what allows us to help ourselves to this knowledge (about ourselves and the world we live in) to estimate the probabilities? Even if we helped ourselves to this (disputed) knowledge, we would still be left with the question: is the estimated improbability enough to say we know the hypothesis is false? I know the probabilities of my winning the lottery are vanishingly small, but that doesn't mean I know I'm going to lose.
- 8 Some philosophers, I know, would not grant me these assumptions. Klein (1981, 1995), for instance, argues that despite having nothing to show in the way of specific evidence for thinking the zoo authorities have not put a disguised mule in the zebra pen, one can nonetheless know they didn't do it because one can see that it is a zebra and one knows that if it is a zebra the zoo authorities didn't do that. Knowing that not-Q is implied by what one knows is good enough to know not-Q even if one cannot distinguish not-Q (it isn't a disguised mule) from Q (it is a disguised mule) and, therefore, has nothing to show for one's belief that not-Q. In the end, it may come down to a matter of taste (I hope not), but this view is as hard for me to swallow as others tell me the rejection of closure is for them.

## References

- BonJour, L. (1987) Nozick, externalism, and skepticism. In *The Possibility of Knowledge*, ed. L. Luper-Foy. Lanham, MD: Rowman & Littlefield.
- Brueckner, A. (1985) Skepticism and epistemic closure. *Philosophical Topics* 13: 89–117.
- Cohen, S. (1988) How to be a fallibilist. *Philosophical Perspectives* 2: 91–123.

- Cohen, S. (1999) Contextualism, skepticism, and the structure of reasons. In *Philosophical Perspectives*, ed. J. Tomberlin. Atascadero, CA: Ridgeview.
- Davies, M. (1998) Externalism, architecturalism, and epistemic warrant. In *Knowing Our Own Minds*, ed. C. Wright, B.C. Smith, and C. Macdonald. Oxford: Oxford University Press.
- Davies, M. (2000) Externalism and armchair knowledge. In *New Essays on the A Priori*, ed. P. Boghossian and C. Peacocke. Oxford: Oxford University Press.
- DeRose, K. (1995) Solving the skeptical problem. *Philosophical Review* 104 (1): 1–52.
- Dretske, F. (1969) *Seeing and Knowing*. Chicago: University of Chicago Press.
- Dretske, F. (1970) Epistemic operators. *Journal of Philosophy* 67: 1007–1023.
- Dretske, F. (1971) Conclusive reasons. *Australasian Journal of Philosophy* 49: 1–22.
- Dretske, F. (1981) *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press/A Bradford Book.
- Feldman, R. (1999) Contextualism and skepticism. In *Philosophical Perspectives*, ed. J. Tomberlin. Atascadero, CA: Ridgeview.
- Fumerton, R. (1987) Nozick's epistemology. In S. Luper-Foy (ed.), *The Possibility of Knowledge*. Lanham, MD: Rowman & Littlefield.
- Goldman, A. (1975) Discrimination and perceptual knowledge. *Journal of Philosophy* 73: 771–791.
- Grice, P. (1967) William James lectures at Harvard.
- Klein, P. (1981) *Certainty: A Refutation of Skepticism*. Minneapolis: University of Minnesota Press.
- Klein, P. (1995) Skepticism and closure: why the evil demon argument fails. *Philosophical Topics* 23 (1): 213–236.
- Lewis, D. (1996) Elusive knowledge. *Australian Journal of Philosophy* 74 (4): 549–567.
- Martin, R. (1975) Empirically conclusive reasons and scepticism. *Philosophical Studies* 28: 215–217.
- Martin, R. (1983) Tracking Nozick's sceptic: a better method. *Analysis* 43: 28–33.
- Moore, G.E. (1959) A proof of the external world. In *Philosophical Papers*. London: George Allen & Unwin.
- Nozick, R. (1981) *Philosophical Explanations*. Cambridge, MA: Harvard University Press.
- Stine, G.C. (1976) Skepticism, relevant alternatives, and deductive closure. *Philosophical Studies* 29: 249–261.
- Tomberlin, J., ed. (1999) *Philosophical Perspectives*. Atascadero, CA: Ridgeview.
- Wright, C. (2003) Some reflections on the acquisition of warrant by inference. In *New Essays on Semantic Externalism, Skepticism, and Self Knowledge*, ed. S. Nuccetelli. Cambridge, MA: MIT Press/A Bradford Book.

# The Case for Closure

John Hawthorne

## Introduction

Fred Dretske's discussion provides us with a case against epistemic closure that is, on its face, rather compelling. In brief, he reasons as follows. (1) There are various "heavyweight" propositions, such as that I am not a brain in a vat, that I cannot come



to know through use of my perceptual capacities. (2) Nor can I come to know them a priori. So (3) I can't know them at all. But (4) rampant skepticism is unacceptable. I can come to know all sorts of ordinary propositions, such as that I have a hand, that there are cookies in my cupboard, that I used to frequent the Harborne fish and chip shop, and so on. But (5) those ordinary propositions are known by me to entail various heavyweight propositions of the sort that (by (1) and (2)) can't be known by me. So (6) it is false that knowledge transmits over known entailments.

Those who deny skepticism but accept closure – of whom there are many – will have to explain how we know various “heavyweight” skeptical hypotheses to be false. As Dretske sees it, this leads us to indulge in “verbal hocus pocus,” twisting and contorting the concept of knowledge to fit the twin desiderata of closure and anti-skepticism. Others reject (4), opting for skepticism. But Dretske, reasonably enough, expects most of us – even closure-lovers – to agree that acceptance of wholesale skepticism is even more “abominable” than a denial of closure. It would appear, then, that we have three choices: (a) abominable skepticism; (b) verbal hocus pocus; (c) the abandonment of the initially plausible sounding closure principle. Dretske hopes that we will come away feeling that the sober option is to abandon closure. Despite its self-confessed costs, his position is a seductive one.

In the remarks that follow, I shall not attempt definitive resolution of the very troubling issues that Dretske's admirable discussion raises. If there were some easily accessible locus of reflective equilibrium in the vicinity, we would surely have reached it by now. But each of the resolutions to the trichotomy carries undeniable costs.

My task here is not, then, to defend a particular resolution to the kind of puzzle that motivates Dretske to abandon closure. Rather it is to question the merits of his own preferred, closure-denying package. Once the force of a suitable epistemic closure principle has been properly appreciated and the untoward consequences of Dretske's own views have been properly exposed, the costs of Dretske's preferred package can be seen to be quite considerable.

In the discussion below, I shall proceed as follows. Having in section 1 said a little about the closure principle I am most interested in, I shall in section 2 underscore the intuitive cost of abandoning that principle. In section 3, I shall point to some difficulties attending to the project of devising a restricted version of closure. In section 4, I shall make some brief remarks about the kind of puzzle that motivates Dretske to abandon closure.

## 1 Versions of Epistemic Closure

Epistemic closure is sometimes formulated as the principle (or, more accurately, schema) that

- (1)  $(S \text{ knows } P \text{ and Necessarily } (P \supset Q)) \supset S \text{ knows } Q$

But this simple formulation does not present a defensible principle: if the relevant conditional is unknown to the subject, we should hardly be surprised if he were to know its consequent though not its antecedent.<sup>1</sup>

Recognizing this, others formulate epistemic closure as the principle that

- (2)  $(S \text{ knows } P \text{ and } S \text{ knows that Necessarily } (P \supset Q)) \supset S \text{ knows } Q.$ <sup>2</sup>

But that doesn't seem right either. If one knows  $P$  and the relevant conditional, one still might not know  $Q$  on account of not having performed the relevant deductive inference. Knowing the logical consequences of what one knows takes cognitive labor.

Dretske is well aware of all this, of course, and formulates a more guarded version of the closure principle:

- (3) If one knows  $P$  and one knows that Necessarily  $(P \supset Q)$ , one has all that it takes, evidentially speaking, to know that  $Q$ .<sup>3</sup>

Here is the picture. Suppose one has a body of evidence such that, were one to base a belief on that evidence, one would thereby come to know  $P$ . Even if one doesn't so form a belief, it is nonetheless natural to say that one has all it takes, evidentially speaking, to know  $P$ .

But version (3) of the closure principle invites at least two sorts of objections other than those that concern Dretske in his discussion.<sup>4</sup>

The first flows from the thought that one cannot always know the logical consequences of the propositions that one knows on account of the fact that small risks add up to big risks. Such a thought flows naturally from the view that knowledge that  $P$  requires that the epistemic probability of  $P$  be above a certain threshold  $N$  (where  $N$  is less than 1). I myself am not especially attracted to this "threshold conception" of knowledge. But if one were to embrace such a conception, then (3) would only be acceptable under the somewhat unrealistic idealization that one knows all necessary truths with certainty. And this is unrealistic. Suppose someone who is quite reliable tells me that the atomic number of calcium is 20. Suppose I know for certain that that if calcium has atomic number 20 then it necessarily has that number. If my rational credence that he is telling the truth is less than one, then the same is intuitively so for the proposition

Necessarily  $(\text{Calcium exists} \supset \text{Something with atomic number 20 exists})$

So if one embraces the threshold conception of knowledge, (3) is unacceptable. This problem can be sidestepped by moving to (3')

- (3') If one knows  $P$  and knows with a priori certainty that Necessarily  $P \supset Q$ , one has all that it takes, evidentially speaking, to know  $Q$ .

A second sort of worry, which confronts both (3) and (3'), turns on some delicate issues about the relationship between logical consequence and rationality. Consider the following case, which builds on Saul Kripke's famous Padarevski example (see Kripke, 1979). Suppose I overhear someone (who is utterly trustworthy) say "Fred Padarevski is Bob Padarevski's brother" and I accept that sentence. I overhear someone else (who is utterly trustworthy) say "Ivan Padarevski is Fred Padarevski's father" and also accept that sentence. However, I do not realize that the token "Fred Padarevski" as used by the second individual corefers with the token of "Fred Padarevski" used by the first. Having accepted the second sentence, I apply the necessity of origins principle and go on to

accept “Necessarily if Fred Paderevski is Bob Paderevski’s brother, Ivan Paderevski is Bob Paderevski’s father.” Now many of us hold a conception of belief attribution according to which an onlooker could truly say “John knows that Bob Paderevski is Fred Paderevski’s brother and John knows that necessarily, if Bob Paderevski is Fred Paderevski’s brother, then Ivan Paderevski is Bob Paderevski’s father.”<sup>5</sup> But it is quite clear that I am in no position to know that Ivan Paderevski is Bob Paderevski’s father. (I am agnostic, after all, as to whether the name “Fred Paderevski” as used by the first interlocutor is coreferential with the name “Fred Paderevski” as used by the second interlocutor.) And without getting fully clear on what “evidentially speaking” is supposed to amount to, it seems at least dubious whether I “have what it takes, evidentially speaking, to know that Ivan Paderevski is Bob Paderevski’s father.” There is plenty more to say here, but because the issues that primarily concern Dretske do not turn on these matters, I will leave things somewhat hanging. Suffice it to say there is a good case to be made that, even if one is ideally rational, one will often be in no position to recognize the logical consequences of one’s knowledge on account of diverse modes of linguistic representation of the propositions known.

If the closure debate were over (3) or (3’), then I would be the wrong recruit for defending closure. Fortunately, the issues just raised about Dretske’s preferred formulation are somewhat tangential to the core of the debate. Here is a version that avoids these issues:

- (4) If one knows P and competently deduces Q from P, thereby coming to believe Q, while retaining one’s knowledge that P, one comes to know that Q.<sup>6</sup>

(Why “while retaining one’s knowledge that P”? Well, one might competently deduce Q but have gotten counterevidence to P by the time one comes to believe Q.)

The “Paderevski case” does not pose problems for (4), since in that case a competent deduction to the consequent of the conditional is, intuitively, not available. And (4) offers in any case a more satisfying development of the closure intuition. The core idea behind closure is that we can add to what we know by performing deductions on what we already know. The idea that knowing both P and Necessarily  $P \supset Q$  is evidentially sufficient for knowing Q is, it seems to me, parasitic on the idea that a priori deduction from P yields knowledge of the propositions so deduced. My concern, then, will be to defend (4), rather than Dretske’s own stated target, (3). To do so is not to frame the issue unfairly. The concerns that Dretske raises about closure carry over perfectly well to (4), and most of what I shall say in response carries over perfectly well to Dretske’s criticisms of (3).

It is worth noting in passing that there is a more ambitious version of closure than (4), one that generalizes to deductions with more than one premise, namely:

- (5) If one knows some premises and competently deduces Q from those premises, thereby coming to believe Q, while retaining one’s knowledge of those premises throughout, one comes to know that Q.

Like (4), (5) has great *prima facie* intuitive plausibility. Indeed, (4) and (5) seem to originate from the core idea that one can add to what one knows by deductive inference from what one already knows. But defending (5) requires additional work. In particular,

one has to confront the concern already mentioned, that “small risks add up to big risks,” a concern that has no force against (4).<sup>7</sup> One has to also confront intuitive counterevidence arising from consideration of particular cases: allowing that a businessman knows that he will meet X on Monday, Y on Tuesday and so forth, it still seems outrageous to allow that he knows some long conjunction of those claims, since the possibility that one of the people on the list pulls out sick seems not to be one that can be disregarded. So defending (5) requires, on the one hand, a conception of epistemic risk according to which knowledge involves no risk<sup>8</sup> and, on the other, some satisfactory resolution of the counterintuitions generated by a consideration of cases. I will not undertake any such defense here.

## 2 In Favor of Closure Principle (4)

Dretske and I would *both* agree that (4) is intuitive. It is unquestionably odd to concede that someone knows some P, to recognize that the person went on to realize that some Q follows from P and to believe Q on that basis, but to deny that the person knows Q.<sup>9</sup>

Now as Dretske points out, there are claims that are “conversational abominations” but that are perfectly true nevertheless. The contrast between truth and conversational propriety is, indeed, a very important one. For example, it is odd to actually assert “I know I have a hand” in nearly all ordinary conversational contexts. But we should not let such facts about conversational impropriety induce us into questioning the truth of “I know I have a hand.” Might a parallel diagnosis be applied here? It seems not. In the case of “I know I have a hand,” one can produce a convincing account in terms of the conversational mechanisms that Grice described concerning why the relevant speech is true though infelicitous. Roughly, we can presume participants in the conversation already know that I know that I have a hand and that, in keeping with mutually accepted conversational maxims, they will reckon that I am trying to be informative in my utterances. But “I know I have a hand” would only be informative if there were some real question as to its truth. So if I utter “I know I have a hand” then I can only be reckoned a cooperative conversant by my interlocutors on the assumption that there was a real question as to whether I have a hand. If there is no real question, my utterance misleadingly suggests there is such a question – hence the infelicity in most contexts. But what sort of mechanism could explain why “He knew P and deduced Q from P but didn’t know Q” is true but conversationally infelicitous? No analogous Gricean narrative seems to be available here.

Moreover, the relevant infelicity intuitions arise not merely at the level of conversation but at the level of thought. If I think someone knows P and deduced Q from P, then I will not merely reckon it inappropriate to say “He still doesn’t know Q”: I will actually form the belief that he does know Q. Correlatively, when I reflect on principle (4), I form the belief that it is obviously correct. If an utterance u means P, then conversational mechanisms may explain away why we don’t utter u even though it is correct, but they will not explain why we believe that which is expressed by the negation of u.

Consider by analogy the transitivity of the *taller than* relation. Suppose, somehow, that the taller than relation isn’t transitive. If God told me such a thing, I would not conclude that I had been guilty merely of conversational abominations in making utterances that presupposed its transitivity. On the contrary, I would conclude that I

had some seriously mistaken beliefs about the nature of the *taller than* relation. The closure principle seems to me like that. If Dretske is right, then I am prone to forming false beliefs – some involving conditionals, some not – about what others know by applying a closure principle that does not have full generality.

So our commitment to something like (4) cannot be explained away by appeal to pragmatic mechanisms. But let me be honest here. While principle (4) is compelling, it is not as *manifestly* obvious as the transitivity of the *taller than* relation. We would think it ridiculous if someone were to write a paper challenging the latter. And while we might say, by way of rhetoric, that Dretske is being ridiculous in denying (4) (or its counterpart (3)), the cases seem different. Perhaps, then, we should work to provide some supplementary considerations that reinforce our commitment to (4)'s intuitive status. Here are two lines of thought to that end.

### Argument (A): closure and distribution

Begin with a principle that seems extremely compelling.

*The Equivalence Principle:* If one knows a priori (with certainty) that P is equivalent to Q and knows P, and competently deduces Q from P (retaining one's knowledge that P), one knows Q.

Interestingly, Dretske's reasons for denying closure have no force against the Equivalence Principle. His argument against closure relies on the following idea. Following recent usage, let us say that R is "sensitive" to P just in case were P not the case, R would not be the case. Suppose one believes P on the basis of R, and that P entails Q. R may be sensitive to P and still not to Q. But notice that where P and Q are equivalent, there can be no such basis for claiming that while R can underwrite knowledge that P, it cannot underwrite knowledge that Q. We may thus safely assume that Dretske will accept the Equivalence principle. Here is a second, equally compelling, principle:

*Distribution:* If one knows the conjunction of P and Q, then as long as one is able to deduce P, one is in a position to know that P (and as long as one is able to deduce Q, one is in a position to know that Q).

(Often a stronger principle is defended: that if one knows P and Q, one knows P and one knows Q. I do not need that here, though I do not wish to question it here either.)

Distribution seems incredibly plausible. How could one know that P and Q but not be in a position to know that P by deduction? Dretske, however, is committed to denying Distribution. Suppose one knows some glass g is full of wine on the basis of perception (coupled, perhaps, with various background beliefs). The proposition that g is full of wine is a priori equivalent to the proposition

g is full of wine and  $\sim$ g is full of non-wine that is colored like wine.

So by equivalence one knows that conjunction. Supposing Distribution, one is in a position to know that

~g is full of non-wine that is colored like wine.

But the whole point of Dretske's position is to deny that one can know the latter in the type of situation that we have in mind.

What does this show? I do not offer the argument as a decisive refutation of Dretske's position. What we learn is that in giving up closure we are forced to give up another extremely intuitive and compelling principle, namely the Distribution principle. If we are to play the game of counting costs, it must surely be reckoned a significant cost of Dretske's view that he must eschew Distribution.<sup>10</sup>

### Argument (B): knowledge and assertion

An idea with growing popularity among epistemologists is that there is a fundamental tie between knowledge and the practice of asserting (see Unger, 1975; De Rose, 1996; Williamson, 2000). Of course, we sometimes assert that which we do not know. But it seems that when we do, we at least represent ourselves as knowing, a fact that is in turn explicable by Timothy Williamson's conjecture that it is the fundamental norm of assertion that one ought not to assert that which one does not know. Combine this thesis with Dretske's view and additional oddities arise. Suppose Q is a "heavyweight" consequence of P and S knows P and also that P entails Q. I ask S whether she agrees that P. She asserts that she does: "Yes," she says. I then ask S whether she realizes that Q follows from P. "Yes," she says. I then ask her whether she agrees that Q. "I'm not agreeing to that," she says. I ask her whether she now wishes to retract her earlier claims. "Oh no," she says, "I'm sticking by my claim that P and my claim that P entails Q. I'm just not willing to claim that Q." Our interlocutor now resembles perfectly Lewis Carroll's Tortoise, that familiar object of ridicule who was perfectly willing to accept the premises of a *modus ponens* argument but was unwilling to accept the conclusion (see Carroll, 1895). If we embrace the thesis that knowledge is the norm of assertion, then a Dretske-style position on closure will turn us into a facsimile of Carroll's Tortoise.

## 3 Heavyweight Propositions and Conclusive Reasons

Dretske does not deny that deductive inference often succeeds in adding to our stock of knowledge. But he restricts its knowledge-extending scope: deductive inference adds to our knowledge in just those cases where the deductively reached conclusion is not a "heavyweight" proposition. *Prima facie*, the picture is quite attractive. Most deductive extensions of our belief set involve no heavyweight propositions. And in those cases, Dretske allows, the conclusions can be known as long as the premises are too. Though closure fails in full generality, it holds as a *ceteris paribus* generalization. Admittedly, in those special cases where we do arrive at heavyweight propositions, deductive inference does not extend knowledge. But it is not as if we are ordinarily misled here. On the contrary, inner alarm bells go off. If, having formed the belief that she is sitting, an ordinary person is invited to infer that she is not a brain in a vat (with an accompanying explanation of what that hypothesis involves), she may well not embrace that conclusion, even though it is a deductive consequence of what she just claimed. On the

contrary, she will likely say, “Gee, I can’t know that.” It begins to look as if our actual practice conforms fairly well to what Dretske’s principles would predict.

But matters become more complicated when we explore the concept of a “heavyweight” proposition. Let us begin with a somewhat vague, but nevertheless workable gloss on the crucial notion. Let *P* be a “heavyweight proposition” just in case we all have some strong inclination to think that *P* is not the sort of thing that one can know by the exercise of reason alone and also that *P* is not the sort of thing that one can know by use of one’s perceptual faculties (even aided by reason). Rough and ready as the characterization is, it captures a familiar category. And Dretske is just right when he points out that we all have some strong inclination to think that we cannot know by perception that we are not looking at a white thing illuminated by a red light (when in fact something looks red), that we are not brains in vats, that we are not looking at a cleverly disguised mule (instead of a zebra), and so on. And he is right when he points out that we all have a strong inclination to say that we cannot know such propositions by rational reflection either. In conceding this, I am not claiming that we do not know such things. But I am conceding that I have an immediate and strong inclination to say that I do not know such things in a way that I do not, for example, have any such immediate inclination to say that I do not know that I am working on a computer right now. Of course, I might derivatively generate some inclination to say that I do not know the latter via reflection on the former. But in the case of the former, doubts are natural and immediate, whereas in the case of the latter, doubts tend to be derivative. We should all at least agree that the relevant contrast is a salient fact about human psychology and that it is one of the bases of epistemic puzzlement. Let us call those propositions that generate immediate epistemic doubt “manifestly heavyweight propositions.”

In response to the phenomenon of manifestly heavyweight propositions, one familiar reaction is to combat the inclinations towards doubt. Another is to allow doubt to spread to more humdrum propositions so that skepticism prevails. A third less familiar response is Dretske’s – to reject closure and to offer a theoretical characterization of those propositions for which closure fails.

But the problem with this sort of approach is that the set of knowledge-eluding propositions identified by the theoretical characterization turns out to be markedly different from the set of manifestly heavyweight propositions that motivated the approach in the first place. Consider, by way of example, Robert Nozick’s approach, in its simplest version. Nozick is initially taken with the idea that a necessary condition for knowing *P* is the following:

If *P* were not the case, *S* wouldn’t believe *P*.

But such a principle will rule out knowledge of many non-heavyweight propositions as well. Suppose a real dog and a fake cat are in a room, the former keeping the latter from view. I look at the dog and form the belief that there is a dog in the room. From this, I infer there is an animal in the room. Suppose further that if there hadn’t been a dog in the room, I would have seen the fake cat and formed the belief that it was a (real) animal. Then my belief that there is a dog in the room passes the Nozick test but my belief that there is an animal in the room does not. But my belief that there is an animal in the room in no way involves a heavyweight proposition. Examples such as

this one put considerable pressure on closure-challengers like Nozick to “fix” their accounts of the necessary conditions on knowledge.<sup>11</sup> The aim of the game is to provide an account that precludes inference from what is known to manifestly heavyweight propositions but that allows inference from what is known to those that are not. If I point out that by the theorist’s lights I can know that I have hands but not that I am not a brain in a vat, the theorist will welcome the result. But if I point out that by the theorist’s lights there are situations where I can know that there is a cat in front of me but not that there is an animal in front of me, the theorist will hardly celebrate this result. On the contrary, he will scurry back to the epistemological laboratory to contrive an account that delivers the welcome result while avoiding the embarrassing one. Time and again, closure-deniers fail to pull this off. I believe that Dretske’s own preferred account of knowledge offers an instance of this pattern of failure.

To begin, let us get clear about the form of his account. Dretske begins with the idea that to know that *P* one has to “rule out” or “exclude” all relevant alternatives to *P*. He claims in passing that this approach to the analysis of knowledge leads “quite naturally (not inevitably, but naturally)” to a denial of closure. I find this claim hard to evaluate without a clear sense of what “ruling out” or “excluding” amounts to. One very natural gloss on “ruling out *P*” is “knowing not-*P*.” On this gloss, the suggestion is that for each *P* that one knows at *t*, there are some propositions incompatible with *P* the falsehood of which one has to know at *t* as well. Whether or not this is true – though I doubt that it is – I doubt that this is the intended meaning. Perhaps “rule out” means “to be able to know not-*P*” or “have what it takes, evidentially speaking, to know not-*P*.” Assuming suitable deductive prowess, the suggestion that knowing *P* entails one can rule out relevant alternatives to *P* hardly encodes a challenge to closure. Indeed, assuming closure, one would, on this construal, be encouraged to think that if one knows *P* one can know the falsehood of any old alternative to *P*, relevant or not.

Fortunately, Dretske offers us a positive constraint on knowledge that bypasses this issue and that does indeed yield the result that closure fails.<sup>12</sup> The constraint is based on the idea of “conclusive reasons.” Suppose I believe *P* on the basis of reasons *R*. Dretske tells us that if I know *P*, then I would not have reasons *R* unless *P*.

Let us say that reasons *R* are conclusive for *P* just in case this condition is satisfied. Suppose it is a necessary condition for knowledge that one’s reasons are conclusive. Then, assuming we are not skeptics, and assuming a very natural gloss on one’s reasons and the relevant counterfactuals, closure does indeed fail. Suppose *A* believes there is a zebra in the cage (and that this is the only animal in the cage) and *A*’s reasons *e* are a set of zebra-ish experiences (coupled, perhaps, with some background beliefs). It is natural to suppose that if there weren’t a zebra there, *A* wouldn’t have had *e* (and if there were other animals there as well, *A* wouldn’t have had *e* (either). So *A*’s reasons are conclusive. Let us assume further that *A* knows on some occasion that the only animal in the cage is a zebra, her reasons being roughly as just described.<sup>13</sup> (This is, of course, an additional assumption. The conclusive reasons principle, as we have formulated it, is only a necessary condition on knowledge.<sup>14</sup>) *A* then deduces and comes to believe that there is no non-zebra in the cage cleverly disguised as a zebra. Let us suppose, as is natural, *A*’s reasons are one or both of *e* and her belief that the only animal in the cage is a zebra. But if there were a cleverly disguised mule in the cage, she might well still have believed there was a zebra in the cage and that it was



the only animal. So A does not have conclusive reasons for her belief that there is no cleverly disguised mule in the cage.<sup>15</sup> The necessary condition is not satisfied and so she does not know.<sup>16</sup>

Note further an obvious advantage of the Dretske approach over the simple version of the Nozick approach, one that Dretske is fully aware of.<sup>17</sup> Return to our “fake cat” case (see Dretske, note 4). One would not wish to concede in that case that one knows that there is a dog in the room but does not know that there is an animal in the room. But Dretske is not forced to such a concession. Suppose one’s reasons for believing that there is an animal in the room are one or both of: (a) doggish experiences; (b) the belief that there is a dog in the room. If there were no animal in the room one would not have had *those* reasons for believing there to be an animal in the room (though one might have had others). So one turns out to have conclusive reasons for believing that there is an animal in the room. No embarrassing concession is mandated.

Nevertheless, embarrassment is close at hand. It seems evident that Dretske’s account is designed to deliver the conclusion that one knows ordinary propositions and their non-manifestly heavyweight consequences, while remaining ignorant of their manifestly heavyweight consequences. But it does not come close to delivering that result. It turns out that we all too often have conclusive reasons for manifestly heavyweight propositions (in which case Dretske has not in fact provided an effective barrier to knowing such propositions), and all too often lack conclusive reasons for a priori consequences of known propositions, even though those consequences are not manifestly heavyweight. In these cases, not surprisingly, it is embarrassing to stick to one’s theoretic guns and deny that the consequences are known.

### Category 1: manifestly heavyweight with conclusive reasons

Case 1. As the discussion above in effect revealed, there are many conjunctions that are manifestly heavyweight but for which I have conclusive reasons. Thus, while I might lack conclusive reasons for the proposition  $\sim$ (I am a brain in a vat), I will (supposing I have a headache) have conclusive reasons for I have a headache and  $\sim$ (I am a brain in a vat). My reasons for that conjunction include my headache. Were the conjunction false, I would not, then, have had my reasons. (The closest worlds where that conjunction is false are ones where I do not have a headache. Stated without recourse to possible worlds language: were that conjunction false, it would be because I didn’t have a headache.)

My putative knowledge of that conjunction cannot, then, be challenged on the grounds that I lack conclusive reasons. But the proposition that I am credited with knowing is every bit as apt to raise inner alarm bells as the proposition that  $\sim$ I am a brain in a vat and will thus come out as manifestly heavyweight.

Case 2. I go to the zoo and see a bird flying around in a cage. I form the belief that I am not looking at an inanimate object cleverly disguised to look like an animate object. That is intuitively a manifestly heavyweight proposition. But suppose (as seems reasonable) that while it is just about possible to make an inanimate object appear like a flying bird, there are much easier ways to make an inanimate object look like an animate object. For example, it is much easier to pull off the task of creating an inanimate object that looks like a turtle than to pull off the task of creating an inanimate object that looks like a flying bird. Suppose then that “at the closest worlds” where there is

an inanimate object in the cage, there is nothing that looks like a bird in the cage: instead there is something that looks like a far more sedate animal. Suppose, as seems plausible, my reason for thinking that there I am not looking at a cleverly disguised inanimate object in the cage is that I have experiences as of a bird flying around. Then I have conclusive reasons, in Dretske's sense, for the proposition that there is no cleverly disguised inanimate object.

Case 3. I see a cookie. I form the belief that there is a mind-independent object that is roughly five feet in front of me. My sense is that Dretske wishes to classify such highfalutin theoretically loaded beliefs as "heavyweight." But that one does not come out as heavyweight by the conclusive reasons test. For presumably the closest worlds where there is no mind-independent physical object five feet in front of me are not ones where some bizarre metaphysics holds but, rather, worlds with laws like this one where there is no physical object at all in front of me. Alternatively put: if there weren't a mind-independent object about five feet in front of me, that would be because that region of space was unoccupied. But if that had happened, then (in the normal case) I wouldn't have had the experiences as of a cookie five feet away from me.

## **Category 2: not manifestly heavyweight but lacking conclusive reasons**

Case 1. I eat some salmon for dinner. I am no glutton. I eat a modest quantity and form the belief that I have eaten less than one pound of salmon. I infer that I have eaten less than 14 pounds of salmon. In fact my perceptual system is very reliable indeed. In those nearby possible worlds where I feast on salmon and eat, say, a pound and a half, I do not believe that I have eaten less than a pound. Dretske, no skeptic, will happily concede that I know that I have eaten less than a pound of salmon. Suppose (and for all I know this is correct) that while it is utterly unlikely that any human being would eat over 14 pounds of salmon, doing this would induce (among other things) severe hallucinations. Indeed, it might even induce the hallucination that one had eaten a rather small quantity of salmon.<sup>18</sup> Thus it might be true that if, God forbid, I had eaten over 14 pounds of salmon, I may have had the eaten-less-than-one-pound-ish visual experiences that I have at the actual world. Thus while it may be true that I would not have had my reasons R for believing that I have eaten less than a pound of salmon unless I had, it is not true that I would not have had my reasons R for believing that I have eaten less than 14 pounds of salmon unless I had. The apparatus (in conjunction with the verdict that I know, in the given case, that I have eaten less than a pound of salmon) delivers the embarrassing result that I know that I have eaten less than a pound of salmon but not that I have eaten less than 14 pounds (even though I believe the latter and have deduced it from the former).

Case 2. I think on Monday that I am going to meet you on Wednesday on the basis of familiar sorts of reasons. (Fill them out in a natural way.) I meet you on Wednesday. It is natural to say, in many such cases, that my belief on Monday was a piece of knowledge. Consider now the counterfactual: if you hadn't met me on Wednesday, I would not have had those reasons on Monday for thinking that you would meet me on Wednesday. I do not find that counterfactual very compelling. (I have a strong inclination not to "backtrack" when evaluating the counterfactual – in effect accepting David Lewis's (1973) advice to hold the past relative to time *t* fixed when

evaluating counterfactuals concerning t.) So my reasons turn out not to be conclusive, since they fail Dretske's counterfactual test.<sup>19</sup>

(There is an additional worry in the vicinity: counterfactuals are slippery and our modes of evaluating them very shifty. Dangerous, then, to build an account of knowledge on them unless one is happy for the truth of knowledge claims to slip and slide with the varying similarity metrics that may guide verdicts about counterfactuals. Let me illustrate with a case. Suppose, unknown to me, the beach is closed when the authorities believe the neighboring waters to be shark-infested. I go to the beach and ask "Is the water shark-infested?" The lifeguard says "no." Now if the lifeguard had believed the waters to be shark-infested, he wouldn't even have been on duty that day – the beach would have been closed. In that case I wouldn't have even gotten to ask him that question. Now consider the counterfactual "If the waters had been shark-infested, the lifeguard wouldn't have said that they weren't." Our evaluation fluctuates according to what we hold fixed. If we hold fixed that I am talking to a lifeguard, then, intuitively, the counterfactual comes out false. Meanwhile, if we are allowed to backtrack, imagining that I wouldn't have even reached the beach if the waters had been shark-infested, the counterfactual comes out true. Yet the parameter on which the truth of the counterfactual depends seems not to be one upon which the truth of the knowledge claim "I now know that the water is not shark-infested" similarly depends.)

The cumulative lesson is clear enough. Dretske's machinery is intended to align itself with our instinctive verdicts about what we can and cannot know by perception and reason. But, quite simply, it draws the can/cannot line in a very different place than intuition does. Of course, one might retreat to an account that gets more intuitively satisfying results by brute force, along the lines of:

If one knows P and deduces Q from P then one knows Q, unless Q is a manifestly heavyweight proposition.

But we should all agree that this would be rather unsatisfying. But it remains an open question whether any reasonably satisfying non-ad-hoc restriction on closure is available. Epistemologists far too often assume that a discrepancy in our epistemic intuitions about pairs of cases can be accounted for in terms of interesting structural differences between the cases – differences in the pattern of counterfactuals and so on – without paying due attention to the myriad ways in which our perceptions of epistemic risk are at the mercy of incidental features of the way that the case is framed, of distortions induced by particularly vivid descriptions of this or that source of error, and so on (see Kahneman, Slovic, and Tversky, 1982; Kahneman and Tversky, 2000). I myself hold out very little hope for a plausible restriction on closure that allows with the skeptic that we are in no position to know we are not brains in vats but that allows deduction to extend knowledge in the normal case.

I close this section with one final intuition pump. Child A visits the zoo and is told by his trustworthy parent that he will get a prize if the first animal he sees is a zebra. The parent adds that the child may get a prize anyway. But whatever else, the child will definitely get a prize if the cage at the zoo entrance contains a zebra. Child B visits the zoo and is told the same thing by his parent. The parents of A and B consult. Parent A is both a philosopher and very generous. He has in fact decided that he will give

child A a prize just as long as the cage at the zoo entrance does not contain a mule cleverly disguised as a zebra. (He resolves to check with the honest zookeeper before dispensing the prize.) The parent of B is less generous and less imaginative. He resolves to give child B a prize iff the first animal is either a zebra or a lion. The parents consult. Since they are walking together, they realize that if the second child wins a prize the first will also – that, indeed, it is easier for A to win a prize than for B to win a prize. They tell this to the children (though they do not explain in detail why that discrepancy obtains). The parents are very trustworthy, transmitting knowledge by testimony. Thus child A knows (a) if the cage at the zoo entrance contains a zebra, he will get a prize, (b) if the cage at the zoo entrance contains a zebra, B will get a prize, and (c) if child B gets a prize, child A will get a prize. Child B knows (a) that if the cage at the zoo entrance contains a zebra, he will get a prize, (b) that if the animal is a zebra, A will also get a prize, and also (c) that if he, B, gets a prize, child A will get a prize.

They see a zebra. Child A knows there is a zebra there. Child B knows there is a zebra there. Child A infers (a) that he will get a prize and (b) that B will get a prize. Child B infers (a) that he will get a prize and (b) that A will get a prize. Apply the Dretske machinery and we get some curious results. A does not know he will get a prize, it would seem. For if A weren't to get a prize it would be because he saw a cleverly disguised mule, in which case A would still think he was going to get a prize. But A knows B will get a prize! Furthermore, A knows that it is easier for A to win a prize than for B to – and in particular, that if B gets a prize, A will get a prize. B, meanwhile, does know that he will get a prize and knows that if he will get a prize, then so will A – since he also knows that the rules of the game are that A is guaranteed a prize if B wins. But, by parity of reasoning, B does not know that A will get a prize. At best, these conclusions are a bit dizzying.

#### 4 Residual Puzzles

Suppose we do not embrace Dretske's rejection of closure. We are still left with the puzzle that led him to the rejection in the first place. Let us focus on an example. Suppose I learn from a reputable newspaper that Manchester United beat Coventry City. That piece of information, in combination with the fact that the newspaper says that Manchester United beat Coventry City, entails that

~Due to a misprint, the newspaper said that Manchester United beat Coventry City when in fact Manchester United lost to Coventry City.

Call this proposition "Misprint." On the one hand, it seems odd to suppose that one can know Misprint in advance of reading the newspaper; on the other, it seems odd to suppose that one can come to know Misprint by looking at the newspaper.<sup>20</sup>

One *prima facie* attractive way out is that of the contextualist. Roughly, the contextualist says that the standards for knowledge ascription vary according to what the ascriber is worried about. If one worries about misprints, one's standards for knowledge ascriptions in testimony cases go up, and that is why we balk at *saying*, in the case described, that one knows Misprint upon looking at the newspaper. But the contextualist saves closure: in contexts where such worries are salient,

the ascription “He knows the Manchester United result” is also false. So we are not left ascribing knowledge to the premises of a deductive argument while denying knowledge of the conclusion.

I am by no means convinced that contextualism holds the key to Dretske’s puzzle. But Dretske’s own way with contextualism is too quick, betraying a flawed understanding of that approach. He attributes to the contextualist the thesis that reasonably experienced adults know such things as that there are cookies only as long as they do not seriously consider the heavyweight implications of such beliefs. This is wrong. Contextualism says that the standards of “know” are *ascriber* sensitive. Suppose that A isn’t considering heavyweight possibilities but that B is. In particular suppose that B believes P and also goes on to believe some heavyweight consequence of P by deduction. Suppose A says (1) “B knows P” and, moreover, (2) “B knows anything he has deductively inferred from P and thereby come to believe.” If B is in a context where “A knows P,” in her mouth, expresses a truth, then she will be in a context where (2), in her mouth, expresses a truth as well. That the *subject* of a knowledge ascription attends to heavyweight propositions does not undermine knowledge ascriptions by an ascriber who does not worry about this or that skeptical scenario. Indeed, in the case just described, the standard contextualist will allow that the extension of “knows” in the mouth of A may include the pair  $\sim B$ ,  $\sim Q$ , where Q is a heavyweight proposition. Relatedly, it is simply not correct that contextualism delivers the result that “philosophers are the most ignorant people in the world.”

Contextualism remains one *prima facie* promising avenue for resolving Dretske’s puzzle. Are there others? Without pretending to be able to resolve perennial problems in a few short paragraphs, let me at least gesture at one other move that is perfectly consistent with closure.

Begin with the humble phenomenon of epistemic anxiety, cases where human beings really begin to have genuine doubts about what they believe. Suppose, for example, I worry that my partner is going to leave me or that the plane I am on is going to crash. In such situations, I may still have the betting inclinations symptomatic of a believer. For example, despite my worries, I may be willing to bet heavily on the plane’s safe arrival. But it is an obvious feature of the way the concept of knowledge works that in such a situation one is not inclined to self-ascribe knowledge of the proposition that is the target of one’s anxiety. This might be accounted for variously. Perhaps knowledge requires a kind of conviction that is incompatible with such anxiety. Or perhaps instead, while knowledge is in fact compatible with such anxiety, the latter is constituted in part by a belief that one doesn’t know.<sup>21</sup>

Notice that such anxiety, once generated and unalleviated – either by informal therapy or by a simple shift of attention – tends to spread. If one has it about Q and P entails Q, one acquires it with respect to P (when the question arises). Someone might know that his partner will never leave him and this may entail that she will never leave him for Mr X. But if he runs through the relevant inference, that may induce anxiety about both the inferred proposition and the original (we all know how to fill in the details!)

On such a picture, there are two sorts of situations. On the one hand, there are standard cases where deductive inference adds to our stock of knowledge. On the other, there are situations where deductive inference brings to light sources of anxiety that may, in some way or other, undercut the epistemic status – or at least the believed

epistemic status – of one’s premises. An account of the human epistemic condition should give due place to both mechanisms. But it remains unclear to me whether a denial of epistemic closure would have any useful role to play in any such an account.<sup>22</sup> On the contrary, the reality of closure explains the epistemic goodness of deductive inference as well as why epistemic anxiety tends to spread.

But still, what of the puzzle itself? Do I know Misprint in advance? And then, by parity of reasoning, can I have contingent a priori knowledge that one is not a brain in a vat? Or is it that, as long as anxiety can be kept at bay, we can know the latter only by perception in combination with deductive inference (intuitions and Bayesian style arguments to the contrary notwithstanding<sup>23</sup>)? Better that we appreciate (with Dretske) the force of this problem than rush too quickly to try to solve it.<sup>24</sup>

## Notes

- 1 There are those who adopt a conception of thought according to which the thought that P is identical to the thought that Q just in case P and Q are true in the same possible worlds. From that perspective, (1) may be more motivated. After all, if P entails Q, then the thought that P is necessarily equivalent to the thought that P and Q. So if one knows P and P entails Q, one automatically knows that P and Q since the thought that P and Q just is, from this perspective, the thought that P. Assuming that knowledge distributes over conjunction – that is, that if one knows P and Q, one knows P and knows Q – then (1) is secured. But, despite its distinguished list of adherents, such a conception of belief is a deviant one, and I shall not pursue it here.
- 2 See, for example, Dretske (1971), Stine (1976), Nozick (1981), Cohen (1988), and Feldman (1999).
- 3 Note that this principle allows for the possibility of some pair P, Q such that P entailed Q but that it was impossible to know P without already knowing Q. Thus the principle at least allows that there be a knowable P and an entailed Q such that deductive inference from P to Q was not a possible route to knowing Q.
- 4 There are other, less decisive, worries as well. Suppose I know P and know that necessarily  $P \supset I$  will never perform a deductive inference again. Is it intuitive to say still that I have all it takes, evidentially speaking, to know the consequent?
- 5 For helpful discussion of the relevant issues, see Kripke (1979) and Soames (2000).
- 6 This is a (slightly improved) version of Williamson (2000, p. 117), restricted to single premise inferences. Williamson has an insightful take on the root of epistemic closure intuitions, namely the idea that “deduction is a way of extending one’s knowledge,” one that I take on board in what follows. The notion of “competent deduction” could obviously do with some elaboration, though I won’t attempt that here. Another general issue worth considering (raised in correspondence by Jonathan Kvanig) is whether, in a case where Q is competently deduced from P, a (misleading) defeater for Q – that destroys knowledge of Q – is automatically a defeater for knowledge of P.
- 7 This may be a little quick: one might worry that an exercise of one’s deductive capacities that in fact is competent carries a risk of being incompetent. I shall not address that worry here.
- 8 Such a conception is, for example, defended by Williamson (2000), who explicitly adduces it in defense of generalized closure.
- 9 I leave out the “retaining all along” clause for ease of exposition.
- 10 A similar line of thought shows that Dretske is committed to denying a seemingly humble, restricted version of closure, namely addition closure – that if one knows P then by deduction one can know P or Q. See my *Knowledge and Lotteries* (2004). Nozick (1981) is willing to abandon distribution along with closure. I await Dretske’s opinion on the matter.

- 11 In Nozick's own case (1981), the "fix" to these anticipated worries famously took the form of an appeal to methods: one knows P, only if, were P false, one wouldn't have believed P by the same method. It is not my purpose to immerse myself in the details of Nozick's account here, though many of the remarks that follow are pertinent.
- 12 The main ideas are laid out in Dretske (1971).
- 13 Of course, we epistemologists are often a little lax in our closure-illustrating examples. We may say that the proposition that I have a meeting tomorrow strictly entails that I will not die beforehand, ignoring the possibility that I reincarnate, or go through with the meeting as a ghost. We may say that the proposition that I am seeing a zebra in the cage entails that I am not seeing a cleverly disguised mule, ignoring the possibility that the cage is itself a cleverly disguised mule or else that the mule is cleverly disguised by being inserted within a zebra that is somehow kept alive. So most illustrations of deductive closure aren't really illustrations of deductive closure. Let us trust that such sloppiness is neither here nor there.
- 14 Best to avoid the pitfalls of devising a necessary and sufficient condition here.
- 15 I am assuming that claims of the form "P wouldn't be the case unless Q were the case" entail "If Q weren't the case P wouldn't be the case." Dretske certainly relies on such an entailment in generating anti-closure results from his formulation of a conclusive reasons requirement. For the record, I am open to being convinced that "unless" counterfactuals have special features that make trouble for that purported entailment, though I doubt that this would bolster Dretske's cause.
- 16 Obviously, there are those who will insist on an externalist notion of reasons according to which, say, the fact that there is a zebra may count as a reason (when I see that there is a zebra). If one opted for this notion of reasons, then I may after all have conclusive reasons for thinking that I am not seeing a cleverly disguised mule. (And, *pace* Dretske, note 5, one's reasons would be logically conclusive.) I shall not pursue the matter further here.
- 17 Note that this simulates some of the work of Nozick's own appeal to methods.
- 18 I could run the case with pints of beer, which may be more easily grasped (as long as it is agreed that the world where I drink enough beer to hallucinate that I have drunk a normal amount is a distant one!).
- 19 Note that Dretske is happy enough to use his conclusive reasons framework to explain why I do not know that I will win the lottery (see Dretske, 1971, and "The case against closure," note 7). But it may well be that for any belief about the future, there are outcomes with a real – though low – probability that, if actualized, will render the belief false. The small chance of winning the lottery means, for Dretske, that we do not know we will lose. But what of the small chance postulated by quantum mechanics of the cup flying sideways if I drop it? And what of the small chance that I will have a fatal heart attack? We are in need of an explanation as to how the cases are supposed to be different, unless skepticism about the future is to be embraced. I discuss these issues at length in *Knowledge and Lotteries* (2004) and "Chance and counterfactuals" (2005).
- 20 A natural application of Bayesian confirmation theory supports this latter contention. Supposing one's conditional probability of Misprint on the newspaper saying that Manchester United beat Coventry City is N (where N is presumably less than 1), then since one's conditional probability of Misprint on the newspaper denying Manchester United beat Coventry City is 1, one's new rational credence on the evidence that the newspaper said Manchester United beat Coventry City will drop from somewhere between N and 1 to N. (One's new rational credence will be obtained by conditionalization. One's old rational credence will be one's rational expectation that the newspaper will deny that Manchester United won plus N multiplied by one's rational credence that the newspaper will say that Manchester United won, assuming (somewhat unrealistically) that it is certain that the newspaper will deliver a verdict on the matter. If not, adjust "deny" to "not say.")

- 21 Perhaps instead, while one knows and believes one knows, this kind of situation is one where certain beliefs, albeit temporarily, no longer control one's behavior. There are no doubt fertile connections here between epistemology and the relevant issues in philosophical psychology.
- 22 Of course, if one thought knowledge required a kind of conviction over and above "belief" in the ordinary sense, then one would slightly reformulate (4) so that it only covers cases where deductive inference produces the suitable kind of conviction.
- 23 One way to resist is to insist that one's body of evidence is misdescribed by such arguments. That, I take it, is the strategy suggested by Williamson (2000). Perhaps, when I see that there is a table, my evidence is not merely that I seem to see a table, but that there is a table. I shall not explore the matter further here.
- 24 Thanks to Fred Dretske, Peter Klein, Jonathan Kvanig, David Manley, Ernest Sosa, Timothy Williamson, and especially Tamar Gendler for helpful comments and conversation. The ideas in this essay overlap in places with material in my *Knowledge and Lotteries* (2004).

## References

- Carroll, L. (1895) What the Tortoise said to Achilles. *Mind* 4: 278–280.
- Cohen, S. (1988) How to be a fallibilist. *Philosophical Perspectives* 2: 91–123.
- DeRose, K. (1996) Knowledge, assertion and lotteries. *Australasian Journal of Philosophy* 84: 568–580.
- Dretske, F. (1970) Epistemic operators. *Journal of Philosophy* 67: 1007–1023.
- Dretske, F. (1971) Conclusive reasons. *Australasian Journal of Philosophy* 49: 1–22.
- Feldman, R. (1999) Contextualism and skepticism. *Philosophical Perspectives* 13: 91–114.
- Hawthorne, J. (2004) *Knowledge and Lotteries*. Oxford: Oxford University Press.
- Hawthorne, J. (2005) Chance and counterfactuals. *Philosophy and Phenomenological Research* 70: 396–405.
- Kahneman, D., Slovic, P., and Tversky, A. (1982) *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.
- Kahneman, D. and Tversky, A. (2000) *Choices, Values and Frames*. Cambridge: Cambridge University Press.
- Kripke, S. (1979) A puzzle about belief. In *Meaning and Use*, ed. A. Margalit. Dordrecht: Reidel.
- Lewis, D. (1973) *Counterfactuals*. Cambridge, MA: Harvard University Press.
- Nozick, R. (1981) *Philosophical Explanations*. Oxford: Oxford University Press.
- Soames, S. (2002) *Beyond Rigidity*. Oxford: Oxford University Press.
- Stine, G.C. (1976) Skepticism, relevant alternatives, and deductive closure. *Philosophical Studies* 29: 249–261.
- Unger, P. (1975) *Ignorance: A Case for Skepticism*. Oxford: Oxford University Press.
- Williamson, T. (2000) *Knowledge and its Limits*. Oxford: Oxford University Press.

## Reply to Hawthorne

*Fred Dretske*

John Hawthorne does an admirable job of describing the “costs” connected with my rejection of closure. These costs are, I admit, significant. I would not be willing to pay this price if I thought there were alternatives that were less expensive. Philosophy is a



business where one learns to live with spindly brown grass in one's own yard because neighboring yards are in even worse shape. From where I sit, the grass on the other side of the fence looks worse than mine.

One option, of course, is skepticism. Though some may be willing to embrace it, John and I agree that this is a last resort. If the *only* choices are skepticism or the rejection of closure, then – too bad – we had better learn to live without closure (distribution, etc.).

As a historical footnote, I wasn't led to deny closure because it represented a way around skepticism. I was led to it because it was a result of what I took to be a plausible condition on the evidence (justification, reasons) required for knowledge. If your reasons for believing P are such that you *might* have them when P is false, then they aren't good enough to *know* that P is true. You need something more. That is why you can't know you are going to lose a lottery just because your chances of losing are 99.99 percent. Even with those odds, you still might win (someone with those odds against him *will* win). That is why you can't learn – can't come to know – that P is true if all you have to go on is the word of a person who might lie about whether or not P is so. This is just another way of saying that knowledge requires reasons or evidence (in this case, testimony) you wouldn't have if what you end up believing were false. You can learn things from people, yes, but only from people who wouldn't say it unless it were true.

John (in section 3) does a beautiful job in showing the problems with this obvious (to me) condition on knowledge. I had grappled with some of these problems myself, but others, I confess (even after 30 years of defending the view) were completely new to me. Contrary to what he suggests (case 2 in category 2, section 3), though, I think the counterfactual condition accords with common and widespread intuitions about knowledge of the future. In my more careful moments, I prefer to say (on Monday) not that I know that I will be in my office on Wednesday, but that I know that is where I intend or plan to be on Wednesday. I retreat to this more cautious claim because I realize that my reasons for thinking I will be in my office on Wednesday are reasons I would have on Monday even if a Tuesday accident were to prevent me from being there on Wednesday. As long as I might not make it to the office, I do not know I will be there. How could I? One doesn't, generally speaking, have conclusive reasons for believing that something will happen. Death may be an exception, but I doubt whether taxes are. I'm a bit of a skeptic about the future.

But John is certainly right when he says that these counterfactuals are a slippery and shifty business. So a theory of knowledge that appeals to them is exposed to examples of the embarrassing sort that he describes. This, I confess, is a problem. It is not *always* a problem. Sometimes this shiftiness in the way we understand the counterfactual reflects quite neatly – or so I think – the shiftiness and subtlety in our judgments about exactly what and exactly when someone knows something. Does S, standing at the stove watching water boil, know that the water is boiling? Can he see that it is boiling? Well, would it look that way if it weren't boiling? If it wouldn't, he can see that the water is boiling. He can know this despite not being able to see – perhaps not even knowing (but just correctly assuming) – that it is water (not gin) he sees in the pot. The relevant counterfactual is: would the water (not just anything) look that way if it weren't boiling. If *this* counterfactual is true, one can know the water is boiling without knowing (at least not by seeing) that it is water that is boiling since,

given the look of things, it could be gin (vodka, pear juice, etc.). I think this example is related in interesting ways to John's lifeguard example, but I won't take the time to spell out the details.

So sometimes the counterfactual condition captures important distinctions. But not always. As John points out, we sometimes seem to have conclusive reasons for believing "manifestly" heavyweight propositions and lack them for clearly lightweight propositions. So things don't sort out as neatly as I would like.

But let's look at the neighbors' yards. How green is their grass? John does not undertake to propose or to defend any particular solution to the problems for which the denial of closure represents an answer. Without endorsing it, however, he suggests that despite my low opinion (and "flawed understanding") of contextualism, this approach to understanding knowledge offers some promise in this regard. We can keep closure and avoid skepticism if we think of knowledge as relative to attributional context. To say that it is relative to attributional context is to say that whether S knows something – that she has two hands, for instance – depends on the context of the person who is saying S knows it. If I, a philosopher, worried about brains in vats and Cartesian demons, say it, then S doesn't know she has two hands. At least what I say when I say she doesn't know it is true. Why? Because, assuming S is in no position to know she isn't being deceived by a Cartesian demon, then, given closure, neither can she know that she has two hands. If she knew she had two hands, she would be able to infer, and thereby come to know, by closure, that she isn't being deceived by such a demon into falsely believing she has two hands. So when I say she doesn't know she has two hands, what I say is true. Neither do I know I have two hands. Nobody knows it. Skepticism is true. At least it is true when I assert it. But if S, an ordinary person on the street, someone without the least tincture of philosophy, says she knows that she has two hands, what she says is true. She attributes knowledge to herself in an ordinary, practical context in which demons and handless-brains-in-vats are not relevant possibilities. They aren't alternatives she means to be excluding, evidentially, when she attributes knowledge to herself. So what she says is true, and it is true, mind you, even when she attributes this knowledge to me, the philosopher pondering skepticism in the seminar room. Whether I know that I have two hands depends not on what alternatives I (the attributee) take to be relevant, but on what possibilities S, the attributor, takes to be relevant. So, if we imagine S (with no tincture of philosophy) asserting that skepticism is false or (given that she probably doesn't understand what skepticism is) that she and most other people know a great many things about the world, she speaks truly. Skepticism is false when S is the speaker.

So who, according to contextualism, is right? Am I, a philosopher, right when I (given my context) say that nobody knows they have hands. Or is S right? We are, I'm afraid, both right. And that is where my low opinion of contextualism comes from. The only people who can truly assert that skepticism is false are those, like S, who don't understand what skepticism is or those who understand it but who steadfastly ignore as irrelevant (for purposes of attributing knowledge to themselves and others) exactly those possibilities the skeptic insists are relevant.

I don't really consider this to be an answer to skepticism or, if it is, it is merely a backhanded way of denying closure. It looks to me as though contextualism succeeds in keeping closure (I use John's improved statement of closure, i.e. (4)) only by denying it a role in reaching conclusions about the heavyweight propositions the skeptic is

concerned with. If you actually try to use closure to conclude from the fact that you have two hands that you are not being massively deceived, then you automatically change the context in which you attribute knowledge to yourself (the context is now one in which the possibility of being deceived in this undetectable way is being explicitly raised) and, given closure (and the fact that you can't know you aren't being deceived in this undetectable way), you cease to know the premise (that you have two hands) you used to reach that heavyweight conclusion. So, despite reaching this heavyweight conclusion in a perfectly respectable way from a premise you (when you started) knew to be true, you do not wind up knowing that the conclusion is true – that you aren't being deceived. Instead, you lose the knowledge of the premise (that you have two hands) you used to reach that conclusion. This grass doesn't look very green to me.

## Chapter Three

# Is Knowledge Contextual?

Consider this triad: (1) I know that those black-and-white equine animals are zebras; (2) I don't know that those animals aren't cleverly painted mules; (3) I know that those animals are zebras only if I know that they aren't cleverly painted mules. Assuming a univocal reading of all terms throughout, the triad is inconsistent. And yet all three of its members seem plausible. Denying (1) seems objectionably skeptical; denying (2) seems naively optimistic; denying (3) strikes many as ridiculous. Something must give. Epistemic contextualists propose that not all terms have a univocal reading throughout. In particular, they propose that knowledge ascriptions (sentences of the form "S knows that P" or "S does not know that P") have contextually variable truth conditions. In normal contexts (2) is false; in skeptical contexts (1) is false; but in no context is (3) false. Stewart Cohen defends contextualism on the grounds that it provides the best solution to the puzzling triad of claims. Earl Conee rejects contextualism. Conee argues that contextualists have failed to sufficiently motivate the semantic hypothesis that their dissolution of the puzzle presupposes. Conee also argues that we don't need the contextualist hypothesis about knowledge ascriptions in order to explain the puzzle; rather, we can equally well explain it based on an invariantist semantics for knowledge ascriptions.

## Contextualism Contested

*Earl Conee*

It should be helpful to situate epistemic contextualism (EC) in a broader context. Epistemology will creep gradually into the discussion.

Most broadly, semantic contextualism is the thesis that something semantic about a symbol varies with some differences in contexts that involve tokens of the symbol. The kind of semantic contextualism of most direct interest to epistemologists asserts

that different sentence tokens of the same type have different truth conditions. That is, some differences in the context where a sentence is used affect which conditions must obtain in the world for the use of the sentence to state a truth.

For instance, the sentence “Smith’s thimble is large enough” uncontroversially displays this variation. Let’s consider a context in which Smith’s friends are discussing whether the thimble fits her. A friend says, “Smith’s thimble is large enough.” Assuming that Smith’s thimble is an ordinary one that Smith has used comfortably for years, the friend states something true by uttering this sentence token. Suppose that the same sentence is uttered simultaneously by a household mover. He has been talking about items that are sizable enough to require two movers to carry, and he mistakenly thinks that Smith has a gigantic thimble. The claim made by that token of the sentence does not state a truth.

In this example the context of the subject of the sentence, the thimble, does not vary. What does vary is the context in which the sentence is used. This is called variation in the “attributer” context, or “attributer relativity.” Variations that affect truth conditions in the context of the subject of the sentence are called “subject” contextual variations, or “subject relativity.” We shall be focusing on claims of attributer relativity.

Indexical expressions provide other illustrations of uncontroversial attributer relativity. Words like “I,” “you,” and “it” in a sentence contribute to the sentence’s having different truth conditions in different contexts. For instance, the first-person pronoun, “I,” is the subject of a sentence that is used to state a truth only if the predicate of the sentence is true of the one who is asserting something with the sentence.

The clear attributer relativity of truth conditions for sentences with a pronoun might seem at first glance to show that attributer relativity is of negligible philosophical interest. No doubt “You are happy” makes a true claim just in case the individual addressed by the attributer is happy. The philosophical questions about persons seem entirely unaffected by the innocuous contextualist claim about how the pronoun “you” refers to them. Is a person a material object, an immortal soul, or something else? Does a person exist as a fetus in early pregnancy? Are the mental properties of a person identical to physical properties? The contextualism is neutral about all of this.

On second thought, though, the general conclusion is too hasty. The contextualism seems to take a side in some philosophical controversies. It claims that sentences of the form “I am F” are true just when the one using the sentence is F. If this is correct, then apparently there has to exist some such entity as the one who is using the sentence for any of these sentences to be true. Yet this runs contrary to no-self theories. They say that there are no such beings as ones who use sentences. According to one no-self view, experiences exist. When these experiences bear certain relations to one another, such as later conscious episodes bearing a “recollective” relation to earlier experiencings, we say that the experiences have “the same subject.” That is not strictly true. There is no one entity that undergoes the experiences. There are just particular events of experiencing and relations among them.

Thus, our supposedly bland semantic claim – that sentences with first-person pronoun subjects say something true just when the predicate is true of one using the sentence – appears to conflict with this no-self view. So the semantics now seems to take a philosophical side. Perhaps the claim that epistemic terms exhibit attributer relativity will likewise turn out to exclude some existing philosophical options.

But on third thought, the assertion of first-person pronoun attributer relativity is no particular threat to the no-self view. The contextualist semantic claim employing the phrase “the one using the sentence” is among the claims that the no-self theory addresses. Our sample version of the theory declares that all such statements are false. To be plausible, the theory must also give some explanation of how such statements are often reasonable to make, in spite of their falsity. Whatever the explanation is, we should expect it to carry over to the assertion of first-person pronoun attributer relativity. The fact that this is a semantic assertion does not make it less likely that the theory can explain its plausibility.

The final merits of no-self views lie in the details of their defensibility. Generally, we see that semantic assertions are neither automatically irrelevant nor immediately decisive.

This illustration should be borne in mind as we consider EC. Many epistemological issues may be unaffected by the existence of attributer relativity in the truth conditions of sentences that include epistemic terms. By itself this is neither a strength nor a weakness in either EC or the unaffected theories. An EC thesis may have some substantial epistemic implication. The implication in conjunction with some epistemological theories may appear to induce surprising truth values. This result need not lead sensible proponents of the affected theories to give them up. They may be able to interpret the sentences to explain away any difficulties.

Another non-epistemic example of purported attributer relativity illustrates how contextualism characteristically copes with certain sorts of conflict. Suppose that Robinson is Chair of the Boring Committee (“BC”). At the time for a meeting of the BC, Robinson sees that all of the members of the BC are in the room. He declares, “Everyone is present.” The members present would be led by this to think that the membership of the BC is present. Yet “everyone,” the subject term of Robinson’s sentence, seems to be about every person in the world. Of course none of the members of the BC is led to believe that every person in the world is present.

One reasonable semantic hypothesis is that quantifiers like “everyone” have a contextually narrowed range. A use of a sentence with the subject “everyone” states a truth just when the predicate of the sentence is true of all of the people who are made relevant by the context of use. In Robinson’s context, his “everyone” would be restricted to all who are members of the BC.

Though this attributer relativity is a reasonable hypothesis, such examples can also be accommodated while allowing that “everyone” always ranges over, well, everyone. One alternative view is that sentences with “everyone” typically are overly general and therefore false. But people who hear such claims immediately ignore those who are irrelevant to the purposes at hand. This practice of ignoring is widely taken for granted. That makes such false claims efficient ways to communicate something that is about only those who are in the group of contextually determined interest.

An asset of this contextually unvarying (“invariantist”) semantic view of such sentences emerges in the following continuation of the example. Suppose that someone on the BC, say Stickler, opposes Robinson’s claim. Stickler says, “No, what you said is not true, because not everyone is here. The Pope is not here.” Probably Stickler’s comment would come off as unamusing, at best. But it is not easy to see that Stickler is saying something false.

Quantifier contextualists can find truth in Stickler’s comment. They can say that Stickler has shifted the conversational context, perhaps by mentioning the Pope,

someone who is not in the group previously assumed to be relevant. So in Stickler's mouth "everyone" ranges over a wider group, perhaps all people worldwide. Thus, in that context it is correct for Stickler to assert "not everyone is here." The contextualist can note that this is compatible with the truth of the claim made by Robinson with "everyone," on the grounds that Robinson's claim had a crucially narrower contextually determined range.

This contextualist response does not wholly avoid the difficulty, though. The response finds a truth in the claim made by Robinson and it finds a truth in a claim made by Stickler. The remaining difficulty is this. It seems that Stickler was entirely accurate (though unamusing, unhelpful, and maybe obtuse). Yet the contextualist view does not find only truth in what Sticker says. Stickler said that Robinson's original claim was untrue. The contextualist view that we are considering must disagree with this. The contextualist must say that Stickler himself is wrong on that point, because the truth conditions of Robinson's original claim were met. The contextualist view implies that the original claim was true if and only if the members of the group in question were all there, and they all were. So our impression that Robinson was entirely accurate must be erroneous.

It is not obvious how best to interpret quantifiers in cases like this. The contextualist position may be overall best. In any event, contextualism applied to quantifiers displays a characteristic feature. Contextualism finds truth at the expense of contradiction. That has a nice constructive ring to it. But it runs a risk of interpretive failure. Concerning some cases in which one sentence seems to be used to deny something that another sentence says, contextualists characteristically allege that they are not really flat denials. This happens whenever there is a difference in what the contextualism counts as a semantically relevant context. Any such contextual difference implies that the two claims do not just contradict one another, even if one sentence used seems to negate the other.<sup>1</sup> Most commonly, contextualist proposals yield compatible truth conditions for the sentences. For instance, they say that Robinson and Stickler can both be right about those whom they respectively call "everyone."

Moral relativists are the contextualists of ethical theory. They hold roughly that something about the context of a moral evaluation, perhaps something in the culture or the personal values of the evaluator, contributes to the truth conditions of the resulting evaluation. Moral relativists have always been a relatively small and besieged group, though not lacking in able philosophers. The reason seems to be this. The prevalence of fundamental disagreements about moral evaluations is a virtually ineluctable element of the ethical situation. Among non-philosophers there are vigorous disputes about capital punishment, the treatment of animals, gun control, and so forth. Or, at least, their exchanges give every appearance of being disputes, and the participants clearly regard them as such. Among philosophers there are vigorous disputes among consequentialists, Kantians, virtue theorists, and divine command theorists. Or, at least, they certainly seem to disagree. In both cases the disputants often speak from more or less different cultural and personal backgrounds. When these differences are ones that a moral relativist claims to be crucial, the relativist cannot find contradictory claims, and may find only compatible claims. Yet it is deeply dubious that any such conciliatory interpretation is correct. Moral relativism has some resources to try to explain away the appearance of disagreement. But often it seems clear both that the disputants are fully competent with the language they are using and that those

on each side fully intend to use it to deny what the other side affirms. These facts are hard for a relativist to explain away. These cases are trouble for moral relativism.

To the extent that apparent disagreement over epistemic claims is similarly witting and heartfelt, it is similarly difficult to find credible a reconciling EC interpretation. Some arguments about skepticism appear to have that character. Like moral relativists, EC theorists can work to explain away the appearance of contradiction or disagreement. But also as in the moral case, often it seems clear that competent speakers use words with the intent to be discussing the same thing and to be respectively affirming and denying something about it. These cases are trouble for EC.

Finally by way of preliminaries, we should be alert to a misleading temptation. It is tempting to think that semantic contextualism is obviously correct for all evaluative expressions. It is a plain fact that people routinely apply differing standards of evaluation in different contexts. For instance, suppose that Jones works on a Sabbath. Smith counts Jones as “evil” because Jones is thereby disobeying a strict edict of Smith’s religious faith. Robinson denies that Jones is “evil” because Jones’s work is intended to promote a social cause that Robinson admires. Numerous similar examples exist for evaluations of beauty, humor, morality, and many other things. In particular, the application of differing standards of judgment plainly occurs as well in epistemic evaluations, such as attributions of knowledge, justification, and good reasoning. Are these contextually varying truth conditions in action?

Not necessarily. The fact that different standards are routinely applied in making an evaluative judgment does not imply the correctness of semantic contextualism about the contents of the judgment. To see this, note that differing standards are routinely applied in making judgments that are uncontroversially context invariant. For instance, usually the genuineness of a purported piece of US currency is judged by a quick look, sometimes it is judged by a more careful inspection, and occasionally it is rigorously judged using a high-tech device. In spite of these differing standards, the content of the uses of “genuine US currency” is clearly the same. Other examples abound.

Thus, we cannot validly infer that contextual differences yield differing truth conditions simply because differing standards are often applied. For semantic contextualism to be correct, there must be something else about an expression, something beyond its susceptibility to differing standards of judgment, that makes the truth conditions of tokens of the expression differ with context.

## **Epistemic Contextualism**

The most widely discussed form of EC holds that the truth conditions for tokens of sentences that include “knows” (and cognate expressions) vary with the attributer’s context. Broadly speaking, what varies with some differences in the context in which a “knows” sentence is used is the strength of the epistemic position that the subject of the sentence must be in, in order for the sentence to assert a truth.

Many versions of EC share this core. Typically, the strength of epistemic position required is said to vary in a range that allows, at its low end, many true attributions in everyday contexts concerning ordinary judgments based on perception, memory, testimony, and perhaps also inductive generalization and high probability. At the high



end of the range of variation, the typical EC truth conditions are demanding enough to make true many skeptical denials of “knowledge” of the external world.

This range corresponds to a genuine phenomenon of “knowledge” attributions. Fluent English speakers regularly make confident “knowledge” attributions about numerous routine matters of daily life: the color of a shirt when it is seen in good light, the occurrence of a death when it is read about in a newspaper, the identity of the first US President when it is recalled from some source or other, the sum of two large numbers when they are totaled on a calculator, and so on. Fluent speakers typically become increasingly hesitant about “knowledge” attributions as the practical significance of the right answer increases. Fluent speakers typically doubt or deny “knowledge” attributions when coming to grips with impressive skeptical arguments.

The heart of the contextualist explanation of this phenomenon is that people are responding in an approximately accurate way to shifting standards for correct “knowledge” attributions. In the context of everyday life, “knowledge” attributions have truth conditions that are weak enough to be met by seeing a color in good light, reading a report in a reliable publication, recalling a famous historical fact, adding on a calculator, and so forth. As the truth value of a proposition matters more, the context shifts so that the truth conditions for “knowledge” of the proposition require somehow excluding ever more potential sources of error. In a context where skeptical considerations are prominent, the skepticism somehow gives rise to truth conditions for “knowledge” attributions that are so demanding that external world beliefs do not meet them.

There are two noteworthy ways in which the contextualist explanation of this phenomenon involving “knowledge” attribution could be misleading.

## Loose Talk

The first way that EC could be misleading is by being fundamentally mistaken. It may be that all “knowledge” attributions have the same truth conditions, but people apply contextually varying standards for making the attributions. The most plausible unvarying standard for truth is very high, but not unreachably high. In ordinary contexts, when nothing much turns on it, people will claim knowledge, and attribute knowledge to themselves and others, in belief and in speech, on a basis that is significantly weaker than what is actually required. This is an efficient way to communicate assurance about the proposition and to facilitate taking it for granted.

According to the view now under consideration, fluent speakers actually realize, at least tacitly, that this knowledge ascription is just loose talk. The realization is shown by the fact that, if asked whether some proposition to which knowledge is ascribed on some such basis is *really* known, or *truly* known, or *really and truly* known, fluent speakers have a strong inclination to doubt or deny that it is.<sup>2</sup> Only the most conspicuous facts of current perception, the clearest memories, triple-checked calculations, and the like will often pass some such “really and truly” test. The present view has it that the answer to this “really and truly” question reveals what a speaker judges to be knowledge when she is trying her best to apply her best thought as to what is the actual standard for the truth of a “knowledge” attribution.

A contextualist has a ready explanation for the results of this “really and truly” test. The explanation is that raising a question about “knowledge” using these words serves

to impose more stringent truth conditions. So it is no wonder that propositions that pass previous, more lenient standards are no longer said to be “known.”

There is an important liability of this contextualist response. There are many big topics that receive similar results on the “really and truly” test. For instance, when fluent speakers consider whether they are “really and truly” happy, or they consider who are “really and truly” their friends, or they consider what is “really and truly” worth striving for, they tend to have doubts about their previous casual attributions of happiness, friendship, and worthwhile goals. It is possible that the English terms for these categories of happiness, friendship, and so forth all have contextually varying truth conditions, and that the standards are raised by these “really and truly” questions. But that is not what seems to be happening as we ponder these questions. It seems that our aim is to set aside loose talk about an important topic and tell the truth about it, the very same topic that was previously more casually discussed. The contextualist who holds that the “really and truly” test changes the truth conditions is bound to count this view of what we are doing as a mistake.<sup>3</sup> That is a liability.

## **Strict Truth**

The second way in which contextualism about “knows” may be misleading would be by being correct. The reason why this might mislead is that it might suggest that the correctness of EC has a significant implication concerning some philosophical issue about knowledge. The fact is that the truth of EC in general would have no such implication.

For instance, suppose that “really and truly” does constitute a context in which “know” has truth conditions with different and higher standards than those of ordinary uses of “know.” It may be that all philosophical discussion of knowledge is carried on in that one context. That is, it may be that every issue about knowledge, from the Gettier problem to the extent of scientific knowledge, from the nature of justification to the merits of external world skepticism, has been discussed solely in this single “really and truly” context. Granting epistemic contextualism, this claim of a single philosophical context has some plausibility. Philosophy generally gets going only when we get serious about a topic. So it may be that “really and truly” is one colloquial way to direct attention to the truth conditions that arise when we take this philosophical sort of attitude toward our investigation.

If all of this is so, then “knows” is context sensitive. But appreciating that fact would tell us nothing new about our philosophical issues. All sides to the philosophical disputes would have been discussing the same thing all along. At most, we would have reason to avoid importing into philosophical discussions of knowledge examples in which “knows” is less scrupulously applied. Given the correctness of EC, these may well not be cases of the philosophical topic of knowledge. Not relying on such cases is sensible. But we have reason not to rely on them in any event. Whether or not EC is correct, relatively casual uses of a term are more likely to be mistakes.

More detailed contextual views have more philosophically consequential implications. For instance, some EC views say enough about what differentiates contexts, and what the differing truth conditions are, to imply that skeptics and non-skeptics typically operate in contexts that make each position correct about what it typically

calls “knowledge.” The present point is only that contextualism in general says nothing about when there are context shifts. So even if they exist, they may not affect philosophical questions about knowledge.

Again, suppose that EC is correct in broad outline, but now suppose that philosophy does not set a single context. In skeptical contexts, the requirements for “knowledge” are severe; in various non-skeptical contexts, including some in which philosophical issues are under discussion, the requirements are variously less severe. There are still no immediate epistemological consequences, even for the merits of external world skepticism. The severe requirements of skeptical contexts might be satisfiable by felicitous external world judgments, the standards actually imposed by skeptics to the contrary notwithstanding. The outline of one such possibility is this. It might be that in skeptical contexts, only 100 percent of the right sort of probability is enough. But it is enough (along with true belief). There need not also be something that skeptics often require: internally possessed evidence that excludes all possibility of false belief. Skeptics see various ways in which our judgments often both lack this probability and are possibly false. Skeptics draw the extreme conclusion that the external world judgments must have the support of entailing evidence. But perhaps external world judgments are in the right way 100 percent probable when they are correctly causally connected to the fact known.<sup>4</sup> This might occur only in optimal cases of perception and memory. The skeptical context would set the 100 percent probability requirement, while the skeptic would mistakenly impose a more stringent standard. EC does not imply that skeptics have infallible theoretical insight into the truth conditions for “knows” that hold when they are being skeptical.

Again, EC leaves the main epistemological issues unresolved. Only versions that are specific about the contextual truth conditions do more. Those versions must be defended on their own particular merits. The general thesis of EC is neutral about the epistemically consequential details.

## A Delicate Balance

EC has never been obviously true. In this way it differs from the contextual variation in truth conditions exhibited by indexicals and comparatives like “large” and “near.” EC is defended not by direct semantic reflection among fluent speakers, but by its explanatory strengths. The classic explanatory role is to explain how both ordinary “knowledge” attributions and skeptical arguments denying “knowledge” can seem so plausible. The classic contextualist view is that each is correct. The reason is that ordinary “knowledge” ascriptions often meet the truth conditions of “knows” in everyday contexts, while skeptical denials of “knowledge” are right according to the truth conditions for “knows” operative in their contexts.

This reason does not quite yield the data to be explained. The data are the *plausibility*, or the *reasonableness*, of both ordinary “knowledge” attributions and skeptical “knowledge” denials. The classic contextualist proposal offers us a way to have those claims come out *true*. In order to use this truth to account for the plausibility in question, the explanation must add something about how the truth engenders the plausibility.

The most straightforward way to do this is to add something along the following lines. When we are thinking about ordinary speakers, and skeptics, are we in some way

cognizant of the truth conditions of their respective “knowledge” ascriptions? We are in some way cognizant that the respective truth conditions are met. And so we find the respective claims plausible.

This explanation attributes to us some sort of grip on the contextual variation in standards for “knowledge” ascriptions. We are not said to think explicitly that there is a contextual variation. But we are supposed to be guided by our understanding of “knows” to count some differences in context as decisive to truth when we make our plausibility judgments.

The problem is to understand how the correctness of EC can be so difficult to recognize that it has been missed entirely until lately, and many philosophers who consider it still sincerely deny it. How can it be that fluent English speakers are well enough attuned to this variation for it to guide some of their plausibility judgments, while none noticed that “knows” is context sensitive until recently, and some are unable to recognize this by semantic reflection, even under the tutelage of a contextualist?

There is no contradiction in this. It is not a hopeless situation for EC. There may be some adequate psychological explanation of how one can have a concept that one commonly applies in a context-sensitive way, while being unable to spot this feature of the concept when one tries. It is difficult to understand this sort of grasp, though. For a start, it would be nice to be shown other examples of concept possession with *almost* these same features, say, examples where the context sensitivity is sufficiently hidden to have been universally overlooked but is eventually recognizable by virtually all proficient speakers. The oddity of this sort of grasp of a concept poses a challenge to EC. It is another liability of the view.<sup>5</sup>

## Notes

- 1 Contextualists find differences in truth conditions. This does not automatically make the claims compatible. However, if there is any contextual variation in truth conditions, then the two claims are not the direct contradictories of one another that they seem to be. Also, denials of contradiction are what a straightforward application of contextualism engenders. It is possible to add a special provision for disputes. A special provision might claim that in a dispute something overrides each speaker’s separate context in determining truth conditions. It might be claimed that this yields one context for both of the pair of syntactically contradictory utterances. This sort of view will have a hard time locating one interpretation that is credible for both utterances, while otherwise interpreting in the spirit of the contextualism. Why does the basis for the unifying interpretation override the particular contexts of use only in disputes, and not everywhere? Furthermore, justifying this sort of special provision is especially difficult when uses of syntactically contradictory sentences are not public, but rather formulate successive thoughts by someone weighing the merits of opposing positions. There is no intuitive third context to appeal to in assigning unifying truth conditions, while the successive contexts seem tied for suitability as context determiners.
- 2 “Really” by itself, and “truly” by itself, sometimes seem to function as intensifiers, as in “a really hot day,” and “a truly long drive.” But “really and truly” seems rather to function as a way to get serious about the truth of the modified claim. Similarly functioning terms include “actually” and “genuinely.” “Really and truly” will be the phrase of choice here.
- 3 The contextualist who denies that the “really and truly” test changes the truth conditions seems to have no reasonable principled way to say why the truth conditions do change in skeptical contexts.

- 4 This is very roughly the position of Fred Dretske about knowledge in general, in his *Knowledge and the Flow of Information* (1981).
- 5 The negative perspective on contextualism that I express in this exchange has been influenced by the work of several philosophers. The following essays have been most influential: Schiffer (1996), Sosa (2000), and Feldman (2001).

## References

- Dretske, F. (1981) *Knowledge and the Flow of Information*. Oxford: Blackwell.
- Feldman, R. (2001) Skeptical problems, contextualist solutions. *Philosophical Studies* 103: 61–85.
- Schiffer, S. (1996) Contextualist solutions to scepticism. *Proceedings of the Aristotelian Society* 96: 317–333.
- Sosa, E. (2000) Skepticism and contextualism. *Philosophical Issues* 10: 1–18.

# Contextualism Defended

*Stewart Cohen*

## 1 Contextualism and Skeptical Paradoxes

Contextualism has been proposed as a way to resolve stubborn epistemological paradoxes. Where P is some commonsense proposition about the external world (e.g., I see a zebra) and H is some skeptical hypothesis (e.g., I see a cleverly disguised mule), the paradox takes the following form:

- 1 I know P.
- 2 I do not know not-H.
- 3 I know P only if I know not-H.

These propositions constitute a paradox because each is independently very plausible, yet jointly they are inconsistent. Because our intuitions about knowledge lead to paradox, skepticism threatens.

Most philosophers who consider this paradox are unwilling to accept skepticism. Instead, they attempt to provide an anti-skeptical resolution of the paradox. But one cannot satisfactorily respond to the paradox by simply denying one member of the inconsistent set. The paradox arises because each proposition seems true. To resolve the paradox, one must explain the intuitive appeal of the denied proposition. Otherwise, we have no explanation for how the paradox arises. So an anti-skeptical resolution must resolve the paradox in a way that preserves the truth of (1) while explaining the appeal of the propositions that threaten a skeptical result, i.e. (2) and (3).

One robust feature of our intuitions regarding (1) to (3) is that they tend to vacillate. We begin by confidently holding (1), that we know many things about the world. Under pressure from skeptical arguments, we see the appeal of (2). Given that (3) is

compelling, and that (3) and (2) entail the falsity of (1), we begin to waver in our assent to (1). But we find this unsatisfactory given our commonsense rejection of skepticism. But again (1) together with (3) entails that (2) is false. But when we consider the skeptical arguments for (2), we tend to fall back into a skeptical frame of mind.

The contextualist resolution of the paradox proposes that we take these intuitions at face value. Rather than rejecting one of these intuitions as mistaken, contextualism attempts to explain away the apparent inconsistency of our intuitions by arguing that they reflect the contextually varying truth conditions for knowledge ascriptions. Contextualism holds that ascriptions of knowledge are context sensitive – that the truth value of sentences of the form “S knows P” depend on contextually varying standards for how strong one’s epistemic position with respect to P must be in order for one to know P.

There are various ways of cashing out this notion of the strength of one’s epistemic position. But contextualist resolutions of the paradox are alike in holding that the context of ascription determines which proposition of the skeptical paradox gets denied. According to contextualism, in everyday contexts (2) is false, in stricter “skeptical contexts (1) is false, but in no context is (3) false. This resolution of the paradox explains the appeal of skeptical arguments by allowing that the claims of the skeptic are true, relative to the very strict context in which they are made. But the resolution is anti-skeptical in that it preserves the truth of our everyday knowledge ascriptions relative to the everyday contexts in which they are made. In essence, contextualism concedes that there is some truth to skepticism, but contains the damage by holding that the skeptical claims are true only relative to atypically strict contexts.

## 2 Alternative Accounts

We can think of contextualism as involving two theses:

- (a) Ascriptions of knowledge are context sensitive.
- (b) The context sensitivity of knowledge ascriptions provides the basis for resolving the skeptical paradox.

Criticisms of contextualism can challenge (a), or grant (a) while denying (b). Earl Conee raises criticisms of both kinds.

Conee challenges (a) by arguing that there are alternative ways of accounting for our intuitions that saddle us with the paradox. He suggests the possibility that the truth conditions for knowledge ascriptions are invariant – most plausibly “very high, but not unreachably high” – but that the standards for making appropriate, though not strictly speaking true, attributions vary with context. In everyday contexts, “when nothing much turns on it” people will ascribe knowledge to themselves and others on a basis that is “significantly weaker than what is actually required.” But competent speakers realize that this is loose talk, as evidenced in their willingness to withdraw their claims when challenged as to whether the proposition is “really and truly” known. Conee notes that “only the most conspicuous facts of current perception, the clearest memories, triple-checked calculations and the like” will pass the “really and truly” test.

Conee's point here is that there is nothing in the intuitive data to support contextualism over this alternative account. But although the account Conee suggests is consistent with the intuitive data, I don't see that this account does explain the intuitive data as well as contextualism. The problem for Conee's alternative is that competent speakers, under skeptical pressure, tend to deny that we know even the most conspicuous facts of perception, the clearest memories, etc. For the contextualist, this is because the strictness of the standards in skepticism make our skeptical denials true, in those contexts. But Conee's alternative incorrectly predicts that we should hang on to such knowledge ascriptions, even in the face of skeptical "really and truly" challenges.

As Conee notes, the contextualist should say that when one applies the "really and truly" test, one is in fact raising the standards, and in this new context, the truth conditions for sentences of the form "S knows P" are so strict as to falsify the ascriptions. This would seem to commit the contextualist to taking a similar line for a whole host of predicates that behave similarly when subjected to the "really and truly" test – for example, happy. But according to Conee, that is not what seems to be happening when we consider questions like "Am I really and truly happy?" It seems to us that we are setting aside loose talk and trying to get to the truth of the matter. According to Conee, this is a liability for contextualism.

There is no doubt that when one is in what the contextualist wants to view as a high standards context, one has a feeling of enlightenment regarding the correct application of the predicate in question. We feel as if we are seeing the truth of the matter that has, up until that point, eluded us. So we have the "witting and heartfelt" sense that, contrary to what contextualism holds, our skeptical judgments conflict with our previous everyday judgments. So the contextualist has to argue that we are in all of these cases mistaken in our meta-judgment that our skeptical judgments conflict with our earlier judgments. Conee holds that this is a liability for contextualism. But the extent to which this is a liability will depend on the plausibility of attributing to us this kind of mistake. I will return to this issue below.

Conee goes on to raise objections to (b). He suggests two ways in which contextualism might be misleading, even if it turns out that ascriptions of knowledge are context sensitive. First, it might be the case that the skeptic (or the skeptic in us) is mistaken in thinking that the stricter skeptical standards are unsatisfiable. Second, it may be that all philosophical discussions are carried out in one strict, "really and truly" context. If either of these possibilities obtains then the fact that ascriptions of knowledge are context sensitive will not, by itself, settle the main *philosophical* questions about knowledge.

Conee is right that even granting the context sensitivity of knowledge ascriptions, nothing strictly follows about what the standards are in particular contexts. That is, (b) is logically stronger than (a). For the contextualist solution to work, it must be that the standards in everyday contexts are low and attainable, and the standards in skeptical contexts are high and unattainable. The contextualist proposes that if we take our shifty intuitive judgments, both skeptical and commonsensical, at face value, then we get just those results.

Of course, the contextualist has no *proof* that our intuitive judgments are correct. What contextualism offers is a non-skeptical way to resolve the skeptical paradox. It proposes an explanation of our inconsistent inclinations that accounts for the appeal of skeptical arguments, while still preserving the truth of our everyday knowledge ascriptions. Now for all that, contextualism may be false. But the fact that it can do

this much shows that there is much to recommend it. In the end, whether contextualism should be adopted depends on how it compares with other proposals for resolving the paradox.

This brings us back to Conee's proposals. The first is that the standards are context sensitive, but we are mistaken in thinking that the strict skeptical standards are not met. Perhaps this is so. But absent some defense of a particular proposal as to how we are mistaken, as to how we can satisfy what appear, in some contexts, to be the very strict standards for knowledge, the proposal does not count for much. (Conee mentions Dretske's information-theoretic account, but I don't think he is seriously endorsing it.) At best, it shows how the falsity of the contextualist solution to the epistemological paradox is consistent with the thesis that ascriptions of knowledge are context sensitive. And this is something the contextualist can concede.

Moreover, even if we had a plausible account of how we could meet the stricter standards, such an account would have to explain why we were inclined to make the mistake of thinking that we cannot meet those standards, when in fact we can. Recall that these are intuitive judgments. It is not clear that our intuitive skeptical judgments are made on the basis of some false theory we hold about what is required for knowledge. If the correct theory entails that our skeptical intuitions are mistaken, the question still remains as to why we have those intuitions. Only with such an explanation would we have a resolution of the paradox.

Conee's second proposal is that although the standards for knowledge vary across contexts, there is a single standard that is always in effect in philosophical discussions of knowledge, "from the Gettier problem, to the extent of scientific knowledge, from the nature of justification to the merits of skepticism all have been discussed solely in the 'really and truly' context." This is another possibility wherein "knowledge" is context sensitive, but that fact has no implications for important philosophical issues.

Again, Conee is correct that nothing philosophically significant strictly follows from the thesis that ascriptions of knowledge are context sensitive. The point of contextualism is to provide an explanation for the intuitive data in a way that can explain the appeal of skepticism while still preserving the truth of our everyday knowledge ascriptions. The fact that contextualism can provide such an explanation does not entail that it is the only possible account, even on the supposition that ascriptions of knowledge are context sensitive. So we need to compare the relative merits of competing accounts.

An important fact about our epistemological intuitions is that even when we are discussing philosophy, we vacillate between thinking that we know and thinking that we do not. When we discuss the Gettier problem, we readily allow that we know all sorts of things. We use these intuitions about when we know as data in constructing our theory of knowledge. But when we discuss skepticism, we have intuitions that we don't know the very things we thought we knew when discussing the Gettier problem.

How do we explain this shiftiness in a way that avoids skepticism while still explaining its appeal? Contextualism proposes a way to do this in terms of contextually shifting truth conditions. The view that Conee mentions provides no such explanation. Since our intuitions vacillate even in philosophical contexts, if, as Conee suggests, there is a single standard for all such contexts, then some of those intuitions are mistaken. But which ones are mistaken, and why, if they are mistaken, do we have them? Conee's suggestion is merely the bare statement of a possibility that leaves



entirely open whether skepticism is true, and if not, why we find it appealing. Thus the contextualist explanation of our intuitions, while not entailed by the thesis that ascriptions of knowledge are context sensitive, has much more to recommend it than the possibility that Conee raises.

### **3 Is the Contextualist Model Coherent?**

Of course, a theory's explanatory power does not count for much if the theory is independently implausible. And Conee argues that contextualism taken as a whole entails an implausible story about competent speakers. He notes that contextualism explains the appeal of the premises of the skeptical paradox by arguing that each is in fact true, relative to a particular context. Thus contextualism must hold that the appeal of the premises of the paradox reflects our grasp of their truth conditions. So it must be that judgments of competent speakers, to the extent that their judgments are so guided, are sensitive to the contextual shifts in the truth conditions.

But at the same time, contextualism involves a kind of error theory with respect to certain meta-judgments about knowledge. According to the view, when we are in skeptical contexts where (1) is false, we mistakenly think our denial of (1) conflicts with our everyday assent to sentences of that form. That is, we fail to recognize the context sensitivity of our own judgments even though we grasp, in some sense, their shifting truth conditions. Conee claims that this account of competent speakers is "difficult to understand" and challenges the contextualist to provide other examples of our use of context-sensitive predicates that fits the model the contextualist proposes in the case of "knows."

As I have argued, there are precedents for the contextualist model (see Cohen, 1999, 2001). Consider ascriptions of flatness. You can lead competent speakers to question their everyday ascriptions of flatness by making salient "bumps" that ordinarily we do not pay attention to. As Peter Unger demonstrated, taking this strategy to the extreme, for example, by calling attention to microscopic surface irregularities, one can lead competent speakers to worry whether anything is really flat (Unger, 1975). But Unger's case for flatness skepticism is interesting precisely because many who feel the pull of flatness skepticism look back on their previous flatness ascriptions and think they may have been wrong.

Should we worry that all along we have been speaking falsely when we have called things "flat"? Surely not. Philosophical reflection will convince most that ascriptions of flatness are relative to context-sensitive standards. Roads that count as flat in a conversation among Coloradans do not generally count as flat in a conversation among Kansans. And while one can truly ascribe flatness to a table in everyday conversations, one might not be able to truly ascribe flatness to that same table when setting up a sensitive scientific experiment. If we implicitly raise the standards high enough (by making salient microscopic bumps), then perhaps, relative to that context, no physical surface really is flat. But of course, that does not impugn our ascriptions of flatness in everyday contexts where the standards are more lenient.

So the controversy over whether anything is flat can be resolved by noting that ascriptions of flatness are context sensitive. And to the extent that competent speakers' ascriptions of flatness are guided by these contextually shifting truth conditions, they

in some sense grasp those truth conditions. But then why can we get competent speakers to question their everyday flatness ascriptions by implicitly raising the standards? It must be that although competent speakers are guided in their flatness ascriptions by contextually shifting truth conditions, such speakers can fail to realize that their flatness ascriptions are context sensitive. And because they can fail to realize this, they can mistakenly think that their reluctance to ascribe flatness, in a context where the standards are at the extreme, conflicts with their ascriptions of flatness in everyday contexts. That is to say, they conflate the proposition expressed by “X is flat” at a strict context, with the proposition expressed by that sentence at a more lenient context. And this is precisely what the contextualist argues occurs in the case of knowledge ascriptions.<sup>1</sup>

Having said that, I should note that there is an important difference between the two cases. Contextualist theories of flatness ascriptions, once proposed, gain easy and widespread acceptance among most people. But, as Conee points out, contextualist theories of justification/knowledge do not. This is something a contextualist – one like me anyway who relies on the analogy – needs to explain.

One thing the contextualist can say here is that there are varying degrees to which competent speakers are blind to context sensitivity in the language. Thus, while everyone can readily see the context sensitivity of indexicals like “I” and “now,” it can take some amount of reflection to convince most that “flat” is context sensitive. And for a term like “knows,” it may be very difficult even after some amount of reflection for competent speakers to accept context sensitivity. It may take subtle philosophical considerations concerning the best way to resolve a paradox in order to “see” the context sensitivity of “knows.”

Here is a somewhat speculative attempt to explain why it is difficult to accept that “knows” (and “justified”) is context sensitive. Justification and knowledge are normative concepts. To say that a belief is justified or constitutes knowledge is to say something good about the belief. We value justification and knowledge. But contextualist theories are deflationary. Contextualism about knowledge says that most of our everyday utterances of the form “S knows P” are true, even though the strength of S’s epistemic position in those instances does not meet our highest standards. In the same way, contextualism about flatness says that most of our everyday utterances of sentences of the form “X is flat” are true, even though X’s surface may fall short of perfect flatness.

In other words, contextualism is a “good news, bad news” theory. The good news is that we have lots of knowledge and many surfaces are flat; the bad news is that knowledge and flatness are not all they were cracked up to be. We find this much easier to accept in the case of flatness than knowledge, because ascriptions of flatness do not have the normative force that ascriptions of knowledge/justification do.

## Note

- 1 A perhaps less clear case involves ascriptions of solidity. With the rise of our understanding of atomic physics, it became clear that many of the objects we think of as solid actually contain quite a bit of empty space. This led some people to conclude that no physical objects are solid. Others found this conclusion to contradict common sense. No doubt some

vacillated between agreeing with the conclusion and rejecting it. Again, I assume that philosophical reflection will convince most that the correct thing to say here is that ascriptions of solidity are context sensitive. Here it seems most plausible to view ascriptions of solidity as involving a kind of implicit quantification and the context as governing a shifting domain. So for  $x$  to be solid, it must contain no spaces between its parts. But what counts as a part depends on the context. If this is correct, then we have another case where competent speakers' judgments are guided by contextually shifting truth conditions, but the speakers are unaware of this fact.

## References

- Cohen, S. (1999) Contextualism, skepticism, and the structure of reasons. *Philosophical Perspectives* 13: 57–89.
- Cohen, S. (2001) Contextualism defended: comments on Richard Feldman's "Skeptical problems, contextualist solutions." *Philosophical Studies* 103: 87–89.
- Unger, P. (1975) *Ignorance: A Case for Skepticism*. Oxford: Oxford University Press.

# Contextualism Contested Some More

*Earl Conee*

Stewart Cohen offers EC as a way to solve an epistemological problem that he counts as a paradox. Here is the standard illustration of the problem. While I see an animal in a zoo enclosure:

- (1) I know that I see a zebra.
- (2) I do not know that I do not see a cleverly disguised mule.
- (3) I know that I see a zebra only if I know that I do not see a cleverly disguised mule.

Stew reports that our intuitions about the truth of (1) to (3) robustly tend to vacillate. We are drawn to (1) by common sense. Skeptical arguments lead us to (2). (3) is compelling all along, and inferences from (3) and each of (1) and (2) lead us to the denial of the other.

Stew's EC theorist has it that each of (1) and (2) is true at a context from which it is found intuitive. The intuitive appeal of (1) is to be explained by the proposal that there are relatively lenient truth conditions for (1) in everyday contexts. The plausibility of a skeptical conclusion is to be explained by the proposal that skeptical claims that imply (2) are true at contexts with relatively strict truth conditions for sentences using "know."

As I write, a painting is in the news. Thoughts about the reports provide a model of an intellectual conflict similar to (1) to (3) that does not allow a contextualist resolution.

A woman bought a painting for \$5. It is a large canvas covered with bright squiggles and blotches of paint. Eventually the woman happened to contact an art professor about it. She was led to believe that the artist might well be the renowned twentieth-century American painter, Jackson Pollock. An expert found a Pollock fingerprint in the paint. He declared it a genuine Pollock. An art institute examined the painting. After careful study, the institute announced that the painting is not by Pollock. There the matter stands.<sup>1</sup>

Considering the case, I vacillate between two conclusions.

C1: The painting is by Jackson Pollock.

C2: The painting is not by Jackson Pollock.

On different bases, each seems true. I am tempted to accept C1 when I think:

How could Pollock's fingerprint get in a painting that looks like a Pollock to an expert, if Pollock didn't paint it? There is no other plausible way.

I am tempted to accept C2 when I think:

The art institute has its reputation at stake. They are aware of the fingerprint evidence. Yet they say that the work is not by Pollock.

I incline toward C1 again when I note that the art institute is relying heavily on the lack of an established history for the work, while I doubt that it is peculiar to lose track of one painting. I incline toward C2 again when I note that I may know little of the reputable art institute's full grounds for their negative conclusion. I alternate between C1 and C2 on the basis of these reasons. I become perplexed and reluctant to judge.

This is no paradox. There is a straightforward empirical fact of the matter (leaving vagueness aside). I happen to have pretty strong evidence for each conclusion with no overriding consideration. I alternate between C1 and C2 by employing different standards of judgment, each of which is normally sufficient for reasonable belief. Some of my standards make a mistake here. This is not mysterious. We have no good grounds to think that such standards are universally correct.

Finally about this case, a reconciling contextualist interpretation of my perplexity is out of the question.<sup>2</sup> The thoughts that I formulate with C1 and C2 are incompatible. Now let's return to Stew's (1) to (3), and its EC resolution.

Intuitive responses do incline us toward accepting (1) and (2), though we can become perplexed and reluctant to judge. It may be misleading to classify the responses as intuitions. Neither (1) nor (2) is just credible on its own. Each seems true, when it does, on the basis of something else. The bases differ. (1) seems true when we naively consider the facts of an ordinary zoo visit. (2) seems true partly as a consequence of plausible skeptical reasoning. For instance, a typical key premise alleges that a visitor to the zoo has no evidence that distinguishes the animal seen from a cleverly disguised mule.

The appeal of (1) and (2) derives in one way or another from other considerations. I shall call any such basis a "standard for judgment." It is of some importance that the plausibility of (1) and (2) derives from standards for judgment, that this joint plausibility is perplexing, and that we have no reason beyond the limited credibility of our standards to regard them as unerring. This all makes (1) to (3) significantly like my C1

and C2. There the truth conditions do not vary. I am drawn to C1 and C2 by credible but fallible standards, while I have no overriding evidence about which standards support a falsehood. The counterpart invariantist view of (1) to (3) beckons.

Does EC have an explanatory advantage? The EC resolution is crucially incomplete. Where (1) to (3) seem true, an EC theorist locates satisfied truth conditions. But satisfied truth conditions do not explain plausibility. What directly explains the plausibility to a person of a sentence is whatever leads the person to regard the sentence as true. An explanation of plausibility must cite something that influences the person's thinking.

Presumably EC theorists will supplement their explanation with claims about our use of commonsense standards for judging (1) and skeptically encouraged standards for judging (2). To involve varying truth conditions in the account, EC theorists will also claim that our standards are sufficient for the truth of (1) and (2) in their respective contexts. They might say that we regard each context's truth condition itself as our standard there.

Imputing a standard for judgment yields the same immediate explanation of plausibility, whether or not we add that it establishes truth. So the distinctive EC claim that our standards establish the truth of (1) and (2) is idle for this explanatory purpose.

Some of the standards have a notable vulnerability that EC overlooks. Our inclination toward (2) results from some more or less philosophical reasoning. It is alluring enough to draw us toward (2). But like most alluring philosophical arguments, it is subject to reasonable objection. At best, its success remains in doubt. EC sweeps the doubts aside and declares that the reasoning in support of (2) yields genuine truth conditions for knowledge. No philosophical reasoning deserves such deference.

Is the EC claim that the relevant standards establish truth an asset on other grounds? A complete account of our responses to (1) through (3) also says why we reasonably and confidently employ the pertinent standards. EC theorists can hold that our standards for judging (1) and (2) are somehow tacitly acquired in learning the meaning of "know." They might say that our learning partly consists in acquiring a rule for accessing a context's truth condition, and the rule yields the standards that make (1) and (2) plausible to us. EC theorists can then hold that the analytic link of these standards to the meaning of "know" explains why the standards seem so trustworthy to us.

Invariantists can offer a better explanation of our reasonable confidence in the standards. Invariantists can observe that although the standards that we apply to (1) and (2) may include ones we are taught in learning "know," they are not therefore semantically guaranteed to be necessary or sufficient conditions for the application of the term. They can be stereotypes offered to guide usage, well supported rules of thumb, or apparent implications of convincing theories. Nothing readily available shows us which appealing standards fail here. An account along these lines explains the attractiveness of (1) and (2) in a way that better accommodates other features of the situation. Allowing our standards to be fallible lets us be right about the conflict that we see among the propositions that attract our acceptance as we consider (1) to (3).

## **Brief Replies**

R1. Stew's EC is the conjunction of

- (a) Ascriptions of knowledge are context sensitive,

and

- (b) The context sensitivity of knowledge ascriptions provides the basis for resolving the skeptical paradox.

Stew writes that I opposed (a) by offering a particular invariantist view of knowledge to account for “our intuitions that saddle us with the paradox.”

I did not offer the invariantist view to explain what Stew regards as our intuitions about knowledge. As may now be clear, I believe that those intuitive responses have no simple connection to the actual conditions on knowledge. I offered the view as a possibility about the nature of knowledge that accords with our having a specific intelligent basis for a variety of the knowledge attributions that the view counts as false. The basis is that those attributions are efficient loose talk. The “really and truly” test is supposed to show that this status is how we ourselves think of the attributions in our best moments.

R2. Stew suggests that there is a telling vacillation in intuitive knowledge judgments within philosophy. During Gettier problem discussions we readily allow that we know many things, while discussions of skepticism give rise to intuitions that we do not know those things. EC can offer shifting truth conditions within philosophy to explain the vacillation. I discussed the possibility of a context-sensitive view of “knows” that finds within philosophy no variation of truth conditions. This view leaves the vacillation unexplained.

I was thinking of EC as Stew’s (a) only. I presented the particular context-sensitive view of “knows” as an example of how (a) could be true while leaving philosophy unaffected. In support of this sort of context sensitivity, I offered just the observation that philosophy seems to occupy a “serious” context.

What does account for the vacillation within philosophy? When we consider the Gettier problem, most of us (some philosophers are spoilsports about this) apply commonsense standards for knowledge. Any of them will do. When we think hard about skepticism, we tend to apply more severe standards. (Here too there have been spoilsports, notably G.E. Moore.) These are reasonable proclivities, if only because the standards are initially plausible. Philosophers’ use of such standards explains the vacillation. Any claimed connection between this variation in our standards and the actual truth conditions requires argument.

R3. Stew offers a speculative account of why some find it difficult to accept EC (where EC is understood as the conjunction of (a) and (b)). We regard knowledge as a good thing. If EC is true, our external world “knowledge” is not so hot. We only meet disappointing truth conditions. Some hold out against EC to avoid this disappointment.

Probably this accounts for some resistance to EC. It does not hit home with me. My resistance derives primarily from two sources. First, it seems to me that skeptical views deny the very knowledge that is otherwise widely affirmed. EC implies otherwise. On further reflection, the thoughts continue to seem incompatible.

Explanatory success could conceivably outweigh this reflective liability. That brings us back to points about explaining our knowledge attribution inclinations. Invariantists can credibly explain our reasonably confident use of the varying standards for judgment that EC theorists invoke.<sup>3</sup> The distinctive EC thesis is that some of our

relatively high standards, and some of our less high standards, are contextual truth conditions, or derive from them. This gives an EC explanation no advantage. It incurs the costs of caving in to dubious skeptical reasoning and requiring some durable impressions of incompatibility to be errors.

## Notes

- 1 Some of these claims about the case may turn out to be inaccurate. By the time you read this, the question may well be settled. In any event, the example as described serves to make the point.
- 2 The sentence types display some contextual variation. “The painting” has different referents in different contexts and probably the name “Jackson Pollock” does too. But to formulate my thoughts with C1 and C2, we must hold fixed the referents of those two expressions. The resulting thoughts are irreconcilable.
- 3 Toward providing an invariantist explanation of our attitudes to the standards, I have just mentioned stereotypes, rules of thumb, and inference from theories. An adequate explanation would elaborate. The point is only that some invariantist explanatory possibilities are at least as initially credible as the envisaged EC account.

# Contextualism Defended Some More

*Stewart Cohen*

## 1 When Should We Appeal to Contextualist Resolutions of Philosophical Paradoxes?

Earl Conee raises important issues about when we should appeal to a contextualist resolution of a philosophical problem. He argues, in effect, that contextualists are too quick to appeal to a contextualist resolution of our conflicting intuitions regarding knowledge. His strategy is to present another case that he argues is structurally identical to the knowledge case where a contextualist resolution looks to be uncalled for. This, he argues, should make us question the validity of appealing to a contextualist resolution of the knowledge case.

Conee’s case involves a conflict over whether a particular painting is by Jackson Pollock. On the one hand, an expert testifies that the painting contains Pollock’s fingerprint. On the other hand, an art institute has testified that the painting is not by Pollock. This leads us to vacillate over whether or not the painting is by Pollock.

About this case, Conee makes the following points:

- (a) Different considerations pull us in different directions regarding a particular matter of fact.
- (b) The considerations in both directions are fallible but we have no overriding basis that tells in favor of one set of considerations over the other.
- (c) We find this situation perplexing.

All the same, according to Conee, “This [case] is no paradox. There is a straightforward empirical fact of the matter (leaving vagueness aside).”

Conee then notes that the skeptical paradox is like his Pollock case in respect of (a) to (c). But, Conee argues, there is no temptation in the Pollock case to appeal to a contextualist semantics for the predicate “... is by Pollock.” Clearly here, the truth conditions are invariant. And this, he argues, should lead us to question whether a contextualist resolution of the skeptical case is correct.

Conee is surely correct that we should not necessarily be attracted to a contextualist solution anytime we have a case that satisfies (a) to (c). And I agree that a contextualist resolution of the Pollock case would be very implausible. But I will argue that there are important disanalogies between the two cases that make a contextualist resolution of the skeptical case plausible.

Perhaps the most important difference is that, as Conee himself points out, there is nothing paradoxical about the question of whether the painting is by Pollock. It is a straightforward empirical matter. Our only problem is that we lack empirical information – information that presumably would settle the matter.

But the knowledge case is not a straightforwardly empirical matter. The question about whether we know is arguably not an empirical question. In fact, there need not be any disagreement between skepticism and common sense about any empirical fact. In the skeptical case, the issue is conceptual. Different considerations pull us in different directions concerning whether or not a particular set of empirical facts is properly classifiable as an instance of knowledge.

Conee questions whether it is correct to call our conflicting responses about the knowledge case “intuitions.” For our each response is not “credible on its own,” but rather based on “something else.” I agree that our responses are based on various considerations. Our response that we know we see a zebra is based on the facts of the case as described. Our response that we don’t know we don’t see a cleverly disguised mule is based on the appeal of various skeptical principles. But I don’t see that they are not intuitions. Our response that we know we see a zebra is an intuitive judgment about whether the concept of knowledge applies in a particular case. And our response that we don’t know we don’t see a cleverly disguised mule is an intuitive judgment about the cogency of a particular argument. The fact that nothing empirical is at stake makes it appropriate to think of these responses as intuitive judgments.<sup>1</sup> And the fact that the Pollock case is clearly empirical makes it clear that our conflicting judgments in that case are not intuitive. We cannot decide by intuition whether or not a particular painting is a Pollock.

So there is an important difference between the Pollock case and the skepticism case. The former involves empirical judgments about an empirical matter of fact, whereas the latter involves intuitive judgments about whether a conceptual matter, or to ascend semantically, about whether a particular predicate applies. Of course, even if we are dealing with intuitive judgments concerning the application of a predicate, it does not follow that a contextualist treatment is called for. There must be a semantical model of the predicate that makes a contextualist treatment plausible. Different contextualists have proposed different models.

Lewis argues that ascriptions of knowledge involve an implicit quantification (Lewis, 1996). *S* knows *P* just in case *S*’s evidence eliminates every alternative. But what counts as an alternative that *S*’s evidence must eliminate depends on context. So for Lewis, the context sensitivity of “knows” is a function of contextual restrictions on the domain of quantification.



On my own view, the context sensitivity of knowledge is inherited from one of its components, i.e. justification (Cohen, 1999). In general, when a predicate has both a comparative form and a *simpliciter* form, context will determine the degree to which the predicate has to be satisfied in order to be satisfied *simpliciter*. So, for example, context determines how flat a surface must be in order to be flat *simpliciter*. Analogously, context will determine how justified a belief must be in order to be justified *simpliciter*. So if justification is a constituent of knowledge, and “justified” is context sensitive, that provides some reason to think “knows” will be context sensitive.

So this makes for another important difference between the skeptical paradox and Conee’s Pollock case. Contextualists about knowledge have provided semantic models for “knows” that explain its context sensitivity in a way that allows the dispute about knowledge to be resolved. But I can see no way to model the predicate “... is by Pollock” that would explain the sort of context sensitivity required to resolve the dispute over whether the painting is a Pollock.

## 2 Explaining the Skeptical Paradox: Contextualism versus Invariantism

Conee also argues that the contextualist does not really explain the appeal of the sentences that constitute the skeptical paradox. On the contextualist view, those sentences come out as true, in those cases where they strike us as true. But Conee notes that “satisfied truth conditions do not explain plausibility.” According to Conee, the contextualist needs an explanation of the plausibility of those sentences involving “... something that influences the person’s thinking.”

But he then goes on to explain quite well how a contextualist can explain the appeal of the sentences that constitute the paradox:

EC theorists can hold that our standards for judging (1) and (2) are somehow tacitly acquired in learning the meaning of “know.” They might say that our learning partly consists in acquiring a rule for accessing a context’s truth condition, and the rule yields the standards that make (1) and (2) plausible to us. EC theorists can then hold that the analytic link of these standards to the meaning of “know” explains why the standards seem so trustworthy to us.

I could not have said it better myself. But Conee goes on to argue that invariantists can offer a better explanation of our intuitions regarding the sentences constituting the paradox.

Invariantists can observe that although the standards that we apply to (1) and (2) may include ones we are taught in learning “know,” they are not therefore semantically guaranteed to be necessary or sufficient conditions for the application of the term. They can be stereotypes offered to guide usage, well supported rules of thumb, or apparent implications of convincing theories. Nothing readily available shows us which appealing standards fail here. An account along these lines explains the attractiveness of (1) and (2) in a way that better accommodates other features of the situation. Allowing our standards to be fallible lets us be right about the conflict that we see among the propositions that attract our acceptance as we consider (1) to (3).

So on Conee's view, it is better to explain our intuitions about the sentences (1) to (3) in a way that allows us to be wrong about them. By so doing we can vindicate our judgment that our skeptical endorsement of (2) conflicts with our commonsense endorsement of (1). Recall that on the contextualist view, that judgment is a mistake that results from conflating contexts.

So here we have a clear contrast between the contextualist response to the paradox, and Conee's invariantist account. On the contextualist view, our first order judgments (1) to (3) are correct. More specifically, our judgment that (1) is true is correct when evaluated at everyday contexts, our judgment that (2) is true is correct when evaluated at stricter skeptical contexts, and our judgment that (3) is true is correct at every context. But our meta-judgment that (1) to (3) are inconsistent is a mistake. On the invariantist view, we are correct in our meta-judgment that (1) to (3) are inconsistent, but we are mistaken in at least one of our first-order intuitive judgments concerning (1) to (3).

Each view provides an explanation for the sort of mistake it attributes to us. On the invariantist view, our mistaken endorsement of both (1) and (2) results from the fallibility of stereotypes, rules of thumb, and our ability to see logical implications.

On the contextualist view, our mistaken judgment that (1) conflicts with (2) results from our conflating contexts. And the contextualist can appeal to the very same kind of mistake in the case of the puzzle concerning flatness to avoid the charge of special pleading.<sup>2</sup>

So which explanation is better, the invariantist or the contextualist? It is important to see that the contextualist explanation of our intuitions in the paradox, if correct, constitutes a resolution of the paradox. It resolves the apparent inconsistency of (1) to (3) in a way that accounts for the cogency of skeptical arguments without compromising our strong sense that we know much about the world.

But the invariantist explanations suggested by Conee do not, in themselves, resolve the paradox. According to those explanations, there is a genuine conflict between our commonsense judgment that (1) is true and our skeptical judgment that (2) is true. Owing to the use of stereotypes, or rules of thumb, and so on, we are mistaken in at least one of those judgments. The problem is that we are unable to determine which is the mistaken judgment. As Conee notes, "Nothing readily available shows us which appealing standards fail here." In the end, contextualism will have to be weighed against worked-out versions of invariantism which explain which standards fail and why.<sup>3</sup>

### **3 Objections to Contextualism**

Conee mentions two reasons in particular to worry about the feasibility of a contextualist resolution of the paradox: "It incurs the costs of caving in to dubious skeptical reasoning and requires some durable impressions of incompatibility to be errors." Let's take these in reverse order.

As Conee says, contextualism holds that our impression that (1) to (3) are inconsistent is an error. But (1) to (3) constitute a paradox, precisely because we have the "durable impression" that each is true and that as a set they are inconsistent. But any response to the paradox will require that at least one of these impressions is erroneous. While the contextualist holds that our impression that (1) to (3) are jointly inconsistent is in error, the invariantist holds that our impression that they are all true is in error. Moreover, as I argued earlier, the contextualist can bolster the plausibility of attributing

such a mistake to us by appealing to an analogous mistake in our sometimes paradoxical thinking about flatness.<sup>4</sup> So I don't see that contextualism here is at a comparative disadvantage.

In what way is the skeptical reasoning dubious? Surely there is no obvious error in the reasoning. Otherwise we would not be stuck with a stubborn paradox. We might say the reasoning is dubious in that we are confident that the conclusion is false. Surely this is true. Again, the reason we are stuck with the paradox is that we find the conclusion of the skeptical argument to be intuitively unacceptable.

Does contextualism cave in to skeptical reasoning? Yes and no – the former because contextualists allow that the sentences expressing skeptical conclusions are true, evaluated at skeptical contexts, the latter because the contextualists hold that those very same sentences are false when evaluated at everyday contexts.

Does this concede too much to the skeptic? Of course, this is something about which reasonable people can disagree. But by my lights, what is truly surprising and intuitively repugnant about skepticism is the thought that all along in our everyday lives, when we have been uttering sentences of the form “S knows P,” we have been speaking falsely. Contextualism denies that this is so.

Contextualism does explain the stubborn appeal of skeptical arguments by allowing that there is some truth in skepticism. But the strategy of the contextualist is to limit the damage. So although contextualism does concede there is some truth in skepticism, it limits the damage by showing how this concession still allows for the truth of our everyday knowledge ascriptions. To me this is no worse than conceding that in certain limited and unusual contexts, we can truly say “Nothing is flat,” while at the same time holding that many of our everyday utterances of the form “X is flat” are true.

## Notes

- 1 Of course there is a general problem about what an intuition is, but to discuss it would go beyond the scope of this essay, not to mention my understanding.
- 2 I discuss the “flat” case in my first contribution to this exchange. The contextualist need not hold that our first-order knowledge ascriptions are infallible. The contextualist need only hold that the explanation of what is occurring in the skeptical paradox is that our judgments reflect our grasp of the correct standards for the context.
- 3 I am indebted to Earl Conee here for helping me to see this issue more clearly.
- 4 Again, see the first response to Conee in this exchange.

## References

- Cohen, S. (1999) Contextualism, skepticism, and the structure of reasons. *Philosophical Perspectives* 13: 57–89.
- Lewis, D. (1996) Elusive knowledge. *Australasian Journal of Philosophy* 74: 549–567.

## Chapter Four

# Do Practical Matters Affect Whether You Know?

What is the relationship between the propriety of an action based on a belief, on the one hand, and the epistemic justification of the belief itself, on the other? Jeremy Fantl and Matthew McGrath argue that the link is captured by the principle that if your belief that *P* is epistemically justified, then you're warranted to act in ways suggested by *P*. And since knowledge requires justification, it follows that if you know that *P*, then you're warranted to act in ways suggested by *P*. In short, knowledge is actionable. Fantl and McGrath's view has the consequence – surprising from the perspective of orthodox epistemology – that whether you know, or are justified in believing, *P* depends in part on how much is at stake in your practical situation. This is called *pragmatic encroachment* or *epistemic impurism*. Justification is not purely a matter of evidence. In his response, Baron Reed argues against the sort of view that Fantl and McGrath defend. Reed argues that there are counterexamples to some of the key principles used to argue for pragmatic encroachment. Reed also argues that pragmatic encroachment overlooks an essential feature of knowledge, that it mischaracterizes practical rationality, and that the relevant intuitions can be explained by alternative means while rejecting pragmatic encroachment.

## Practical Matters Affect Whether You Know

*Jeremy Fantl and Matthew McGrath*

William Clifford (1886) famously argued that “it is wrong always, everywhere, and for any one, to believe anything upon insufficient evidence” (1886: 346). In arguing for this he asked us to imagine that:

A shipowner was about to send to sea an emigrant-ship. He knew that she was old, and not over-well built at the first; that she had seen many seas and climes, and often had needed repairs. Doubts had been suggested to him that possibly she was not seaworthy. These doubts preyed upon his mind and made him unhappy; he thought that perhaps he ought to have her thoroughly overhauled and refitted, even though this should put him to great expense. Before the ship sailed, however, he succeeded in overcoming these melancholy reflections. He said to himself that she had gone safely through so many voyages and weathered so many storms that it was idle to suppose she would not come safely home from this trip also. He would put his trust in Providence, which could hardly fail to protect all these unhappy families that were leaving their fatherland to seek for better times elsewhere. He would dismiss from his mind all ungenerous suspicions about the honesty of builders and contractors. In such ways he acquired a sincere and comfortable conviction that his vessel was thoroughly safe and seaworthy; he watched her departure with a light heart, and benevolent wishes for the success of the exiles in their strange new home that was to be; and he got his insurance-money when she went down in mid-ocean and told no tales. (339)

Asks Clifford, "What shall we say of him? Surely this, that he was verily guilty of the death of those men" (339).

Let us agree that Clifford's verdict is correct. The shipowner – sending the ship out as he does while stifling legitimate doubts about the ship's seaworthiness – is guilty of any deaths that result when the ship sinks because it unseaworthy. What, though, is the source of the shipowner's guilt? We might think that the source of the guilt is the mere fact that the ship sank and the passengers died – that it is because the passengers died that the shipowner is culpable. But, says Clifford, that cannot be the whole of the matter. For even if the shipowner lucked out and the ship returned safely, the shipowner would have been just as culpable: the shipowner "would not have been innocent, he would only have been not found out" (340).

Perhaps, then, it is not the results of the act that contribute to the shipowner's guilt. Perhaps the shipowner's guilt is constituted simply by his culpability for the act itself: on such meager evidence, the act of sending the ship out to sea is a culpable act. This, Clifford agrees, must be admitted. Regardless of the shipowner's belief, the shipowner had an obligation to inspect the ship more carefully before sending it out. But, he continues, the culpability of the act cannot exhaust the shipowner's culpability. That is because, in believing as the shipowner does, the shipowner makes it much more likely that the act will follow – by believing as he does, the shipowner increases the likelihood that he will act irresponsibly. In fact, if the shipowner's attitude did not conduce toward the irresponsible action, it would not really amount to a genuine belief:

No man holding a strong belief on one side of a question, or even wishing to hold belief on one side, can investigate it with such fairness and completeness as if he were really in doubt and unbiased; so that the existence of a belief not founded on fair inquiry unfits a man for the performance of this necessary duty.

Nor is that truly a belief at all which has not some influence upon the actions of him who holds it. He who truly believes that which prompts him to action has looked upon the action to lust after it, he committed it already in his heart. (342)

For these reasons, Clifford concludes that there is a link between the normative status of a belief and the normative status of an action suggested by that belief: "it is not possible

so to sever the belief from the action it suggests as to condemn the one without condemning the other" (342). If that's right, then in condemning the shipowner's acting upon the belief that the ship is seaworthy, we unavoidably condemn the belief itself.

We think that Clifford's Link is intuitively powerful. It also, when properly spelled out, has some surprising and counterintuitive consequences. The fact that it has such consequences means that the intuitive power of his link needs to be supplemented by argument if it is to be acceptable. We'll do that here as well. But we don't want the initial intuitive power of Clifford's Link to be lost by the fact that we are supplementing it with principled argument. For it is clear that when it comes to the condemnation of action, we often feel forced to condemn the belief that suggests the action. James Montmarquet (1993), for example, points out that:

There are times when we want to hold an agent morally to blame for conduct which, from that agent's *own* point of view, seems quite justified. Cases in point abound, including many of the murderous activities of tyrants, terrorists, racists, and religious fanatics, not to mention the less murderous activities of plain self-righteous hypocrites ... But how are we entitled to cast such moral blame and even fairly punish these individuals unless we can sometimes find these individuals culpable for having these beliefs in the first place? ... If we cannot assign culpability for holding such beliefs, how can we assign culpability for acts premised on their (assumed) truth? (1–2)

If there were no such link of the sort that Clifford alleges, it would be hard to see how the need to condemn beliefs in order to condemn action could be made coherent.

But what precisely does it mean to say that it is not possible to sever the belief from the action it suggests so as to condemn the one without condemning the other? One thing it means is that if an action suggested by some belief is condemnable, then so is the belief itself. Of course, there are many kinds of grounds for condemnation: you might condemn an action because it harms others, or harms the agent, or harms yourself. You might condemn an action because it is uncouth, or immoral, or unseemly. Some criticisms of action do not filter back down to every one of the beliefs that suggest them. For example, a thief might rob a bank because of a rather well-informed belief that the bank has a particularly large supply of cash and a particularly poor security system. The action is morally condemnable, but the belief is not. There may, of course, be other beliefs operant in the thief's action – that, for example, it is permissible for the thief to rob the bank – and these beliefs may inherit some of the culpability of the thief's action, but certainly not all the beliefs suggestive of the thief's actions do so.

Still, we can say this on behalf of Clifford: the reason the shipowner's action of sending the ship out to sea is condemnable is that the shipowner did not have enough evidence that the ship was seaworthy. This is not so in the case of the thief; the thief's action is condemnable, but not because the thief lacks evidence that the bank has a poor security system. To introduce a term of art, let's say that in the shipowner case, the proposition that the ship is seaworthy is not "warranted enough" to justify sending the ship out to sea. On the other hand, in the thief case, that the bank has a poor security system is warranted enough to justify robbing the bank (though, of course, robbing the bank was not justified for other reasons).<sup>1</sup> According to Clifford, then, when a proposition isn't warranted to justify some action suggested by the proposition, the agent's belief in that proposition is itself condemnable.

Clifford's paper is entitled "The Ethics of Belief," but the condemnation of the shipowner's belief is not just an ethical condemnation. If there is a moral requirement to trust the word of your closest friends no matter what evidence is against them, then belief in your closest friend's guilt in a murder case might be, in some sense, ethically condemnable, even if decisively supported by evidence. Here the fault of the belief doesn't stem from a lack of evidence. Not so with the shipowner's belief: the grounds of the condemnation are epistemic – had the belief been sufficiently supported by evidence, the action it suggests would *not* have been condemnable and, therefore, neither would have the belief. That means that, when Clifford says that an action is only condemnable if the belief that suggests it is condemnable, to the extent we agree with Clifford, we should think that in the cases he has in mind, if an action is condemnable, then the belief suggestive of it is, not just ethically, but epistemically condemnable; in the terminology of contemporary epistemology, the belief is "unjustified." According to Clifford, then, if a proposition suggestive of some action is not warranted enough to justify that action, then belief in that proposition is unjustified. Or, contraposing:

If your belief that *p* is justified, then *p* is warranted enough to justify actions suggested by *p*.

According to epistemological orthodoxy, agents know that *p* only if they are justified in believing that *p*. That means a simple syllogism delivers what we will call "Clifford's Link":

(Clifford's Link) If you know that *p*, then *p* is warranted enough to justify actions suggested by *p*.<sup>2</sup>

What does it take for an action to be "suggested" by *p*? Right now, suppose, you believe that you have the spice cumin at home. Nothing much hinges on this, suppose (you're not planning a dinner that requires cumin). As such, you're perfectly justified in continuing as you are – working, flying your kite, reading, or whatever you happen to be doing. You don't need to rush home and check to see if you indeed have cumin at home. But there are hypothetical situations in which much hinges on whether you have cumin. There are situations in which you are planning a cumin-involving dinner. Worse, there are situations in which you are going to have a relative visit who is deathly allergic to cumin and will fall ill even if there is an unopened bottle in the house. And the stakes might be raised even higher than this. In such situations, your evidence that you have cumin may not be sufficient for you to continue as you are. You may need to go home and check. Does Clifford's Link require that, to know that you have cumin at home, that you have cumin at home must be warranted enough to justify you in not going home in your actual, low-stakes situation? Or does it require that, to know that you have cumin at home, that you have cumin at home must be warranted enough to justify you in not going home in all possible hypothetical situations, no matter how high the stakes? If it requires the latter, then it seems that you don't know you have cumin at home. More generally, for almost any proposition we take ourselves to know, it looks like there are hypothetical cases in which the stakes would be so high that we wouldn't be able to justifiedly perform actions suggested by that proposition. If so, and if Clifford's Link requires *p* to be warranted enough to justify all actual and hypothetical actions suggested by *p*, then it looks like no propositions will be known by anyone.

This skeptical conclusion might be a cost some are willing to bear. But we think it is better to understand Clifford's Link not as meaning that if you know that  $p$  then  $p$  is warranted enough to justify, in any hypothetical situation, actions suggested by  $p$ , but rather that if you know that  $p$ , then  $p$  is warranted enough to justify, in your practical situation, actions suggested by  $p$ . This allows you to have all sorts of knowledge of propositions about which there is a reasonable amount at stake in what you do depending on whether the proposition is true. It allows you, for example, to know that you have cumin at home.

However, allowing knowledge of this sort has surprising consequences. It means there is what Jon Kvanvig (2004) has called "pragmatic encroachment" on knowledge. On this view, whether we know something depends on practical features of our context: whether we know can vary depending on what's at stake in whether the proposition is true.<sup>3</sup> To see this, consider a variant of Clifford's shipowner example. In this variant, suppose that you are not a shipowner but a visitor at a nautical museum. During your tour of the museum you pass by a full-size three-masted ship. Interested, you read the placard<sup>4</sup>:

### **The Star of Italy**

This iron three-masted ship was built by Harland and Wolff, out of Belfast. Weighing 1784 tons, it was launched in 1877 and taken out of service in 1927. It is still seaworthy.

You nod your head approvingly, think, "Huh! It's still seaworthy. How about that!", and move on. Here, you satisfy the condition for knowing that the ship is seaworthy required by Clifford's Link. And, assuming as we are that knowledge is possible even if there are some hypothetical situations in which doing what is suggested by the proposition is unjustified, we should say that you do know that the ship is seaworthy.

However, evidence of this quality would not have been enough for Clifford's shipowner to be justified in sending the ship out to sea. The say-so of a museum's placard isn't sufficient to risk the hundreds of emigrant-lives the shipowner risks by sending the ship out to sea. Therefore, if all that Clifford's shipowner had was the same evidence you have, by Clifford's Link, the shipowner would not have known that the ship was seaworthy. In short, even with the same evidence, you know, but the shipowner does not, that the ship is seaworthy. And notice that this is not because in the one case it's true that the ship is seaworthy and in the other case it's not: it may very well be true in both cases. As Clifford points out, it's not the fact that the shipowner is wrong that makes him at fault for sending out the ship or believing it's seaworthy. It's the fact that he believed and sent the ship out on insufficient evidence. Nor is belief lacking in one case but not in the other. Both you (the museum visitor) and Clifford's shipowner believe that the ship is seaworthy. What's preventing Clifford's shipowner from knowing is that he doesn't have enough evidence to know, even if he has the same evidence that you do, and we're supposing you do have enough evidence to know.

The result, we think, is not defused by bickering over this specific case – by saying, for example, that you (the museum visitor) can't know on such evidence that the ship is seaworthy. We're allowing – as we must to avoid skepticism – that you can know



that  $p$  even if there are some hypothetical cases in which you have the same evidence for and against  $p$  and you can't act on  $p$ . In such cases, even though you have the same evidence for and against  $p$  in the actual and hypothetical cases, you will differ in whether you know that  $p$ .

This possibility seems counterintuitive. If two subjects share the same evidence for and against  $p$  (and also, as is possible, satisfy all the other standard conditions on knowledge – truth, belief, and the absence of Gettierish features), it seems that one subject knows only if the other does. Knowledge can be lost or gained by gaining or losing information – by acquiring new evidence or forgetting old evidence. But you can't lose knowledge, it might seem, by changing what you care about, your available options, or the expected costs and benefits of acting in various ways. If Clifford's Link is correct, though, and skepticism about knowledge is false, it seems you can.

Such a result might prompt some, despite the initially compelling power of Clifford's Link, to deny Clifford's Link or, at the very least, wonder what more there is to be said in its favor. There is, at least, this: Clifford's Link explains nicely various linguistic habits we have – habits of citing and asking about knowledge to defend, criticize, and explain action. For instance, you might say, after someone criticizes you for taking a left instead of a right to get to the airport, "I know that there's a shortcut to the left." This habit is hard to make sense of unless, in citing knowledge, you're saying that the proposition you're claiming to know is good enough to justify the action suggested by it – namely, taking a left instead of a right. Likewise, with questions: if your spouse tells you that you should get ready to leave for the 7:50 movie, you might well ask, "Do you know that's when it starts?," implying that if the answer is yes, then leaving will indeed be justified. These habits do not vanish when the stakes are high. You might desperately need to get to the airport as soon as possible and defend your taking a left instead of a right by saying "I know this is a shortcut." Again, when the stakes are high, you might ask, "the evidence is strong, but do the scientists know this drug is safe?" But some might think this linguistic data can perhaps be explained in other ways, so for the rest of this paper, then, we'd like to offer a more principled argument in favor of Clifford's Link.

What might seem to be the problematic feature of Clifford's Link is that it makes what should be a purely epistemic concept – knowledge – subject to practical conditions. So we'd like to start our argument for Clifford's Link by arguing for a more mundane-seeming epistemic requirement on knowledge. According to closure principles on knowledge, what you know can justify you in believing all sorts of things that are entailed by what you know. So, if you know that the liquid on the table is white, that can justify you in believing that the liquid isn't red. This seems plausible, but we don't want to assume anything even this strong. One way to weaken this requirement is to talk not of what justifies what, but about what reasons knowledge can provide. If you know that the liquid spilled on the table is white, for example, then you have at least some reason to believe that the liquid isn't red. And, if you know that the liquid on the table is white, then you even have some reason for thinking that the liquid on the table is milk: you have more reason for thinking that it's milk than that it is orange juice, say. Of course, you might know that the liquid spilled on the table is white but not thereby have a reason to believe other things – for example, that the ship is seaworthy. But, in such cases, the target belief isn't suggested by what you know – that the liquid is white.

So, let's take as our first premise in our argument for Clifford's Link this: if you know that  $p$ , then  $p$  is warranted enough to be a reason you have to believe some propositions suggested by  $p$ .<sup>5</sup> If you know that the ship is seaworthy, for example, then that the ship is seaworthy is a reason you have to believe that the ship will return safely from a modest voyage on normal seas, and if it is a reason, then of course it is warranted enough to be such a reason.

But what about other beliefs – beliefs about the potential consequences, costs, or benefits of your own actions? It's one thing to say that what you know can be a reason you have to believe some neutral consequences of what you know. It's another thing to say that what you know can be a reason you have to believe some action-regarding consequences of what you know. Or so it seems. In fact, though, it is quite bizarre to use what you know as a reason to believe some neutral propositions suggested by what you know while refusing to use what you know as a reason to believe some action-regarding propositions suggested by what you know. Consider this dialogue between the shipowner and his son, where the inspection in question is a simple "seaworthiness" inspection that a ship passes (if seaworthy) or fails (if not):

- SHIPOWNER: The ship is seaworthy.  
 SON: So, it'll return safely from a modest voyage on normal seas.  
 SHIPOWNER: Yes, so it's a waste of time and money to get it inspected.  
 SON: Wait a second – what reason do you have to believe that?  
 SHIPOWNER: It's seaworthy!  
 SON: I granted that; after all, it's your reason for believing it'll return safely from a modest voyage on normal seas. I just don't see what reason you have to believe that it's a waste of time and money to get it inspected.  
 SHIPOWNER: Look, if – as you grant – it's seaworthy, the inspection will simply reveal that it's seaworthy and I will only have wasted time and money. Right?  
 SON: Yes.  
 SHIPOWNER: So that the ship is seaworthy is a reason I have to think it's a waste of time and money to have it inspected.

The son's third speech is absurd to our ears, while the shipowner's response is utterly reasonable. What the shipowner's third remark makes clear is this: that the ship is seaworthy suggests that it's a waste of time and money to have the ship inspected. Further dialogues only reinforce this point. Suppose you're at the edge of a frozen pond of an oval shape. You're at the mid-point of the long side of the oval. It's a long way around but not far across. You want to get to the other side.

- YOU: The ice is thick enough to hold me.  
 YOUR OBVIOUSLY LIGHTER SIBLING: So it's thick enough to hold me, too.  
 YOU: So, I won't fall through the ice if I walk across it.  
 SIBLING: Hold on a second! What reason do you have to believe that you won't fall through if you walk across it?  
 YOU: Well, the ice is thick enough to hold me!  
 SIBLING: Look, I agreed that's your reason for believing it's thick enough to hold you! But I was asking what reason you have for believing that you won't fall through if you walk across.

Here again your sibling's third reaction seems absurd. Just as epistemologists don't restrict closure principles to propositions that are "neutral" regarding action, we shouldn't restrict principles about reasons for belief in this way either.

Thus, we have our **second premise**: if  $p$  is warranted enough to be a reason you have to believe some propositions suggested by  $p$ , then  $p$  is warranted enough to be a reason you have to believe *any* proposition suggested by  $p$ .

Our next step links reasons for belief with reasons for action. Here especially we attempt to draw on the insights of Clifford. When we appeal to some consideration as a reason for belief, we tend to treat it as available as a reason for action as well. We do not keep two reasoning streams in our head, one with propositions available as reasons for belief and another with propositions available as reasons for actions. We treat something as available as a reason for action when it is available as a reason for belief. And not only do we do this, it seems absurd to segregate the two. Consider the following dialogue:

- SHIPOWNER: The ship is seaworthy.  
SON: Yes, so it's a waste of time and money to get it inspected.  
SHIPOWNER: Right, so I'll send it for the emigrant trip without bothering with an inspection first.  
SON: Wait a second – what reason do you have to do that?  
SHIPOWNER: It's seaworthy!  
SON: I granted that: after all, it's your reason for believing that it is a waste of time and money to get it inspected. I just don't see what reason you have to send it out without an inspection.  
SHIPOWNER: I don't understand. The ship is seaworthy, and as you admit, if it is seaworthy, then the inspection is a waste of time and money, as the inspection will simply come back "seaworthy" and we will be out the fee for the inspection. So, this is why its being seaworthy is a reason I have to send it out without inspection.

Here, again, the third comment from the son seems bizarrely uncomprehending and the shipowner's final remark clarifies exactly why, if the ship's being seaworthy is a reason the shipowner has to believe it's a waste of time and money to get it inspected before sending it out, then its being seaworthy is also a reason the shipowner has to send it out without inspection. A similar dialogue for the ice case would lead to the same conclusion. It would be bizarre to say or think, "I grant that the ice's being thick enough to hold you is a reason you have to believe you'll be fine and save some time to cross the frozen pond rather than walk the long distance around, but I just don't see what reason you have to walk across." The proper response is: "I just explained why it is a reason I have: since it is a reason I have to believe that I'll be fine and save time to walk around, it is a reason I have to walk across rather than around."

These reflections support our **third premise**: if  $p$  is warranted enough to be a reason you have to believe any proposition suggested by  $p$ , then  $p$  is warranted enough to be a reason you have to perform actions suggested by  $p$ .

Reasons to perform a given action provide at least some support for performing that action, but they needn't *justify* you in performing the action. Perhaps the reason is only *pro tanto*, that is, only weighs partially in favor of performing the action. So, if you are prone to dangerous falls on ice, the fact that it is thick enough to hold you (and is shorter across than around) is only a *pro tanto* reason to cross it, and can be

outweighed by a reason not to cross it, that you're prone to disastrous falls on ice. One thought, then, is that knowing something to be true does make it a reason you have for action, but this reason can be defeated if the practical stakes are high enough (and your warrant for the reason is less than absolutely conclusive). So, perhaps when you read that the ship is seaworthy in the nautical museum, you know it is seaworthy and this can be a justifying reason to do various things in that low-stakes situation, such as admire it; whereas, if you were in the shipowner's situation, evidence of that same quality would give you less usable knowledge – you would know that the ship is seaworthy and you would thereby have a reason to send it out on the emigrant voyage but this reason, because of the stakes involved, wouldn't justify you in sending it out, because it would be defeated by something having to do with the chance of error together with the high stakes involved.

To plug this gap, we need to argue for a principle about reasons for action, and this will be our **fourth premise**: if  $p$  is warranted enough to be a reason you have for performing an action suggested by  $p$ , then it is warranted enough to justify you in performing that action.

The shipowner case seems like a perfect test case. The shipowner, assume, has a reason to send out the ship, namely that the ship is seaworthy. The stakes are high, though. Perhaps this is a case in which this reason is defeated by factors having to do with the fact that there is a chance that the ship isn't seaworthy, and that if it isn't, horrible results might ensue by sending it out on the seas with many emigrants in it. Perhaps, that is, that the ship is seaworthy is outweighed or defeated by the *serious risk* that the ship isn't seaworthy. Suppose that this is so: that the serious risk that it isn't seaworthy outweighs or defeats its being seaworthy, with the result that the shipowner isn't justified in sending out the ship. Given all this, it ought to be perfectly acceptable to weigh the two reasons explicitly and to conclude that the serious risk reason wins out. But consider statements expressing such weighing:

There's a serious risk that the ship isn't seaworthy, so that's a reason I have not to send it out. But the ship is seaworthy, so that's a reason I have to send it out. Which is more important, the serious risk that it isn't seaworthy or the fact that it is seaworthy? The serious risk that it isn't. So, I shouldn't send it out.

This speech sounds absurd, not merely absurd to say but absurd to think. It's not that it is absurd to weigh reasons having to do with facts with reasons having to do with chances. It's fine to weigh the cumbersomeness of the umbrella (a reason having to do only with facts) with the chance that it will rain. What's absurd is to weigh a proposition against the serious risk that that very proposition is false. In fact, it's absurd to weigh propositions *about chances* against the serious risk that those very propositions are false: "On the one hand,  $p$  might not be true. On the other hand,  $p$  is true. Which is more important?" Whatever the content of the proposition is, it seems absurd to weigh it against the serious risk that it is false.

One might think that this sounds strange because it is so obvious that the serious risk reason wins out. But this won't do, for two reasons. First, it *isn't* obvious that, if you could have both reasons at once, the serious risk reason would win. What we care about fundamentally is what will happen, not what is likely to happen. We care about actual results first and expected results derivatively. Second, even supposing that you could

have both reasons and that the serious risk reason would win, this wouldn't explain why the weighing statements sound absurd. It should rather sound all too obvious – pedantic – to make or think the weighing statements. A pedant isn't being absurd.

If Premise 4 is false, we ought to expect the weighing statements to be perfectly acceptable, at least in high stakes cases, but they are not. They seem absurd. This is grounds for thinking Premise 4 is true.

We want to address a remaining worry one might have about Premise 4. If this premise is true, the weighings should sound absurd, and they do. But why is it, then, that we sometimes find ourselves saying the likes of “it is seaworthy, but I'd better not send it out, just in case” and “the ice is thick enough to hold me, but I'd better not walk on it, just in case”? Don't these statements – call them “yes, but” statements – suggest Premise 4 is false? There is a puzzle here. The absurdity of the weighing statements is reason to think Premise 4 is true, and the fact that we sometimes assert and think “yes, but” statements seems to be a reason to think Premise 4 is true; and *yet* the appropriateness of “yes, but” statements seems to imply the appropriateness of the weighing statements!

We think the puzzle is best resolved by taking the “yes, but” statements to express vacillation or two-mindedness. Notice how odd it would be, and closer to a problematic weighing statement, after saying, “it is seaworthy, but I'd better not send it out, just in case” to follow-up with “and, again, it is seaworthy; I realize that.” Why would we go in for “yes, but” statements, though? What is the point of expressing vacillation in this way? We offer some speculative guesses. For one thing, it might be useful to express our inner conflict to others and to ourselves, and “yes, but” statements allow this. For another, we might not merely wish to express vacillation; rather, we might want to “try out” a possible commitment to the “yes” part, either to reassure ourselves or in figuring out whether it's really so, but then upon finding that it doesn't stick, that it feels wrong, we back out of this commitment by asserting the “but” part. Finally, we might be anticipating a sort of regret for playing it safe and finding out that the precautions weren't necessary, that all would have turned out all right. Later, we might say to ourselves, “I knew it was all going to be fine. I shouldn't have been so worried.” In saying, “the ship is seaworthy, but I'll not send it out just in case,” one might well be anticipating a future regret for playing it safe. These are speculative explanations. What seems to us hard and fast is that if the “yes, but” statements expressed single-mindedness about the issue at hand, the weighing statements would not be absurd, which they clearly are.

Clifford's Link follows from our four premises. Suppose you know that  $p$ . Then by Premise 1,  $p$  is warranted enough to be a reason you have to believe some propositions suggested by  $p$ . By Premise 2, it follows that if you know that  $p$ ,  $p$  is warranted enough to be a reason you have to believe any proposition suggested by  $p$ . By Premise 3, which connects reasons for belief with reasons for action, it follows that if you know that  $p$ , then  $p$  is warranted enough to be a reason you have for actions suggested by  $p$ . Finally, given all this, Premise 4, which takes warrant enough to be a reason to be sufficient for warrant enough to justify, enables us to conclude that Clifford's Link is correct: that is, if you know that  $p$ , then  $p$  is warranted enough to justify you in actions suggested by  $p$ .

Supposing that the level of warrant sufficient to justify a  $p$ -suggested action can vary with stakes, a plausible fallibilist assumption, then if we accept Clifford's Link we will also have to admit that knowledge varies with stakes. And as we noted above, this is a

counterintuitive claim. There are two possibilities we want to mention, briefly, about how one might cope with this counterintuitiveness. First, one might wish to open up the “philosophy of language toolbox” to explain the counterintuitive results. It is not implausible to think that, while the core relation of knowledge to the practical is stated in Clifford’s Link, it is sometimes useful to apply standards appropriate to one party’s situation more generally. Thus, if what I care about in my situation is using an informant who can give me information I can act on, I might not care about the informant’s stakes; so, I might not call the informant a knower unless I can take what she tells me as a reason for action in my situation. One could then say, if one wished, that my knowledge attribution ‘S knows that  $p$ ’ is true in my context of speech only if the warrant for  $p$  that S possesses is strong enough to make  $p$  a reason for action in my stakes situation. We cannot discuss all the details here of working out such a proposal here.<sup>6</sup> A second possibility is that, after all is said and done, you just might not be able to accept that knowledge can vary with changes in stakes. This would be a reason for thinking, however surprisingly, that Premise 1 is false. However, Premises 2 through 4 say nothing about knowledge. And they themselves jointly entail, given suitably fallibilist assumptions, that something of real epistemological importance can vary with changes in stakes, namely the status of having a proposition as a reason for belief. If we are right, then, there are compelling grounds for thinking that something of epistemological importance depends on practical matters – if not knowledge, then reasons for belief.<sup>7</sup>

## Notes

- 1 This is a case in which something is warranted enough to justify an action but does not justify it, because it isn’t a good reason for the action. In general, being warranted enough to justify a certain action doesn’t suffice for justifying that action, just as being old enough to be the US president doesn’t suffice for being the US president. But if  $p$  is warranted enough to justify you in an act  $\phi$ , then any shortcomings in your warrant for  $p$  do not stand in the way of  $p$  justifying you in  $\phi$ -ing.
- 2 What is it for a belief that  $p$  or action  $\phi$  to be suggested by  $p$ ? In the case of belief, we can think of it as a matter of there being a good argument of the form “ $p$ , therefore  $q$ .” In the case of action, we can think of  $p$  being such as, if it were true, to make a good case for  $\phi$ -ing. Thus, the shipowner, whether he thinks he knows or not, might say “well, if it is seaworthy, that’s a reason to send it out.” Here the shipowner is not saying he *has* a good reason to send it out, but only that if a certain fact obtains then there is a reason to send it out.
- 3 Pragmatic encroachment is defended by, among others, Fantl and McGrath (2002, 2009), Hawthorne (2004), Stanley (2005), and Hawthorne and Stanley (2008). Critics include Brown (2008), Nagel (2008), and Reed (2010).
- 4 Information found at [http://commons.wikimedia.org/wiki/Category:Star\\_of\\_Italy\\_\(ship,\\_1877\)](http://commons.wikimedia.org/wiki/Category:Star_of_Italy_(ship,_1877))
- 5 Except perhaps in cases of reasoning which would be question-begging, such as the famous reasoning, “I have hands, and so I am not a brain in a vat.” When one proposition is warranted enough to be a reason you have to believe another suggested by it, it is a reason you have to believe another suggested by it. Since no worries about question-beggingness arise in all the cases we are considering, we will mostly ignore the difference between “warranted enough to be a reason you have to believe” and “is a reason you have to believe.”
- 6 See Fantl and McGrath (2009, pp. 55–58, 211–212).
- 7 The line of argument developed here is developed in more detail in Fantl and McGrath (forthcoming) and especially Fantl and McGrath (2009, chapter 3).

## References

- Brown, J. (2008) Subject-sensitive invariantism and the knowledge norm for practical reasoning. *Noûs* 42: 167–189.
- Clifford, W. (1886) The ethics of belief. In *Lectures and Essays by the Late William Kingdon Clifford*, F.R.S., ed. L. Stephen and F. Pollock, 339–363. London: Macmillan.
- Fantl, J. and McGrath, M. (2002) Evidence, pragmatics, and justification. *The Philosophical Review* 111: 67–94.
- Fantl, J. and McGrath, M. (2009) *Knowledge in an Uncertain World*. Oxford: Oxford University Press.
- Fantl, J. and McGrath, M. (forthcoming) Pragmatic encroachment: it's not just about knowledge. *Episteme*.
- Hawthorne, J. (2004) *Knowledge and Lotteries*. Oxford: Oxford University Press.
- Hawthorne, J. and Stanley, J. (2008) Knowledge and action. *Journal of Philosophy* 105: 571–590.
- Kvanvig, J. (2004) Pragmatic aspects of knowledge? Blog. March 18, 2008. <http://certaindoubts.com/?p=13> (last accessed March 28, 2013).
- Montmarquet, J.A. (1993) *Epistemic Virtue and Doxastic Responsibility*. Lanham, MD: Rowman and Littlefield.
- Nagel, J. (2008) Knowledge ascriptions and the psychological consequences of changing stakes. *Australasian Journal of Philosophy* 86: 279–294.
- Reed, B. (2010) A defense of stable invariantism. *Noûs* 44: 224–244.
- Stanley, J. (2005) *Knowledge and Practical Interests*. Oxford: Oxford University Press.

# Practical Matters Do Not Affect Whether You Know

*Baron Reed*

One of the central questions that epistemologists have attempted to answer is this: when someone knows something, what explains or grounds that knowledge? Many different answers have been given – for example, clear and distinct perception, evidence, sense experience, reliable belief-producing processes – and the differences between them should not be ignored. But it is interesting to note that all of these answers agree in at least one respect: they are all focused on something that is *truth directed*. For example, evidence is thought to be an indication of what the truth is, reliable belief-forming processes are reliable insofar as they tend to produce true beliefs, and so forth.

Traditionally, this point of agreement was so widely and deeply shared that epistemologists never really thought about it. It has been given a name – *intellectualism* or *purism* – only recently and only by the relatively small number of philosophers who have argued against it.<sup>1</sup> According to these philosophers, knowledge can be only partially explained by (or grounded in) truth-directed things like those mentioned above. A full explanation also has to include the practical stakes of the person in question. In this sense, the pragmatic encroaches on the epistemic.<sup>2</sup>

The thesis of pragmatic encroachment has been defended in three ways. First, it is thought to be the best explanation of how we use knowledge attributions and knowledge denials in ordinary conversation, when we defend, criticize, or excuse actions. Second, it has been argued that pragmatic encroachment is needed to make sense of the way in which we use knowledge in practical reasoning. And, third, pragmatic encroachment is thought to be supported intuitively by consideration of pairs of cases in which subjects who are identical with respect to truth-directed things (like evidence and reliability) differ in their practical stakes and also differ epistemically – that is, one of the subjects has knowledge that the other lacks.

In what follows, I shall present each of these arguments for pragmatic encroachment. I shall then raise several objections to the view, which will encourage a reconsideration of those arguments. Finally, I will sketch an alternative view that better explains the connection between knowledge and practical reason while remaining within the bounds of traditional, truth-directed epistemology.

## 1 The Case for Pragmatic Encroachment

Defenders of pragmatic encroachment typically argue for a principle that links knowledge with practical rationality. For example, John Hawthorne and Jason Stanley defend the following Reason-Knowledge Principle (RKP):

Where one's choice is *p*-dependent, it is appropriate to treat the proposition that *p* as a reason for acting if and only if you know that *p*.<sup>3</sup>

Suppose that I am trying to decide whether to go to the library because I need to meet my friend, Richard. In that sense, my choice of whether to go to the library is dependent on the proposition that Richard will be there. Using this proposition in my deliberation is permissible just in case I know it to be true.

Notice that RKP specifies that knowledge is both *sufficient* and *necessary* for action.<sup>4</sup> That is, when one has the relevant knowledge, one's epistemic position is good enough to make the action in question rational. (This is compatible with it failing to be rational for some non-epistemic reason – for example, the action isn't optimal for attaining one's goal.) And an action is rational only when it is grounded in knowledge. In other words, if one does not have the relevant knowledge, one does not have the epistemic grounding needed to make the action rational.

We are now in a position to fully appreciate the sense in which the pragmatic is thought to encroach upon the epistemic. If RKP, or something like it, is correct, then knowledge is sufficient for practical rationality: if one knows that *p*, then one may treat the proposition that *p* as a reason for acting. There is a logically equivalent way of reading this conditional: one knows that *p* only if one may treat the proposition that *p* as a reason for acting. Read in this way, it makes clear that practical rationality is necessary for knowledge. If our concern is to provide the correct analysis of knowledge, then, we have to include in the list of necessary conditions, not only things like truth, belief, and justification, but also suitability for use in practical reasoning. On this view, the practical is embedded in the very nature of knowledge.



## 1a Practical talk about knowledge

We make claims about knowledge – attributing or denying it, to ourselves and to others – for a variety of reasons. Consider the following examples.

- (1) Maria knows when World War I occurred.
- (2) Ask Ronald – he knows where I parked the car.
- (3) Why did you lock the door? You knew I forgot my key.
- (4) I knew they wouldn't mind sitting on the balcony.
- (5) I didn't know you were allergic to peanuts.

Sometimes, as in (1), our aim is merely to describe reality. Maria's knowing when World War I happened is a fact in just the same way it is a fact that the Amazon is the biggest river in the world. But there are also many uses of knowledge claims that serve a more obviously practical purpose. For example, the speaker of (2) is flagging Ronald as a good source of information, given the hearer's practical needs.

In other instances, the connection between knowledge and action is more intimate still. Some of the most interesting practical uses of knowledge claims are to criticize, defend, or excuse actions, as in (3), (4), and (5), respectively. The thought is that, if RKP were true, we would have an explanation for these natural ways of talking. In (3) and (4), the speaker is attributing knowledge to the subject in question. In doing so, she has indicated that the subject has no epistemic obstacle to performing the relevant action. If the action is also the one that best serves the subject's ends, it is rational for the subject to perform the action – for example, buying tickets for the balcony, as in (4) – and a failure of practical rationality if she does *not* perform the action – like failing to leave the door unlocked, as with (3). In both of these cases, the speaker appears to be relying on the (alleged) fact that knowledge is sufficient for rational action.

A speaker can also criticize a subject for acting with an inadequate epistemic grounding – for example:

- (6) How could you leave the kids home alone? You didn't know I would be home early.

In this sort of case, the subject lacks the knowledge that is (purportedly) necessary for rational action. Finally, a speaker can also excuse behavior – her own or someone else's – by pointing out that the person in question did not have the requisite knowledge. With respect to (5), for example, the speaker is closing out the possibility of criticism – like that made in (3) – by asserting that she did not have the knowledge she would have needed in order to make rational a different action (in this case, offering a different dish to her guest). Not only would the missing knowledge have been sufficient to rationalize a different action, it is also necessary to make rational the choice of a different course of action.

These common ways of talking about knowledge are offered as evidence for pragmatic encroachment because they seem to presuppose the truth of RKP. Why would we focus on what the agent knows or doesn't know? It makes sense to do so on the assumption that knowledge is sufficient for practical rationality.<sup>5</sup>

## 1b Knowledge as a safe reason

Jeremy Fantl and Matthew McGrath offer an argument for pragmatic encroachment that is grounded in the way we conduct practical reasoning. One of the two premises they employ is *Safe Reasons*:

If  $p$  is a reason you have to  $\phi$ , then  $p$  is warranted enough to justify you in  $\phi$ -ing, for any  $\phi$ .<sup>6</sup>

The basic idea is that, when we weigh the reasons we have (for acting or for forming beliefs), we regard them as safe to rely upon. We do not take into account how probable they are – or, if we do come to consider their probability, we no longer regard them as reasons we can rely upon. As evidence for this claim, Fantl and McGrath point out that we never weigh against one another reasons like “the horse I bet on won the race,” and “there’s a chance that the newspaper reported the race results incorrectly and the horse didn’t win.” If I am trying to decide whether I can afford to eat a fancy dinner tonight, I simply take it as a fact that I have the money from betting on the winning horse. (Or, if I am genuinely concerned that the race report is incorrect, I don’t treat this as a reason at all.) I would never think, “Well, on the one hand, I have the money from winning at the track. On the other hand, I may not have that money, since there’s a chance that my horse didn’t really win. What should I do?” Instead, “[o]ne reason is defeated by another, and the defeat has nothing to do with how probable either reason is.”<sup>7</sup> Fantl and McGrath describe this as a “ledger-keeping” picture of having reasons: when an agent has a reason, “it gets put in the ledger with countervailing reasons and *weighed* against them. But the probabilities of these reasons don’t get recorded alongside.”<sup>8</sup>

The other premise in their argument is the *Knowledge-Reasons* Principle, KR:

If you know that  $p$ , then  $p$  is warranted enough to be a reason you have to  $\phi$ , for any  $\phi$ .<sup>9</sup>

Not just any proposition can serve as a reason. It wouldn’t after all, be practically rational to buy a particular book simply because one has guessed that it will be assigned by one’s instructor. It makes sense to buy it only if one has the proper epistemic relationship to that proposition. What *is* the proper epistemic relationship? KR identifies knowledge as being good enough epistemically to provide us with reasons, both for acting and for forming beliefs.

Together, *Safe Reasons* and KR entail that, if you know that  $p$ , then  $p$  is warranted enough to justify you in  $\phi$ -ing, for any  $\phi$ .<sup>10</sup> That is, knowledge is sufficient for practical rationality. Again, it is logically equivalent to say that being suitable for use in practical reasoning is necessary for knowledge.

## 1c High stakes and low stakes

Perhaps the most common way of defending pragmatic encroachment involves comparing two subjects who are identical in all truth-directed ways but differ in their practical situations. For example, suppose that Jamie hears the waiter say that the cake she has just been served does not contain any nuts. “Too bad,” she thinks, as the waiter returns to the kitchen. Jamie has no reason to distrust the waiter, and she comes to

believe what the waiter has said. Most philosophers would agree that the following assertion is true:

- (7) Jamie knows the cake does not contain nuts.<sup>11</sup>

Sasha, who has been served a piece of the same cake, also overhears the waiter. But her reaction is quite different from Jamie's – Sasha is extremely allergic to nuts and could die if she eats any. Many, perhaps most, people think that Sasha should not simply start eating the cake. Given the risk of dying, the rational thing for her to do is to check with the waiter and have him confirm the recipe with the chef. But surely, the defenders of pragmatic encroachment say, this would be unnecessary if Sasha really *knew* the cake was nut-free. For that reason, they think, most speakers would not assert:

- (8) Sasha knows the cake does not contain nuts.

More strongly, some speakers may even deny (8) or assert its contradictory. Notice, however, that Sasha and Jamie are identical in all truth-directed ways. They have the same evidence that the cake does not contain nuts and, let us suppose, their relevant cognitive faculties – perception, memory, reason, and so on – are equally powerful and reliable. But they differ in that Sasha has much more at stake practically than Jamie does. Practical rationality seems to require Sasha (but not Jamie) to confirm that the cake doesn't have nuts in it. A natural explanation is that Sasha needs to put herself in a stronger position than Jamie's, with respect to truth-directed things like evidence, in order to have the knowledge that Jamie has. It is only when her epistemic position is strong enough, given her practical interests, that her belief counts as knowledge and she is permitted to treat it as a reason to act.

## 2 Objections to Pragmatic Encroachment

There are a variety of objections that can be raised to the thesis of pragmatic encroachment. Some offer counterexamples to RKP. Others argue that pragmatic encroachment fails to capture something essential about knowledge. And, finally, others argue that the view gives an inadequate or incorrect picture of practical rationality.

### 2a The variety of ways we criticize, justify, and excuse

Recall that argument 1a for pragmatic encroachment is grounded in our practice of using knowledge attributions to criticize, justify, and excuse actions. The thought was that there must be a unique and special link between knowledge and practical rationality to explain this sort of behavior. In fact, however, criticism, justification, and excusing are connected to a variety of epistemic properties, both weaker and stronger than knowledge. Here are some examples:

- (9) When I wrote my letter of recommendation, I had every reason to believe she would graduate with honors.

- (10) How could you go rock climbing on Widow's Peak without being completely certain that your gear was in good shape?
- (11) I've done a bit of research on this, so I am quite sure that vaccines are not linked to autism.

In some cases, we even defend or excuse our actions by talking about what we believed, without any explicit mention of the epistemic properties of those beliefs:

- (12) I believed him when he said he wasn't married.

Given this variety of ways in which we criticize, justify, and excuse behavior, there is no reason to think that knowledge bears a special relationship to practical rationality.

These data also give us reason to deny RKP, with respect to both the necessity and the sufficiency of knowledge for rational action. Take (9), for example. In defending his letter of recommendation, the professor is letting his interlocutor know that it was practically rational for him to write the letter, even though he didn't know that the student in question would graduate with honors. (He couldn't know this because, as it turned out, it was false.) In other words, his action was rational even though it was grounded in a reason that failed to be knowledge. Hence, knowledge wasn't necessary to make it rational.

Consideration of (10) makes clear that knowledge isn't always sufficient for rational action, either. Consistent with making this assertion, the speaker might grant that the climber knows her gear is in good shape. The problem is that mere knowledge – an epistemic status that falls short of certainty – might not be good enough when someone's life is at risk.

## 2b Crazy counterfactuals

One of useful things about epistemic concepts is that they allow us to think about how our cognitive position could be different from what it is. For example, if I have heard vague remarks about a presidential candidate's views on immigration, I can acquire knowledge of his views by reading newspaper articles about them. Doing so will allow me to gather relevant evidence. I can also recognize that, if I hadn't been watching television, I wouldn't have had any evidence at all about the candidate's views.

There is nothing at all surprising about this. It should be obvious that changing one's standing with respect to truth-directed things like evidence will also change the epistemic properties of one's beliefs. This is consistent with pragmatic encroachment. However, it is a consequence of RKP that changing one's practical interests will also change the epistemic properties of one's beliefs. The higher the stakes, the more justified one's belief will have to be to count as knowledge. For this reason, assertions like the following are true, given RKP:

- (13) If she really loved me, she wouldn't know that I cheated on her.
- (14) If I didn't have so much money riding on the game, I would know the Bears had won.

Defenders of pragmatic encroachment have acknowledged that counterfactuals like these will be true, if their theory is correct.<sup>12</sup> Being invested in a relationship or gambling

large sums of money make it the case, given RKP, that the subject has to do much more to have knowledge than someone who does not have anything at stake in the situation. This is a serious problem: not only is it extremely implausible that these counterfactuals are true, but it is also clear that people never assert anything like them. If pragmatic encroachment is motivated, in part, by paying careful attention to the way people use attributions of knowledge, then surely we need to consider the way they use counterfactual knowledge attributions, too.

## 2c Knowledge and high stakes

Perhaps the most straightforward objection to pragmatic encroachment is grounded in counterexamples to principles like RKP. Here is one such case:

I am taking part in a psychological study that tests the effects of stress on memory. I am asked a question: when was Julius Caesar born? If I give the correct answer, I get a jelly bean. If I give an incorrect answer, I am given a horrible electric shock. Nothing happens if I give no answer. I remember that Caesar was born in 100 BC, but I am not so sure of it that it is worth risking electrocution. Nevertheless, I quietly say to myself, "I know it's 100 BC."<sup>13</sup>

Notice that in this scenario it is plausible that I retain my knowledge even though the practical stakes are so high it isn't rational for me to act on it. If this is right, it is a case in which knowledge isn't sufficient for rational action. Notice, too, that it is natural for me to continue to attribute knowledge to myself, even when I recognize that I shouldn't treat it as a reason to act.<sup>14</sup>

## 2d The relativity of knowledge

The defenders of pragmatic encroachment sometimes claim that their view has an advantage in being able to fully capture the value or importance of knowledge.<sup>15</sup> Showing how knowledge has an intimate link to practical rationality explains why it is fitting for knowledge (rather than, say, justification) to be the central focus of epistemology. But the sort of objection raised in section 2c can be extended to show that pragmatic encroachment forces us to view knowledge in a way that actually undermines our sense of its importance:

The psychological study changes so that I am playing two games at the same time. The first is as before: I'll get a jelly bean for a correct answer and a horrible electric shock for one that is incorrect. But, in the second game, I get \$1000 for a correct answer and only a gentle slap on the wrist for a wrong answer. In both games, nothing happens if I fail to answer, and I can take different strategies in the two games. Both games begin when I am asked: when was Julius Caesar born? I give no answer in the first game, but I answer "100 BC" in the second.<sup>16</sup>

In this scenario, I am faced with two decisions, both of which are *p*-dependent, where that *p* is the proposition that Julius Caesar was born in 100 BC. But I am rational to treat that *p* as a reason for only one of the decisions. If RKP is true, this means that I both know and don't know that *p* at the same time.<sup>17</sup>

The only way an object can have contradictory properties at the same time is if they are relativized – for example, to time or to place. In this case, I both have and lack the same knowledge in the same place and at the same time. It would seem that the only thing left to which they might be relativized is something like *practical situation*. This would be an unwelcome consequence for several reasons. First, we do not ordinarily talk as though knowledge is relativized in this way. Second, our account of how to reason with knowledge would become rather complicated, as we would have to be sure that pieces of knowledge combined in deliberation have not been relativized in incompatible ways. And, finally, knowledge would no longer seem to be the stable accomplishment it is usually thought to be. Not only would it “come and go with ease,” as Hawthorne suggests, it would also be confined to practical situations in a way that would make it difficult to rely upon.<sup>18</sup>

## 2e Safe reasons and the risk of error

Recall that Fantl and McGrath offer what they call a “ledger-keeping” picture of reasoning, according to which we weigh reasons against one another without taking into account their probabilities. This picture is suggested by the fact that we do not weigh that *p* against the chance that not-*p*. They conclude that our reasons are safe: they are warranted enough epistemically to justify us in anything we do or believe. They then argue that knowledge provides us with reasons that are safe in this sense.

If fallibilism is correct, however, knowledge is distinguished from certainty in that knowledge carries with it the risk of error.<sup>19</sup> When I see a medium-sized black bird in a nearby tree, that’s good evidence for me to think it is a crow. A fallibilist will go further and say that the evidence is good enough for me to *know* that it’s a crow. Still, the evidence I have is logically compatible with it being something else – an escaped mynah, perhaps, or merely a figment of my imagination.

Ordinarily, we ignore the chance of error because it is quite low when we have knowledge. But there are times when it becomes relevant after all. When we combine many fallibly known premises in a single piece of reasoning, the risk of error rises until we would be irrational to continue ignoring it. This is something that Fantl and McGrath acknowledge; for this reason, they say, “no fallibilist can say that when you have a reason, you get to put it to work in *any* reasoning in which it figures.”<sup>20</sup> But this admission does not fit well with the ledger-keeping picture of reasoning. If knowledge really did provide safe reasons, then it *ought* to be something that we could put to work in any reasoning at all.

Fantl and McGrath are right to point out that we do not weigh against one another that *p* and the chance that not-*p*, but they draw the wrong conclusion from it. When we begin to worry about the chance that not-*p*, we stop treating that *p* as a reason. (We might substitute for it the probability that *p*, but that is a different fact.) So, it’s not that knowledge provides us with a reason that is always safe – rather, we can use knowledge as a reason only when it is safe to do so. As we saw in section 2c, there can be cases in which knowledge is not good enough to act on because the agent has stakes that are too high for this to be rational. To this we can also add cases in which the stakes are low but the risk of error has simply grown too great through the use of too many fallibly known premises. The upshot in either sort of case is the same: the thesis of pragmatic encroachment does not capture how we reason with knowledge because knowledge is not always good enough for practical rationality.

## 2f The Dutch book

If RKP were true, it would license some highly irrational behavior. Consider the following case:

I have a broker who is extremely reliable at picking stocks. She tells me that a biotech stock, BXD, is a good long-term investment and that she can move a fourth of my assets into BXD stock. Given her testimony, I know that it will go up in value, so I agree. An hour later, she tells me that she can now move another fourth of my assets into BXD. I know it will go up in value, so I agree again. An hour later, the same thing happens. When she calls me for the fourth time, she offers to move my remaining assets into BXD. But she also points out that I would then have all of my assets tied up in a single stock, which is a very risky thing to do. The stakes have become too high, and so it's not rational for me to buy more shares of BXD. Given RKP, this means that I no longer know that the stock will go up in value. I no longer have the knowledge that would permit me to keep the stock, so I tell my broker to sell all of it in favor of other investments I know to be safe. After an hour, she calls back to remind me that BXD is an excellent long-term investment. Having sold all my shares, this is no longer a high-stakes proposition for me. I reflect that she is reliable in her stock tips, and I again come to know that BXD will go up in value. So, I take her up on her offer to move a fourth of my assets into BXD. And so on.<sup>21</sup>

This is a bad deal for me. Because my broker charges a fee for every transaction, she has set up a Dutch book against me: my money is guaranteed to melt away.

Of course, no one would continue making trades once it has become clear that losses are guaranteed. The problem for pragmatic encroachment, though, is that RKP permits every step and never gives me a reason to stop. I can get out of the Dutch book only by ignoring RKP. That's good reason to think the principle provides us with an incorrect account of knowledge and practical reason.

## 2g Putting the cart before the horse

If RKP were true, it would mean that knowledge and action could interact with one another so that practical deliberation turns out to be incapable of reaching a stable end point. The agent's knowledge might license action that raises the stakes in such a way that the agent ceases to have the knowledge in question. The Dutch book objection shows this occurring over a sequence of events, but it can also happen in a way that prevents the agent from taking any action at all. Consider the following dialogue:

*ME:* If BXD will go up in value, I should invest in it. So, should I invest in it?  
*FRIEND:* That depends – do you know it will go up in value?  
*ME:* That depends – should I invest in it?

If I don't invest in BXD, I have nothing at stake relative to knowing whether it will go up in value. So, this is relatively easy for me to know. But, if I invest heavily in BXD, the stakes may go so high that I can't know its value will increase.

Presumably, the defenders of pragmatic encroachment put forward a principle like RKP because they think we can discover what the practically rational thing to do is by using knowledge in deliberation. But, if RKP were true, it would mean that the reverse

is true: we can't tell what our knowledge is until we have first determined what the practically rational thing to do is.<sup>22</sup> But this is to put the cart before the horse. Moreover, it is hard to see how knowledge would retain any importance at all – once we have made an independent determination of what the practically rational thing to do is, knowledge has no role left to play.

### 3 An Alternative Proposal

Even if the above objections show that pragmatic encroachment cannot be correct, it doesn't yet mean that a traditional, truth-directed epistemology can provide a plausible account of knowledge and practical rationality. In particular, we would like to have an explanation of the intuitive data (from section 1c) regarding the different ways we sometimes talk about people who differ only with respect to their practical stakes.

Toward this end, notice that justification comes in degrees, ranging downward from perfect justification. If fallibilism is correct, there is some degree of justification lower than perfect justification that serves as a threshold for knowledge. In other words, knowledge is compatible with having a variety of degrees of justification, ranging from the threshold to perfect justification. Corresponding to these different degrees of justification, we can also say that there are different degrees of knowledge.

If we take knowledge to rest on truth-directed things like evidence or reliability, then Jamie and Sasha, the two subjects we compared in section 1c, both know that the cake doesn't have nuts in it. Because we reject RKP – we don't hold that knowledge is always sufficient for rational action – we can say that only Jamie is rational to act on her knowledge. Sasha needs a higher degree of knowledge, perhaps something approaching certainty, before it is rational for her to eat the cake.

That explains the behavior of both subjects. Advocates of pragmatic encroachment, however, will argue that the intuitive data include, not only how the subjects behave, but what we say about them. Ordinary speakers will assert (7) but not (8) – in fact, some people will even positively deny (8). On the view I have sketched, it seems that denial should be false. So, why would speakers say such a thing?

The simplest explanation is that “knows” is ambiguous: corresponding to the various degrees of knowledge are various meanings of the term. Each meaning is available for use in any conversational context; what the term ends up meaning on any occasion of use is determined only by the speaker's intentions. “Knows” is also strongly associated with a threshold usage, where the threshold can be set either at the minimal threshold to count as knowledge at all or at some higher degree of knowledge.<sup>23</sup> This threshold will also be set by the speaker's intentions alone, where the speaker will usually intend to talk about the degree of knowledge that is currently practically useful. Because this is the usual practice, one's interlocutors will expect one's knowledge attributions to be about only practically relevant degrees of knowledge. So, when a speaker denies (8), her interlocutor will naturally hear the speaker as denying that Sasha has a *high* degree of knowledge – not necessarily as denying that Sasha has knowledge of *any* degree.

It is worth noting that this expectation can be ignored; speakers can choose to talk about practically useless degrees of knowledge, too. Hence, in the psychological study case in section 2c, I could say – both to myself and as an unofficial side remark to the researcher – that I know the answer even though I don't want to make an official



response to the question. This, too, is an intuitive datum about how we think and talk about knowledge, and it is one that the defenders of pragmatic encroachment simply cannot accommodate.<sup>24</sup>

Of course, much more could be said about these and other, related issues. But it should be clear enough that practical matters do not affect whether you know. The thesis of pragmatic encroachment simply faces too many debilitating objections to be correct. At most, then, practical matters affect what degree of knowledge it is rational to use in deliberation and it is most useful to talk about. Knowledge itself is still best understood in accordance with traditional, truth-directed epistemology.

## Acknowledgments

For helpful discussions of these issues, I am grateful to Jessica Brown, Jeremy Fantl, Matthew McGrath, Ernest Sosa, Sarah Wright, and, most especially, Jennifer Lackey.

## Notes

- 1 For “intellectualism,” see Stanley (2005); for “purism,” see Fantl and McGrath (2009). See also Fantl and McGrath (2002), Hawthorne (2004), and Hawthorne and Stanley (2008).
- 2 Fantl and McGrath (2009) characterize the view as *pragmatic encroachment*.
- 3 Hawthorne and Stanley (2008, p. 578).
- 4 Fantl and McGrath (2009, p. 66), defend a somewhat different principle, KJ: If you know that  $p$ , then  $p$  is warranted enough to justify you in  $\phi$ -ing, for any  $\phi$ . Notice that KJ takes knowledge to be sufficient but not necessary for rational action. Where it is relevant to an argument or objection, I will indicate this difference between RKP and KJ. Also, Fantl and McGrath do not limit their principle to choices that are  $p$ -dependent, as Hawthorne and Stanley do with RKP. As Jessica Brown notes, this means that the subject will need to have an extraordinarily high degree of epistemic justification for all of her beliefs; see Brown (2011, p. 170). It is hard to see how KJ, in its unrestricted form, would not lead to almost total skepticism.
- 5 For this sort of argument, see Fantl and McGrath (2007) and (2009, p. 63), and Hawthorne and Stanley (2008, pp. 571–574).
- 6 Fantl and McGrath (2009, p. 77).
- 7 Fantl and McGrath (2009, p. 79).
- 8 Fantl and McGrath (2009, p. 79).
- 9 Fantl and McGrath (2009, p. 69).
- 10 This is KJ (see footnote 4 above). Remember that it does *not* take knowledge to be necessary for practical rationality, unlike Hawthorne’s and Stanley’s RKP.
- 11 There is a complication here. Contextualists would say that the truth of (7) depends on the epistemic standards in play in the speaker’s conversational context. Defenders of pragmatic encroachment have argued forcefully against contextualism and, though this debate is far from over, I shall simply follow them here in holding that the truth of knowledge attributions depends on what is happening with the subject of the attribution, not with the speaker.
- 12 See Hawthorne (2004, p. 177, n.40), and Fantl and McGrath (2009, pp. 208–212). Both Hawthorne and Fantl & McGrath suggest that there are also crazy counterfactuals that are true in virtue of the Gettier problem, so this is not a problem unique to pragmatic encroachment. Hawthorne, for example, says that the counterfactual (said of a person looking at a real barn but in an area where there are many barn facades), “If there weren’t so many fake barns

- around, he would know that he's looking at a barn," sounds odd. I am inclined to say that it sounds *less* odd than the pragmatic encroachment counterfactuals. More than that, any discussion of the Gettier problem – couched in counterfactuals or not – will sound odd in ordinary discourse, given that it is a technical problem in epistemology. The same cannot be said about the pragmatic counterfactuals, which one would think ought to play a role in ordinary deliberation, given how frequently they would be useful if true.
- 13 This case is drawn from Reed (2010). See also Brown (2008).
  - 14 Fantl and McGrath concede that it might be natural to continue to attribute knowledge to oneself in this scenario, but they also argue that "it wouldn't raise eyebrows" if the subject denied that he had the knowledge in question (2009, p. 62). That may be true, but it doesn't change the fact that the case as described (in which the subject *can* continue to attribute knowledge to himself) is straightforwardly a counterexample to pragmatic encroachment.
  - 15 See, for example, Stanley (2005, p. vi), and Fantl and McGrath (2009, chapter 6).
  - 16 See Reed (2010).
  - 17 The situation is not quite as clear for Fantl and McGrath's KJ, given that it does not take knowledge to be necessary for practical rationality. For more on the implications of this objection for KJ, see Reed (2012).
  - 18 Hawthorne (2004, p. 176).
  - 19 More precisely, if fallibilism is correct, knowledge carries with it the risk of being either false or true only by accident. See Reed (forthcoming a) and (2002).
  - 20 Fantl and McGrath (2009, p. 82).
  - 21 See Reed (2012).
  - 22 Recall that, if RKP is true, being suitable for use in practical reasoning is a necessary condition on knowledge.
  - 23 "Knows" has this threshold usage because we often are interested in whether someone knows *well enough*, given the practical circumstances.
  - 24 For a much more detailed presentation of this view, see Reed (forthcoming b).

## References

- Brown, J. (2008) Subject-sensitive invariantism and the knowledge norm for practical reasoning. *Noûs* 42: 167–189.
- Brown, J. (2011) Fallibilism and the knowledge norm for assertion and practical reasoning. In *Assertion*, ed. J. Brown and H. Cappelen, 153–174. Oxford: Oxford University Press.
- Fantl, J. and McGrath, M. (2002) Evidence, pragmatics, and justification. *Philosophical Review* 111: 67–94.
- Fantl, J. and McGrath, M. (2007) On pragmatic encroachment in epistemology. *Philosophy and Phenomenological Research* 75: 558–589.
- Fantl, J. and McGrath, M. (2009) *Knowledge in an Uncertain World*. Oxford: Oxford University Press.
- Hawthorne, J. (2004) *Knowledge and Lotteries*. Oxford: Clarendon Press.
- Hawthorne, J. and Stanley, J. (2008) Knowledge and action. *The Journal of Philosophy* 105: 571–590.
- Reed, B. (forthcoming a) Fallibilism. *Philosophy Compass*.
- Reed, B. (forthcoming b) Fallibilism, epistemic possibility, and epistemic agency. *Philosophical Issues*.
- Reed, B. (2012) Resisting encroachment. *Philosophy and Phenomenological Research* 85: 466–473.
- Reed, B. (2010) A defense of stable invariantism. *Noûs* 44: 224–244.
- Reed, B. (2002) How to think about fallibilism. *Philosophical Studies* 107: 143–157.
- Stanley, J. (2005) *Knowledge and Practical Interests*. Oxford: Oxford University Press.

## Chapter Five

# Can Skepticism Be Refuted?

A standard form of skepticism rests on the premise that the choice between our ordinary beliefs and skeptical hypotheses is evidentially underdetermined. To make a case against this premise, Jonathan Vogel invokes inference-to-the-best-explanation and considers what he calls the “isomorphic hypothesis,” which matches the cause and effect relationships of the real world hypothesis, according to which things are by and large what we take them to be. The isomorphic hypothesis’s explanatory power seems to be no less than that of the real world hypothesis. On what grounds, then, can we prefer the latter over the former? Vogel argues that the isomorphic skeptical hypothesis does not deepen our understanding of our perceptual experiences. The isomorphic hypothesis is more complex than the real world hypothesis because it replaces genuine shapes with pseudo-shapes and genuine locations with pseudo-locations. This increased complexity is not offset by a commensurate increase in explanatory scope and depth. Hence, Vogel suggests, we may prefer the real world hypothesis over the skeptical isomorphic hypothesis and thus reject the evidential underdetermination premise on which skepticism relies.

In his response essay, Richard Fumerton explores the challenge of refuting skepticism within the traditional framework of inferential internalism, according to which knowing propositions about the external world requires (a) that we have evidence E making such propositions probable and (b) that we know E makes such propositions probable. For inferential internalists to avoid skepticism, they must meet two serious challenges. First, to avoid an infinite regress, they must show that we can have direct knowledge of what makes what probable. Second, they must identify what body of evidence enables us to know that skeptical hypotheses are false. Does inference-to-the-best-explanation help with the latter challenge? Fumerton expresses doubt about that. Skeptical hypotheses might provide simpler explanations than their non-skeptical counterparts. Moreover, the skeptics have every right to question that inference-to-the-best-explanation is a guide to truth.

# The Refutation of Skepticism

Jonathan Vogel

## I

We take ourselves to know a lot about the world, and it would be profoundly disturbing if we didn't. To deny that we enjoy such knowledge is to endorse *skepticism* about the external world.<sup>1</sup> Skepticism is philosophically important because a gripping line of thought seems to show that it is correct.

The argument that supports skepticism is one of the most famous in the history of philosophy.<sup>2</sup> It turns on the possibility that we might be victims of some kind of massive sensory deception. Consider two situations. In one, everything is normal, and you see a bridge. In the other, you are the subject of an awful experiment. Your brain is isolated in a laboratory vat and fed completely delusory sensory inputs. These inputs make it appear to you as though there were a bridge before you, even though there isn't. It may well seem that you have no way of knowing which of these two situations you're in. So, in particular, you don't know that there is really a bridge before you. This line of thinking generalizes. If you don't know that you're not the victim of massive sensory deception, then you're unable to know virtually anything at all about the world. Skepticism prevails.

That is the gist of the "deceiver argument" (as I shall call it), but there are various details that need to be taken into account. I think we can understand better what is at stake by comparing the deceiver argument with an ordinary case in which someone fails to know. Imagine that you're in the kitchen, and the toaster suddenly stops working. It could be that the toaster itself has burned out. However, it could also be that the toaster is all right, but a fuse has blown instead. You are, then, faced with two competing hypotheses as to what caused the toaster to shut off.<sup>3</sup> If you have no further relevant information, you have no basis for accepting one hypothesis over the other. Any choice on your part would be arbitrary. I will say that, in such a situation, your choice of hypothesis is *underdetermined*. If, despite such underdetermination, you guess that the toaster shut off because the fuse blew, you wouldn't *know* that – even if, luckily, your guess turned out to be correct. This sort of example supports the view that knowledge is governed by the following general principle:

*Underdetermination principle* (UP): If *q* is a competitor to *p*, then one can know *p* only if one can non-arbitrarily reject *q*.

Rejection of *q* would be arbitrary in the relevant sense just in case *q* is, from an epistemic standpoint, no less worthy of belief than *p*. So, we have:

*Underdetermination principle* (UP, alternate version): If *q* is a competitor to *p*, then a subject *S* can know *p* only if *p* has more epistemic merit (for *S*) than *q*.

What factors add to or subtract from epistemic merit is a crucial, but controversial, matter. I address it below.

With these points in hand, we can set out the deceiver argument more rigorously, as follows.

- (1) Consider any proposition *m* about the world I ordinarily believe (hereafter, “mundane propositions”). In order to know that *m*, my belief that *m* must not be underdetermined.
- (2) My belief that *m* is underdetermined.
- (3) Therefore, I don’t know that *m*.

Premise (1) may be seen as a straightforward application of the underdetermination principle. According to that principle, underdetermination is inimical to knowledge. Premise (1) makes the point that we fail to know any mundane proposition for which there is a competitor of equal epistemic merit. Premise (2) is the claim that all mundane propositions face a competitor of just that sort. The competitor is that one is the victim of massive sensory deception, as described above (I will call this the “skeptical hypothesis”). The conclusion (3) is that we lack knowledge of mundane propositions in general, which is to say that skepticism about the external world holds. Now, the deceiver argument is logically valid. To refute it, we would have to successfully challenge one of its premises. Premise (1), and the underdetermination principle that underwrites it, seem unassailable.<sup>4</sup> So, a satisfactory response to the deceiver argument will have to show that premise (2) is false, or at least not adequately supported.

But satisfactory to whom? The deceiver argument explicitly or tacitly depends on various epistemic principles. One is the underdetermination principle. In determining whether underdetermination exists, other principles will be brought to bear. Now, I take it that we are committed to a body of epistemic principles that govern what we count as knowledge, justified belief, and the like. The deceiver argument may be construed as (putatively) proceeding from just those principles. Thus understood, the argument is an attempt to show that, by our own lights, we lack the knowledge of the world we think we have. I will call this position *domestic skepticism*. Domestic skepticism is concessive, but dangerous. It is concessive, in that it doesn’t contest the legitimacy of the epistemic principles we embrace and employ. But it is also dangerous, in the sense that it would be deeply unsettling, or worse, if we have no knowledge of the world, according to our own accepted view of what knowledge is and what it requires. On the other hand, if it emerges that our epistemic principles don’t have that consequence after all, domestic skepticism has been refuted.<sup>5</sup>

My primary concern here will be with domestic skepticism. However, there are other forms skeptical thinking might take. Someone might contest not only our ordinary judgment that we have knowledge of the world, but also the legitimacy of the principles on which we rely in making that judgment. I will call this more extreme position “exotic” skepticism. It won’t matter to an exotic skeptic that our knowledge claims accord with our epistemic principles, since the latter themselves are supposed to be questionable in some way. Thus, a reply that answers the domestic skeptic will seem to the exotic skeptic like our pointlessly patting ourselves on the back. I will return to the difference between domestic and exotic skepticism in section III.

## II

The case for skepticism depends on the status of premise (2) of the deceiver argument, and the status of that premise depends on whether mundane propositions have more epistemic merit than the skeptical competitors they face. Philosophical opinion varies widely and sharply on this point. One can distinguish the following positions, among others. (a) We have no basis whatsoever for rejecting skeptical hypotheses, if such hypotheses really do compete for our acceptance. Skepticism is at least “conditionally correct,” and it can’t be refuted.<sup>6</sup> (b) While no particular evidence counts against skeptical hypotheses, epistemic rationality permits us or requires us to reject such hypotheses out of hand. “A reasonable man does not have certain doubts,” as Wittgenstein said (Wittgenstein, 1972, Remark no. 220). (c) What licenses us in maintaining belief in mundane propositions, and rejecting skeptical hypotheses, isn’t a special anti-skeptical epistemic principle, but a broader principle of methodological conservatism. Methodological conservatism is, roughly, the doctrine that we are entitled to maintain beliefs we already have, all other things being equal. Since we have already accepted mundane propositions, rather than any skeptical competitor, the former have a kind of epistemic merit the latter don’t. (d) Experience itself provides immediate justification for the acceptance of mundane propositions. On one way of framing this approach, it is a fundamental epistemic principle that if it appears to S that F, S is justified in believing that F. Insofar as experience justifies believing F, F is epistemically more meritorious than the hypothesis of massive sensory deception, and there is no underdetermination (see Pryor, 2000). (e) It has been argued that skeptical hypotheses are “dialectically” or “pragmatically” self-refuting, or even logically inconsistent. If this is so, and our mundane beliefs aren’t similarly defective, then our mundane beliefs are certainly superior to their competitors.<sup>7</sup>

I regard these views as worthy of serious consideration, but, in my judgment, all of them are ultimately untenable.<sup>8</sup> The alternative I favor may be outlined as follows: when one is choosing between competing candidates for belief A and B, one has good reason to accept A rather than B if A provides a better explanation of a relevant body of facts than B does. That is, I uphold the legitimacy of what is known as *inference to the best explanation* (hereafter, “IBE”). Skeptical hypotheses are less successful than our mundane beliefs in explaining a relevant body of facts, namely, facts about our mental lives.<sup>9</sup> Because of this disparity, we have reason to favor mundane propositions over their skeptical competitors, and the choice between them isn’t underdetermined after all. Therefore, premise (2) of the deceiver argument is false, and skepticism stands refuted.<sup>10</sup> This approach faces some significant obstacles, however. There is unclarity about what counts towards explanatory goodness, and about how such goodness could be assessed or compared – there is even substantial disagreement as to what constitutes an explanation at all.<sup>11</sup> I certainly don’t propose to resolve these difficulties here. Rather, in what follows, I will make some specific assessments of explanatory goodness that I take to be plausible in their own right.

A further complication is that the skeptical hypothesis that one is the victim of massive sensory deception could be developed in any number of ways. Presumably, these will vary with respect to their degree of explanatory goodness. Consider what I will call the “minimal skeptical hypothesis” (MSH). The content of the MSH is just

that your experience is caused in a delusory manner, and no more: with respect to every Z, if it appears to you that Z, then something causes it to appear to you *falsely* that Z. (If you like, you can embellish the MSH a little by adding that the cause is an evil demon or the computer of a mad neuroscientist.) It seems to me that the MSH has little, if any, explanatory power. The putative explanations it offers are impoverished and ad hoc. According to the MSH, it appears to you that Z because something causes it to appear to you falsely that Z. That is like explaining that you fell asleep because of the action of something with a dormitive virtue – that is, like saying that you fell asleep because something caused you to fall asleep – which is practically no explanation at all.<sup>12</sup>

However, I have offered no reason to suppose that all skeptical hypotheses have to fail as badly as the MSH. The possibility apparently lies open that a skeptical hypothesis could posit a nexus of causes and regularities that is as cohesive and economical as the body of our ordinary beliefs about the world. Call this body of beliefs the “real world hypothesis” (RWH). In fact, it seems that the skeptic could take over the causal-explanatory *structure* of the RWH, but substitute within it reference to objects and properties other than the ones we take to be real. For example, suppose you move your hand so as to pet a cat, which purrs in response. The motion of your hand (H) causes your (V) visual experience as of a hand moving, and it causes (C) the cat to purr. The cat itself causes you to have (P) an auditory experience of a gentle rumbling. But now imagine that the brain-in-a-vat skeptical hypothesis were true, and that the computer has a file (H\*) that stands in for your hand and another file (C\*) that stands in for the cat. (H\*), rather than a real hand, causes (V) and also activates (C\*) the cat-file in a particular way. The cat-file, in turn, causes (P) your auditory experience of a purring sound.<sup>13</sup> I will call a skeptical competitor of this sort the “isomorphic skeptical hypothesis” (ISH). The relationships among causes and effects according to the ISH match those of the RWH. To that extent, it seems, the explanations provided by the one are no better or worse than the explanations provided by the other. There is, then, no difference in explanatory success that favors the RWH over the ISH.

So far, we have considered two ways the skeptic might proceed. Either he deploys the MSH (or some variant of it) as a competitor to the RWH, or he puts forward a skeptical hypothesis that is structurally identical to the RWH. The MSH differs in structure from the RWH, but is explanatorily inferior to it. The ISH is supposed to have the same structure as the RWH, and, consequently, explain just as well as the RWH does. But it may seem that the inadequacies of the MSH needn’t carry over to every other skeptical competitor that differs in structure from the RWH. So there could be some such competitor that equals the RWH in explanatory merit.

There are several points to make in response to this suggestion. First, a certain class of skeptical hypotheses (the MSH and its ilk) may be rejected on explanatory grounds. Formulations of the deceiver argument typically invoke skeptical hypotheses of just this kind, and those versions of the argument *are* refuted by the appeal to explanatory considerations. Second, the claim that, for some particular hypothesis H, there is a competitor that differs in structure from H, but equals H in explanatory merit, is by no means trivial and cannot be taken for granted. Adducing an equally good, structurally different competitor to a given hypothesis requires one to undertake substantive theory-construction, which is by no means guaranteed to succeed.<sup>14</sup> Finally, as I will suggest below, the RWH enjoys certain explanatory advantages over

skeptical hypotheses. There is no reason to suppose that these advantages are offset by unspecified *disadvantages* attaching to the RWH, which a structurally different skeptical competitor will somehow manage to avoid. Hence, it is doubtful that one could cobble together a structurally different skeptical competitor to the RWH that will succeed in matching the RWH in explanatory merit.

To provide a setting for further argument, it will be helpful to say something more about the nature of our problem and its solution. As I see things, the underlying motivation for skepticism is that perception, the ultimate source of whatever empirical knowledge we have, is a causal relation. Our perceptual experiences are the effects of the world upon us. In general, causal relations are contingent, so there seems to be no bar in principle to anything causing a given effect. Various skeptical hypotheses serve to make us vividly aware of this point in the perceptual case. Our perceptual experiences could be caused by familiar sorts of things in the world, but those experiences could also be caused by radically different sorts of things (a computer, a demon). In the same way, it seems that specifying any wider structure of causal relations and regularities doesn't fix which properties and objects participate in that structure. Unfamiliar properties (e.g., a computer disk's magnetic patterns, a demon's mental states) could stand in for the familiar ones invoked by the RWH, without any additional loose ends or explanatory regularities. The result would be a skeptical hypothesis that has the same structural explanatory virtues as the RWH.

At the same time, though, there is a natural impression that a good explanation for why X behaves like something that is F is that X is, indeed, F. An explanation of why something that *isn't* F nevertheless behaves as though it were F seems bound to involve greater complications. However, someone with Humean scruples may well find such talk, as it stands, mysterious and misguided. She will think that X's being F by itself implies nothing about how X "behaves." X's being F doesn't determine what regularities X enters into, including whether X appears to us to be F or not.

The issues here are deep and hard, but I want to suggest that explanatory considerations can still do the requisite anti-skeptical work.<sup>15</sup> Even though the ISH is supposed to share the structure of the RWH, it is meant to be a thoroughgoing competitor to it. That is, the ISH is required to explain the character of our experience as successfully as the RWH does, while positing a very different disposition of matter in space. Rather than there being ordinary arrangements of ordinary-shaped objects, there is just your brain, its vat, and the computer with which it interacts.<sup>16</sup> Now, the ascription of shapes and locations to things does a great deal of explanatory work for us. The ISH has to achieve the same explanatory success by ascribing properties other than shape and location to some other kind of object. Call these alternative properties "pseudo-shapes" and "pseudo-locations." My proposal is that skeptical hypotheses are bound to suffer an explanatory disadvantage insofar as they have to make do with pseudo-shapes and pseudo-locations instead of genuine shapes and locations.

The disadvantage is a lack of simplicity. It is, in principle, a defect to proliferate explanatory apparatus without any commensurate increase in the depth or scope of what is explained. So, other things being equal, a hypothesis that invokes fewer explanatory regularities is preferable to a competitor that invokes more. I take it that we are able to recognize differences of this sort and choose between hypotheses accordingly. For example, we find a version of mechanics that provides a uniform account of terrestrial and celestial motion preferable to a mechanics that invokes one



set of laws governing the former and another set of laws governing the latter. Similarly, we accept that most people who seem comfortable really are so, rather than “super-spartans,” that is, people who are in great pain, yet have the desire and fortitude not to manifest it in any way.<sup>17</sup> I believe that the RWH is more economical than the ISH in much the same way, and so the RWH may be favored over the ISH on explanatory grounds.

In previous work, I tried to establish that result as follows (Vogel, 1990). It is, I claimed, a necessary truth that two distinct objects can’t be in the same place at the same time.<sup>18</sup> This truth about location is deeply ingrained in our understanding of the world. If I walk two blocks north and then two blocks west, while you walk two blocks west and then two blocks north, we encounter the very same thing, whatever it may be, when we arrive. In the RWH, such facts are explained by the necessary truth that if X has the property of being located at a genuine location L, and Y is distinct from X, then Y doesn’t have the property of being located at L. Now, where the RWH says that two objects don’t have the same location at the same time, the ISH must say that two pseudo-objects don’t have the same pseudo-location at the same time. But in so doing, the ISH must invoke an extra empirical regularity, since it is metaphysically possible for distinct objects to share any property other than location.<sup>19</sup> To make this point more vivid, imagine that the nefarious computer in the skeptical scenario keeps track of the “location” of “your hand” by writing coordinates in the hand-file, and likewise for the pseudo-location of other pseudo-objects. The same coordinates could be written in two different files. It is therefore an additional empirical regularity, rather than a necessary truth, that different pseudo-objects have different pseudo-locations, that is, that they are not, so to speak, “double-booked.”<sup>20</sup>

The details of this argument raise a number of questions.<sup>21</sup> But even if this particular line of thought doesn’t serve as a definitive response to the skeptic, it illustrates a crucial difficulty the skeptic faces. Some relations between properties are necessary. To that extent, properties have what might be called a “modal configuration.” If F and G are properties whose modal configurations are different, one can’t simply substitute reference to G for reference to F within a hypothesis, and assume that the structure of the original can be preserved. My argument about the role of location-properties in the RWH was meant to exploit that fact. The idea was that the modal configuration of location-properties includes their being uniquely instantiated, that is, that necessarily, if x is located at L, then nothing else can be located at L. No other genuine property P is such that, necessarily, if x is P, then nothing else is P. Hence, a version of the ISH which adverts to other properties in place of location-properties must add some empirical “exclusion principle” for these pseudo-locations to explain why no two items share the same pseudo-location. The upshot would be that substitution of reference to pseudo-locations rather than genuine locations in the ISH doesn’t permit preservation of the explanatory structure of the RWH. More generally, the ISH will fail to match the RWH in explanatory adequacy unless the properties invoked by the ISH have exactly the right modal configurations. I see no reason to believe that this requirement can be met, and substantial reasons to doubt it.<sup>22</sup> Considerations of explanatory adequacy will, then, favor the RWH over the ISH.<sup>23</sup>

Here is where we stand. Skeptical hypotheses usually discussed, such as the MSH, are underdeveloped, explanatorily impoverished, and thus defective. The question then arises whether there is a superior skeptical hypothesis which could compete on equal terms with the RWH. My answer is no.

### III

The ground gained so far is significant, but not unlimited. We have a refutation, or the outline of a refutation, of a certain kind of skepticism: domestic skepticism about the external world, motivated by the underdetermination principle. I have assumed that, since the skepticism under consideration is domestic skepticism, a reply may freely appeal to whatever epistemic principles we ordinarily accept, including principles of IBE. But the status of IBE itself is controversial, and some will doubt that an appeal to explanatory considerations has much force against the skeptic.

A principal criticism directed against IBE is that epistemic justification would be worthless unless beliefs that are so justified are thereby more likely, or have a greater tendency, to be true. Now, suppose  $H_1$  provides appropriately simpler, hence better, explanations of a given domain than  $H_2$  does, so that IBE underwrites the choice of  $H_1$  over  $H_2$ . This choice seems to presume that the simplicity of an explanation is a guide to truth, that is, that the world is likely to be simple in the ways we appreciate. But, the criticism goes, such a presumption is nugatory – it is a kind of wishful thinking that the world is nice and neat. Since simplicity and truth seem to be independent, explanatory success gives us at best pragmatic, rather than epistemic, reasons to favor one hypothesis over another. The upshot for skepticism is that, if the hypothesis that you are the victim of massive sensory deception is more complex than the real world hypothesis, the former may suffer some pragmatic disadvantage (it may be more cumbersome, inelegant, or the like). Nevertheless, you don't have genuinely epistemic justification for rejecting that hypothesis, and the deceiver argument can still go through.

The issues raised by this kind of argument are various and difficult, and I can't hope to do them justice on this occasion. But I would like at least to register the opinion that, in the end, IBE is no more suspect than other kinds of inductive confirmation. IBE licenses the acceptance of hypotheses that aren't entailed by the evidence that supports them, as any sort of inductive confirmation does. For example, consider an application of induction by enumeration. Observations that the roses in the yard have all been red in the past is evidence for the conclusion that the roses in the yard will be red next season. But the proposition that all the roses in the yard have been red in the past doesn't entail that all the roses in the yard will be red next season. IBE and enumerative induction have still more in common. The criticism of IBE just considered was that acceptance of a hypothesis because of its simplicity presupposes that simplicity is a guide to truth, and that such a presupposition amounts to a kind of wishful thinking that the world is nice and neat. But one might say with equal justice that enumerative induction presupposes that the observed is a guide to the truth about the unobserved, and that this presupposition, too, is a kind of wishful thinking that the world is nice and neat. To this extent, IBE appears to be on a par with other forms of inductive confirmation.<sup>24</sup>

Given the assumption that knowledge requires justification, there are now three positions in play: skepticism about the external world, skepticism about IBE, and skepticism about induction in general. In my view, the relations among them are as follows. If you aren't a skeptic about IBE, you can reject skepticism about the external world. If, however, you are skeptical about IBE for the reasons given above, you will also be a skeptic about induction of all kinds (and presumably about the external world, too).

So far, I haven't offered any reply to skepticism about induction, nor will I attempt to do so. Instead, I'll make some comments that are meant to illuminate, rather than settle, some of the issues now before us.<sup>25</sup>

(1) One can take a sunny view of the situation, according to which there can be no rationally compelling challenge to our basic epistemic principles. A critic might reject our principles, or be committed to principles that conflict with ours. But the mere fact that someone else disagrees with us hardly shows that we must be wrong. If anything, we should stand by our own principles, and regard the critic's stance as mistaken. Assuming then that our epistemic principles include or imply the canons of IBE and induction in general, their legitimacy isn't open to serious question. To use the terminology introduced above, skepticism about IBE is exotic and, consequently, may be ignored. The same goes for skepticism about induction more broadly.<sup>26</sup>

(2) This response may strike some as unsatisfactory. For one thing, it may be disappointing that we are unable to establish the correctness of our epistemic principles vis-à-vis others, from some kind of "neutral" standpoint. In addition, the line of criticism directed against IBE and induction at the beginning of this section may seem to have some weight, and just setting it aside may appear superficial or obtuse. But exactly what is at issue here is somewhat cloudy. One sharpening of the concern about the connection between justification and truth is what I have termed the *problem of misleading evidence* (see Vogel, 2004, 2007). Suppose your evidence E inductively confirms hypothesis H, so that you're justified in believing H. Presumably, you're justified in believing propositions logically weaker than H, including  $(\tilde{E} \ \& \ \tilde{H})$ .<sup>27</sup> If  $(E \ \& \ \tilde{H})$ , then H is false despite your having good evidence for H. We could say that, in such a case, your evidence is misleading with respect to H. Now, we're supposing that you're justified in believing  $(\tilde{E} \ \& \ \tilde{H})$ , that is, that your evidence *isn't* misleading with respect to H. That is a proposition about the world, so you're not justified in believing it unless you have evidence that supports it. It may appear, though, that the only pertinent evidence you have is E. That looks like trouble, because it is hard to see how E could be evidence that E itself isn't misleading with respect to H. But if you have no evidence at all for  $(\tilde{E} \ \& \ \tilde{H})$ , it seems that you're not justified in believing  $(\tilde{E} \ \& \ \tilde{H})$ . It follows that you're not justified in believing H, and the result is a thoroughgoing skepticism about induction. As far as I'm concerned, whether the problem of misleading evidence is a serious threat to inductive knowledge, and, if so, how to respond to it, remain open questions.

(3) In his contribution to this volume, Richard provides a searching and distinctive treatment of motivations for a general skepticism about (non-deductive) "inferential justification." According to Richard, someone who rejects externalist accounts of justification must accept that

When one's justification for believing P involves *inference* from E [one's evidence for P], a constituent of that inferential justification is a justified belief that E does indeed make probable P ... We could, of course, *infer* that the probability connection holds from some other proposition we believe, F, but then we would still need reason to think that F makes likely that E makes likely P. Eventually, however, inferential internalists will need to find some proposition of the form E makes likely P that we can accept without inference.

Substantive propositions of that sort look to be very hard to come by.<sup>28</sup> In that event, skepticism about induction immediately follows. Now, I'm not confident that I grasp exactly what Richard means by "makes probable," but structurally this challenge resembles the problem of misleading evidence. The former makes having evidence for "E makes probable H" a condition for justified belief that H; the latter makes having evidence for "E isn't misleading with respect to H" a condition for justified belief that H. The two requirements are also similar in spirit, or so it seems to me.<sup>29</sup> Thus, I'm inclined to think that there will be a satisfactory response to Richard's challenge if and only if there is a satisfactory response to the problem of misleading evidence – assuming that a response is necessary at all.<sup>30</sup>

This reservation reflects uncertainty as to the status of the demand that, if S's believing E justifies S's believing H, then S must have evidence for the proposition that E makes probable H (hereafter, "Fumerton's requirement," or FR). I take Richard's view to be that, given the falsity of externalism, FR holds. This means, at a minimum, that FR is a condition on what *we take to be* inferentially justified belief. FR is an epistemic principle we accept. Richard maintains that FR leads to skepticism, in the absence of a priori, non-inferential justification for judgments of the form *E makes probable H*. Suppose all that is so. The result is *domestic* inductive skepticism, insofar as such skepticism follows from epistemic principles to which we are committed.

Now, there is at least some reason for denying that FR is a condition for justification, as we understand it. Consider the total body of evidence (TE) available to a subject S. When S has inductive justification for a proposition H, TE confirms but doesn't entail H.<sup>31</sup> According to FR, S still needs reason to think that TE makes probable H. And, since TE doesn't entail H, S presumably needs some additional *evidence* that the making-probable relation holds between TE and H. But S can have no such evidence, since, by assumption, TE is S's total evidence. FR, then, seems to be unsatisfiable in principle. So, if we have a *coherent* conception of justification, it can't include FR. That result, in turn, favors the view that skepticism stemming from FR is exotic, not domestic.<sup>32</sup>

But maybe justification as we conceive it is subject to FR, and it undoes itself in the way just described. Our situation with respect to epistemic justification would then be something like the one we face with respect to the concept of a set, in the face of Russell's Paradox, or with respect to the concept of truth, in the face of the semantic paradoxes. After the bad news is in, we can try to minimize our disappointment by seeking out a successor concept that does a good part of the work done by its untenable predecessor. The question in the present case would then be whether our concept of justification, shorn of FR, is enough like the original for us to care about it as before. If so, the encounter with the sort of skepticism Richard envisions might end in partial concession rather than full capitulation.

#### IV

In this essay, I have argued that some kinds of skepticism need to be refuted, and can be. Other kinds of skepticism can't be refuted, but needn't be. I hope, and want to suggest, that the division is exhaustive.

## Acknowledgments

I would like to thank Philip Bricker, Richard Feldman, Michael Glanzberg, Ned Hall, Mark Johnston, Tom Kelly, Tim Maudlin, Richard Moran, Bernard Nickel, Ted Sider, and Steve Yablo for their kind help. I have been very fortunate to work with Richard Fumerton, Ernest Sosa, and Matthias Steup on this project. Finally, I'm deeply grateful to Neil Harpster for his extraordinary efforts on the side of the angels.

## Notes

- 1 Generically, skepticism is the denial that we have some kind of knowledge (or, maybe better, the denial that we have some kind of knowledge one might sensibly suppose we have). The rejected body of knowledge might be identified by subject matter (skepticism about the external world, skepticism about the past) or by mode of acquisition (perception, induction). When I talk about "skepticism" as such, I mean skepticism about the external world. In addition, I assume here that knowledge requires justification, so that the denial that we have justification for a body of beliefs implies the denial that those beliefs count as knowledge. For more on these matters, see Richard's helpful discussion, and section III below.
- 2 What follows is an updated version of Descartes's "Evil Genius" argument in the First Meditation.
- 3 For simplicity, I'm ignoring the possibility that the toaster shut off because it burned out *and* also because the fuse blew. I'm proceeding as though the two hypotheses are *competitors*, in the sense that the truth of one is logically incompatible with the truth of the other.
- 4 For discussion of these issues, see Vogel (2007).
- 5 Richard considers the investigation of various conditional claims of the form *If such and such principles hold, skepticism follows*. His characterization of the way I understand the situation is helpful, except for two things. First, I don't think I've misidentified the epistemic principles to which we're committed, being "stingy" about excluding "*sui generis* epistemic principles of perception" (see Vogel, 1997). Second, I'm not sure how Richard distinguishes "conceptual requirements for knowledge and justified belief" from "the legitimacy of certain basic sorts of inference." Thus, I don't see that there are two different ways of drawing a distinction between domestic and exotic skepticism.
- 6 For extended, critical discussion, see Williams (1992).
- 7 For extended, critical discussion, see Greco (2000).
- 8 For my evaluation of these views, see Vogel (1993, 1997).
- 9 Whether there is such a body of facts to be explained is disputed by Williams (1992) and Byrne (2004). For a response, see Vogel (1997).
- 10 Other authors have advanced different replies to skepticism that, in one way or another, appeal to IBE. Two recent contributions are those due to BonJour (1998) and Peacocke (2004).
- 11 For different views about what an explanation is, see Pitt (1988).
- 12 My discussion blurs some important distinctions. At least some explanations are explicitly causal, where the *explanans* is a cause of the *explanandum*. I assume that at least some explanations involve the subsumption of the *explanandum* under some law or regularity, either explicitly or because causation itself requires nomological support. In the text, I make no attempt to sort this out.
- 13 The terms of the causal relation are usually taken to be events, not things. The scrupulous reader may make the necessary adjustments to the text.

- 14 Otherwise, science would be even harder than it is. For example, proponents of steady-state cosmology would have been able to explain the presence of background radiation as well as proponents of the Big Bang. Then, too, the job of defense attorneys would be far easier than it actually is.
- 15 According to the view known as “structural realism,” we can know that things in the world have at least some relational properties, but we are unable to know what intrinsic properties they have. A structural realist might then deny that we have reason to accept the RWH rather than the ISH. I take it Richard is sympathetic to such a position. But if shape is an intrinsic property of bodies, then a structural realist of this stripe has to say that we don’t know what shapes things have, which is a more skeptical outcome than I would like to see. Matters here are complicated, however, involving among other things questions about the identity conditions for properties that I cannot discuss in any detail.
- 16 The difference in purely spatial features of our environment doesn’t exhaust the disagreement between the ISH and the RWH. But I think that if skepticism about the spatial configuration of the external world can be refuted, dealing with whatever else remains of external world skepticism would be relatively easy.
- 17 The example is from Lewis (1980).
- 18 By “distinct” I mean at least that the objects have no parts in common at the time in question. So, if you think that a statue and the mass of clay from which it is fashioned aren’t identical, the statue and mass of clay still wouldn’t count as distinct, in the way that I’m employing the term. If your favored ontology allows for object-stages, the thesis is meant to imply that there can’t be two distinct object-stages at the same place at the same time.
- 19 On the relation between simplicity and necessary truth see Glymour (1984), and below (compare Sober, 1988).
- 20 A bit more precisely: an account according to which X has the property of being at a location L will explain more economically why X behaves as though it has that property than an account according to which X really doesn’t have the property of being at location L, but rather has a pseudo-location which corresponds to L.
- 21 Mark Johnston and Philip Bricker have cautioned that there may be counterexamples to the claim that the co-location of distinct objects is impossible. Ned Hall and Tim Maudlin have suggested that the propensity of bodies to exclude one another is a matter of the dynamics governing those bodies, and thus not necessary.
- 22 Here is the sort of thing I have in mind. I am inclined to think that spatial properties are such that, necessarily, they conform to some kind of geometry, and their so doing enters into the modal configurations of those properties. A relatively clear – although somewhat out of the way – case in point may be the following. Consider the property of being a perfect solid. In Euclidean space, there are exactly five perfect solids, so the determinable *perfect solid* has precisely five determinates. Suppose the RWH invokes the fact that something is a perfect solid in giving an explanation (say, of why the chances that each side of an object will come up on a toss are equal). To maintain structural parity with the RWH, the ISH would have to invoke some other property (the property of being a “pseudo-perfect-solid”) with exactly five determinates, while avoiding other difficulties. I see no reason to think that this condition can be met. The sort of difficulty I am raising here is analogous to that facing the possibility of spectrum inversion: color-properties bear complex relations to one another, such that a simple substitution of one for another in an account of color experience may not be feasible. See Hardin (1997).
- 23 There may be another way of arguing for the same result, which moves in something like the opposite direction from the foregoing. The ISH isn’t merely a competitor to the RWH, but is something like a simulacrum of it. As the ISH more closely resembles the RWH in structure, the question of what makes the ISH *different* from the RWH becomes more pressing. The problem, loosely speaking, is that the ISH needs to specify how things that behave as though they are F still *aren’t* F, without taking on unwanted encumbrances.

- 24 There are other notable objections to IBE, including those due to Van Fraassen. For a survey and discussion of these, see Ladyman et al. (1997). I should note that Van Fraassen would reject the view that IBE is no worse off than any other kind of ampliative inference, and that Richard agrees with Van Fraassen at least that far.
- 25 One important difference is that the deceiver argument works by invoking the underdetermination principle. Richard examines skepticism about induction (and the external world) motivated by what he calls the principle of inferential justification. For a discussion of these issues, see Vogel (2007), and below.
- 26 This position is similar to the one adopted by P.F. Strawson (1952), although it doesn't include Strawson's view that claims about inductive support are analytic. Strawson undertakes to show in some detail how various motivations for inductive skepticism covertly distort or diverge from our ordinary understanding of what is required for epistemic justification. Those efforts bolster the judgment that skepticism about induction is invariably exotic in character.
- 27 That is, it is not the case that E is true while H is false. This is a more precise rendering of what it is for the world to be "nice and neat."
- 28 Richard also stresses that the justification for the underpinning propositions must be a priori, which if anything makes it less plausible that there are such propositions.
- 29 See Vogel (2007) for discussion of what I take to be the likeness of the two. But there are differences, as I note there.
- 30 But some caveats are called for. First, adopting some kind of holism might block the threat Richard envisions, although it is somewhat harder to see how holism could help with the problem of misleading evidence. Second, the link between justification for H and justification for ( $\tilde{E} \ \& \ H$ ) is secured by the eminently plausible principle that justification is closed under known logical implication. What underwrites the connection between justification for H and justification for (E makes probable H) is somewhat harder to see, and is perhaps more tenuous.
- 31 I take it Richard would not protest at this point, in light of his reply to a criticism by Michael Huemer.
- 32 One way to read FR is that it covertly appeals to a deductivist view of confirmation, as follows. Whenever E is inductive evidence for H, FR finds E insufficient, and demands further evidence (for the proposition that E makes probable H). So, the import of FR is that inductive evidence is never sufficient for justification; only deductive evidence would really do. Richard agrees that the standard of justification for knowledge should not rise that high, although I take it he would reject the way I have characterized FR. Points related to this one, and the one in the text, may also provide some refuge from the problem of misleading evidence.

## References

- BonJour, L. (1998) *In Defense of Pure Reason*. Cambridge: Cambridge University Press.
- Byrne, A. (2004) How hard are the skeptical paradoxes? *Noûs* 38: 299–325.
- Glymour, C. (1984) Explanation and realism. In *Scientific Realism*, ed. J. Leplin. Berkeley: University of California Press.
- Greco, J. (2000) *Putting Skeptics in Their Place*. Cambridge: Cambridge University Press.
- Hardin, C.L. (1997) Reinverting the spectrum. In *Readings on Color, Volume I*, ed. A. Byrne and Hilbert. Cambridge, MA: MIT Press.
- Ladyman, J., Douven, I., Horsten, L. and van Fraassen, B. (1997) A defence of Van Fraassen's critique of abductive inference: reply to Psillos. *Philosophical Quarterly* 47: 305–321.
- Lewis, D. (1980) Mad pain and Martian pain. In *Readings in Philosophy of Psychology, Volume I*, ed. N. Block. Cambridge, MA: Harvard University Press.

- Peacocke, C. (2004) Explaining perceptual entitlement. In *The 'Challenge' of Externalism*, ed. R. Schantz. Berlin: de Gruyter.
- Pitt, J., ed. (1988) *Theories of Explanation*. Oxford: Oxford University Press.
- Pryor, J. (2000) The skeptic and the dogmatist. *Noûs* 34: 517–549.
- Sober, E. (1988) Let's razor Occam's razor. In *Reconstructing the Past: Parsimony, Evolution, and Inference*. Cambridge, MA: MIT Press.
- Strawson, P.F. (1952) *Introduction to Logical Theory*. London: Methuen.
- Vogel, J. (1990) Cartesian skepticism and inference to the best explanation. *Journal of Philosophy* 87: 658–666.
- Vogel, J. (1993) Dismissing skeptical possibilities. *Philosophical Studies* 70: 235–250.
- Vogel, J. (1997) Skepticism and foundationalism: a reply to Michael Williams. *Journal of Philosophical Research* 11–28.
- Vogel, J. (2004) Subjunctivitis. *Philosophical Issues* 14: 426–455.
- Vogel, J. (2007) Skeptical arguments. *Philosophical Studies* 134: 73–88.
- Williams, M. (1992) *Unnatural Doubts*. Oxford: Blackwell.
- Wittgenstein, L. (1972) *On Certainty*. New York: Harper and Row.

# The Challenge of Refuting Skepticism

*Richard Fumerton*

One of the keys to success in battle is to find and hold ground favorable to the engagement. The skeptic, it is often claimed, insists on unacceptably strong standards for knowledge or justified belief and proclaims victory when those standards are not met. In his paper Jonathan distinguishes what he calls *exotic* skepticism from what he takes to be its more interesting *domestic* versions. I join with Jonathan in an attempt to make clear what a skeptic might reasonably demand of a successful refutation of skepticism. As will be obvious, Jonathan and I agree on a great many issues, but may part company on just exactly how to understand epistemic probability and the need for access to that probability as a requirement of acquiring ideal justification.<sup>1</sup> I will conclude with a brief discussion of the prospects of avoiding skepticism within the framework of the domestic skepticism that Jonathan discusses.

## Defining the Skeptical Challenge

When addressing a skeptical challenge one should probably first attempt to discover whether the skeptic is bent on undermining *knowledge*, or has the more ambitious goal of undermining *justified belief*. We can also distinguish *global* skepticism from its *local* varieties. The global skeptic with respect to *knowledge* claims that we don't know anything – we can't even know that we don't know anything (though we might be justified in believing that). The global skeptic with respect to *justified belief* claims that we have no justified beliefs – we are not even justified in believing that we have no justified beliefs. There have been relatively few global skeptics with respect to knowledge and fewer still with respect to justified belief. The global skeptic with respect to justification



must obviously deal with the charge of epistemic self-refutation. Any argument for global justification skepticism will have premises that cannot be justifiably asserted if the conclusion of the argument is true. It is not clear that that fact entitles one to ignore the arguments, but it puts the most extreme of skeptics in an awkward position to say the least. In any event, even rather radical skeptics tend to at least implicitly suppose that they can know truths about their subjective experience and recognize the distinction between legitimate and illegitimate reasoning.

Local skeptics are skeptics with respect to knowledge or justified belief about some particular *kind* of truth: for example truths about the external world, the past, the future, other minds, theoretical entities posited by physics, and so on. Like Jonathan, I will often take, as our paradigm of a local skepticism, skepticism with respect to truths about the external world. The skeptic with respect to knowledge of the external world claims that one cannot know (or sometimes with added emphasis, know *with certainty*) any truths about the physical world. Before investigating this claim, we should decide what conditions we need to meet in order to satisfy our skeptic that we do indeed possess the knowledge in dispute. Suppose that our skeptic insists that we know only when we possess justification that precludes any possibility of error. If we agree, the skeptic might seem to face relatively smooth sailing. The justification I have for believing that I am seated in front of a computer is precisely the same sort of justification that I would have had were there no computer there and were I instead suffering a massive and vivid hallucination (or were I a brain in a vat, a victim of demonic machination, an inhabitant of the *Matrix* world, etc.). My justification for believing that the computer exists, therefore, does not entail the existence of the computer.

The above argument is not uncontroversial. Williamson (2000), Brewer (1999), and others, for example, argue that we might be unable to *recognize* crucial and dramatic differences between epistemic situations in which we find ourselves. We might not be able to tell, for example, the difference between a situation in which we are directly confronted with the computer (a situation in which the computer itself is literally a *constituent* of our visual experience) and a situation in which we are merely hallucinating the computer. Our perceiving that there is a computer in front of us might *in fact* generate a kind of justification that guarantees the computer's existence and, trivially, provide us with evidence of precisely the sort Descartes sought that guarantees the computer's existence.<sup>2</sup> For those of us who find the skeptic's challenge interesting, however, it is difficult to take this idea seriously. While we may be unable to tell the difference between epistemic situations that are only marginally different, it still seems utterly puzzling as to how we could be so incompetent as to fail to notice the difference between two experiential situations that Williamson (and other direct realists) claim should get *radically* different philosophical analyses.

In any event, for the purposes of this debate I will simply presuppose (and Jonathan, I understand, will not dispute) that we can easily imagine skeptical scenarios that are epistemic counterparts to veridical perception, scenarios in which we would possess precisely the same kind and strength of justification for the same kind of beliefs about our physical environment that we would possess were our perceptions veridical. We will agree here, therefore, that we cannot avoid knowledge skepticism with respect to the physical world if we understand knowledge as requiring justification so strong that it eliminates the possibility of error.

Of course, the moral most contemporary epistemologists draw is that we should reject Cartesian standards for knowledge. Our only hope of avoiding skepticism with respect to knowledge of the external world (the past, the future, other minds, and so on) is to accept more relaxed standards for knowledge. While Jonathan and I may end up disagreeing on whether or not one can defend weaker requirements for knowledge, that disagreement is probably unimportant for this debate. I think we do agree that the most interesting epistemological question is whether we can find justification that makes likely for us the truth of what we believe. As a result I'm going to frame the skeptical challenge employing the concept of justification with the understanding that if one can know without possessing justification that guarantees the truth of what one believes, the discussion could just as easily take place employing the concept of knowledge.

## Underdetermination

If we switch our focus from knowledge to justified belief then what should one be expected to show in order to defeat skepticism? Here, I agree completely with Jonathan that the issue is best understood in terms of underdetermination. He suggests two principles. While he states the principles in terms of knowledge, one can easily reformulate them to concern justification instead:

- UP1: If  $q$  is a competitor to  $p$ , then there is justification for one to believe  $p$  only if one can non-arbitrarily reject  $q$ .
- UP2: If  $q$  is a competitor to  $p$ , then there is justification for  $S$  to believe  $p$  only if  $p$  has more epistemic merit for  $S$  than  $q$ .

For these purposes, competitors to  $p$  include all contraries of  $p$  and any other proposition whose truth would render highly unlikely the truth of  $p$ .<sup>3</sup>

UP1 and UP2 state only a *necessary* condition for  $S$ 's possessing justification for believing  $P$ . A skeptic might want to remind us that for  $S$  to possess justification for believing  $P$ , there must be an epistemically non-arbitrary way of rejecting the *disjunction* of competitors to  $p$ . This is particularly important to remember in evaluating arguments to the best explanation. To have justification for accepting a given explanation  $h$  for some data  $e$  one must have more reason to accept  $h$  than the *disjunction* of competing explanations. When the skeptic comes up with competitors to the hypotheses of commonsense beliefs, therefore, one must be very careful not to assign these competitors *any* sort of significant possibility of being true. The disjunction of a large number of very unlikely hypotheses is still often very likely to be true.

The nice thing about framing the debate in terms of underdetermination is that we beg no question against the skeptic. We meet the skeptic on a level playing field. If we are to take seriously the epistemic challenge of skepticism, we must insist on no burdens of proof. The beliefs of common sense are not epistemically innocent until proven guilty, and skeptical scenarios are not epistemically guilty until proven innocent.

## Choosing from among Competitors

How are we supposed to figure out whether a hypothesis of common sense satisfies the above requirements? Well, if we are foundationalists, we need both a meta-epistemological account of non-inferential justification and a normative epistemological account of what we are non-inferentially justified in believing.<sup>4</sup> If beliefs about the external world end up in the class of beliefs that are non-inferentially justified, and the skeptic's competitors to those beliefs do not, then the skeptic loses. One takes the propositions of common sense and employs closure to reject the skeptic's competitors.<sup>5</sup>

Given my own views about non-inferential justification, propositions about the physical world (the past, the future, other minds) do *not* get included in the class of propositions for which we possess non-inferential justification. I can't argue here for my version of foundationalism. The view I defend (1995 and elsewhere) is very traditional and now very unpopular. In short, I hold that one is only non-inferentially justified in believing a proposition *p* when one is directly acquainted with the fact that *p* while one has the thought that *p* and one is acquainted with a correspondence between the thought and the fact. Because I think that one can possess in hallucinatory experience the *same* justification that one possesses in veridical experience for believing truths about one's physical environment, I don't think one can be directly acquainted with facts about physical objects.

Jonathan wants to remain neutral on many of the most fundamental questions concerning the *structure* of justification (the debate between foundationalists and coherentists), and even neutral on the question of what we should take to be the class of beliefs non-inferentially justified if we endorse foundationalism. But because he thinks that his approach to meeting the skeptical challenge through argument to the best explanation works from within quite different frameworks, I think he would not object to my presupposing for the purposes of this debate a foundationalism that takes as its base propositions restricted to the phenomenal (the world of appearance).

So how do we move beyond our foundational justification for believing some proposition *e* to justified belief in a proposition of common sense, say a proposition describing our immediate physical environment? The two most obvious suggestions are these:

- 1: For *S* to have justification for believing *p* on the basis of *e*, *e* must make epistemically probable *p* (where *e*'s entailing *p* can be viewed as the upper limit of *e*'s making *p* probable).
- 2: For *S* to have justification for believing *p* on the basis of *e*, *S* must be aware of the fact that (have justification for believing that ) *e* makes probable *p*.

Elsewhere, I have called (2) *inferential internalism* (see Fumerton, 1995, chapter 3). Notice that the inferential internalist does *not* endorse the general principle that to be justified in believing *P* one must be justified in believing that one is justified in believing *P*. Nor does he endorse the principle that to be inferentially justified in believing *P* one must be justified in believing that one is inferentially justified in believing *P*. *Those* principles seem destined to generate vicious regress.<sup>6</sup> The inferential internalist insists only that when one's justification for believing *P* involves *inference* from *E*, a *constituent* of that inferential justification is the justified belief

that E does indeed make probable P. One must see the connection between one's evidence and one's conclusion before one can justifiably draw the conclusion from one's evidence.

If we are foundationalists and we accept inferential internalism, then there isn't just one potentially vicious regress of justification we need to end with non-inferential justification. To be sure, in order to have inferential justification for believing P, we must be able to trace our justification for believing P back to some proposition for which we have non-inferential justification – a proposition that will serve as our premise. But we also need to cauterize a threatening regress concerning the justification we need for believing that our premise makes probable our conclusion. We could, of course, *infer* that the probability connection holds from some other proposition we believe, F, but then we would still need reason to think that F makes likely that E makes likely P. Eventually, however, inferential internalists will need to find some proposition of the form E makes likely P that we can justifiably believe without inference. And, indeed, the *key* to meeting the skeptical challenge for the inferential internalist centers on the ability to find non-inferential justification for accepting probability connections between our available evidence and the propositions of common sense.

Leaving aside for now the issue of how to understand epistemic probability, why should one embrace inferential internalism? Why shouldn't we view the inferential internalist's demands on inferential justification as exotic, as demands that simply *invite* skepticism? Well, consider the astrologer who predicts the future based on beliefs about the positions of planets relative to one another. It seems obvious that a *sufficient* condition for rejecting astrological inferences as rational is that the astrologer has no reason to believe that the positions of planets relative to one another makes *likely* any of his predictions. If a detective infers from someone's appearance that the person is guilty of a crime, one will surely demand evidence for supposing that appearances of this kind are correlated with guilt before one will concede the rationality of the conclusion. These commonplace examples and indefinitely many others like them surely indicate that we *do* embrace the inferential internalist's account of what is necessary for inferential justification.

Mike Huemer (2002) has objected plausibly to the above arguments for inferential internalism. He argues that the examples used to make initially attractive the principle are misleading in that they inappropriately characterize the evidence from which one infers the relevant conclusion. Even astrologers don't think that they can legitimately infer their predictions from propositions describing the positions of planets and the birth dates of people, *and from that information alone*. It should be a truism that much of the argument we actually give outside of a philosophical context is highly compressed, highly enthymematic. As we ordinarily use the term "evidence," we certainly do characterize litmus paper's turning red in a solution as evidence that the solution is acidic. The approach of very dark clouds is evidence of an approaching storm. A footprint on a beach is evidence that someone walked on the beach recently. But it is surely obvious upon reflection that one's evidence for believing that the solution is acidic, for example, is not the color of the litmus paper *by itself*. To legitimately draw the conclusion one would need an additional *premise*, most likely a premise describing a correlation between the color of litmus paper in a solution and the character of that solution.

Once one realizes that the reasoning discussed above is enthymematic, one is positioned to respond to that appearance of an argument for inferential internalism. It is necessary to have some justification for believing that there is a connection between positions of planets and the affairs of people, the approach of dark clouds and storms, footprints on a beach and the recent presence of people before drawing the respective conclusions, but *only* because propositions describing connections or correlations of the relevant sort are implicitly recognized as critical *premises*. Internalists and externalists alike share the foundationalist's insight that inferential justification is parasitic upon the justification we possess for believing the relevant premises of our arguments. If the astrologer is relying on an unstated, but critical, premise describing correlations between astronomical facts and the lives of people in reaching her conclusion, she will, of course, need justification for believing that *premise*. But that in no way suggests that when we have fully described *all* of the relevant premises from which a conclusion is drawn, we should require that the person who draws that conclusion have additional evidence for believing that the premises make probable the conclusion. The existence of the relevant connection between premises and conclusion is enough.

One can make just as strong a case for inferential internalism, however, by focusing on non-enthymematic reasoning. Consider the case of someone who infers P from E where E logically *entails* P. Is the inferential internalist right in maintaining that in order for S to believe justifiably P on the basis of E, S must be aware of the fact that (or at least have a justified belief that) E entails P (or alternatively, that the inference in question is legitimate)? The answer still seems to me obviously yes. We can easily imagine someone who is caused to believe P as a result of believing E where E does in fact entail P, but where the entailment is far too complicated for S to understand. Unless S sees that P follows from E, would we really allow that the inference in question generates a justified belief? Or to make my case a bit stronger, would we allow that the person who reaches the conclusion has philosophically relevant justification or ideal justification – the kind of justification one seeks when one searches for philosophical assurance.

Summarizing, I'm not at all sure that the inferential internalist is imposing unreasonably exotic requirements on justification, at least if the justification we seek is justification that provides assurance from the first-person perspective. I would allow that one might well acknowledge *derivative*, less demanding concepts of justification. Elsewhere (Fumerton, 2004), I have argued that one might allow that when one is caused by the fact that E to believe P, when the fact that E is the truth maker for the proposition that E and the proposition that E makes probable P (either alone or with other propositions in our background evidence), one has a *kind* of justification for believing P. It is not the kind of justification I think the *philosopher* seeks, and to determine if we have it, we still need a justified belief of the sort sought by the inferential internalist.

## **The Analysis of Epistemic Probability**

Whether or not we adopt inferential internalism, we need an analysis of the probability connection that by itself or as the object of awareness is partially constitutive of inferential justification. There are at least two quite different approaches one might take to

analyzing the epistemic probability with which philosophers are concerned. One approach attempts to understand what it means to say of our evidence that it makes probable a conclusion in terms of epistemic evaluation of belief. Crudely, E makes probable P when one's justification for believing E gives one justification for believing P (were E the only evidence one possesses). A philosopher taking this approach must analyze epistemic concepts without appealing to the concept of epistemic probability.

On the other approach, one appeals to an allegedly prior and more fundamental understanding of the relation of making probable holding between propositions in order to explicate inferential justification. On this approach, one's inferential justification for believing P on the basis of E is constituted by either the existence of, or, according to the inferential internalist, our *awareness* of, a probability connection holding between propositions. Just as one proposition or conjunction of propositions can entail another, so also one proposition or conjunction of propositions can make probable another. Just as there are deductively valid arguments whose deductive validity is not analyzable by epistemic concepts, so also there are valid non-deductive argument forms whose legitimacy depends on relations between the *contents* of premises and conclusions.

In what follows, I'm going to assume without argument that this last approach is correct. It is simply hard for me to believe that the justificatory status of inferentially justified beliefs is not fundamentally derived from relationships between that which is believed. Put another way, it is surely a feature of the *arguments* whose premises and conclusions are believed that is key to understanding the justificatory status of the beliefs formed in the conclusions as a result of justified belief in the premises.

But what is the best way of understanding the relation of making probable that holds between certain propositions? This debate has a long history, one that predates, but in many ways foreshadows, the now more familiar contemporary internalist/externalist controversies in epistemology.<sup>7</sup> Painting with a very broad stroke, one can attempt to analyze probability claims in epistemology on the well known model of relative frequency that is offered as a way of interpreting claims about the probability of an individual or event having a certain characteristic. On a very crude interpretation of the frequency theory, to say of something that it is probably G is always elliptical for a more complex relativized claim of probability. One must refer the individual *a* that is G to some reference class F, and the more perspicuous statement of the probability claim is one about the probability of *a*'s being G relative to its belonging to the class F. The truth conditions for the claim of relative probability are determined by the percentage of Fs that are G. The higher the percentage of Fs that are G the more likely it is that something is G relative to its being F.

One could borrow at least the spirit of the relative frequency interpretation of probability and apply it to relations between propositions in the following way. We could suggest that in claiming that P is probable relative to E we are simply asserting that E and P constitute a pair of propositions, which pair is a member of a certain class of proposition pairs such that, when the first member of the pair is true, usually the second is. Thus in saying that *a*'s being G is probable relative to its being F and most observed Fs being G, I could be construed as claiming that this pair of propositions is of the sort: most observed Xs are Y and this is X/this is Y, and most often it is the case that when the first member of such a pair is true, the second is. Similarly, if I claim that my seeming to remember eating this morning (E) makes it likely that I did eat this morning (P), I could be construed as asserting that the pair of propositions

E/P is of the form S seems to remember X/X, such that most often when the first member of the pair is true, the second is.

The above view obviously resembles, at least superficially, the reliabilist's attempt to understand justified belief in terms of reliably produced beliefs. And it encounters many of the same difficulties. Just as the relative frequency theory of probability must inevitably move beyond actual frequencies in defining probability, so both the above account of epistemic probability and the reliabilist will inevitably be forced to move beyond actual frequencies in order to define the relevant epistemic probability/reliability. Just as reliabilism must deal with the generality problem, so the above approach to understanding epistemic probability as a relation between propositions must deal with the problem of how to choose from among alternative ways of characterizing the class of propositions pairs to which a given pair belongs. Just as many reliabilists are troubled by the implications of their view for what to say about worlds in which demons consistently deceive epistemically faultless believers, so a frequency theory of epistemic probability must deal with similar alleged counterintuitive consequences about what is evidence for what in demon worlds. Lastly, and most importantly for this debate, both reliabilism and the frequency theory of epistemic probability will be anathema to the inferential internalist who is convinced that one needs non-inferential *access* to probability connections in order to gain philosophically satisfying inferential justification. The inferential internalist who is a foundationalist will need to end a potential regress when it comes to gaining access to probabilistic connections. If one's model for foundational knowledge is something like knowledge of truths made true by facts with which one is directly presented, there seems no hope that one will get *that* kind of access to either the reliability of a belief-forming process or a probability relation (understood in terms of frequency) holding between propositions.

One of the historically most interesting alternatives to the frequency interpretation of epistemic probability is a view developed some 80 years ago by Keynes (1921). Keynes wanted to model epistemic probability on entailment. He held that just as one can be directly aware of the entailment holding between two propositions, so one can also be directly aware of a relation of making probable holding between two propositions. There are, of course, obvious differences between entailment and making probable. From the fact that P entails Q it follows that the conjunction of P with any other proposition entails Q. From the fact that P makes probable Q, it doesn't follow that P together with anything else makes probable Q. But for all that, we could still take making probable to be an a priori internal relation holding between propositions (where an internal relation is one that necessarily holds given the existence and non-relational character of its relata). P and Q being what they are, it cannot fail to be the case that P makes probable Q. (It might also be true that P, R, and Q being what they are it cannot fail to be the case that (P and R) makes probable not-Q.)

Against the Keynesian, one might argue that it is patently absurd to suppose that making probable is an internal relation holding between propositions. Such a view yields the absurd consequence that claims about evidential connections are necessary truths knowable a priori. If anything is obvious it is that the discovery of evidential connections is a matter for *empirical* research. While the objection might seem initially forceful, one must remember the point we conceded in considering Huemer's objections to inferential internalism. There is certainly no necessary evidential connection between litmus paper's turning red in a solution and the solution's being acidic, between

dark clouds and storms, between footprints on a beach and the prior presence of people. But then on reflection, we decided that it is misleading to characterize the litmus paper, dark clouds and footprints as the evidence from which we infer the respective conclusions. What we call evidence in ordinary parlance is just a *piece* of the very elaborate fabric of background information against which we draw our conclusions. So we shouldn't *expect* to find Keynesian probabilistic connections holding between, for example, the proposition that the litmus paper turned red and the proposition that the solution is acidic.

Where *should* we look for a plausible example of Keynes's relation of making probable? The obvious, though perhaps not all that helpful, answer is that we should look for it wherever we have what we take to be legitimate, *non-enthymematic* and *non-deductive* reasoning. The trouble, of course, is that philosophers don't agree with each other about what constitutes legitimate though deductively invalid reasoning. One might look at the relationship between the premises and conclusion of an enumerative inductive argument. Less plausibly, perhaps, one might think about the connection between the proposition that I seem to remember having an experience and the proposition that I had the experience. Still more problematically, we might suggest that my seeming to see something red and round necessarily makes probable that there is something red and round.

The view that there is an internal relation of making probable that holds between propositions is just what the inferential internalist *desperately* needs in order to avoid vicious regress. As we saw, if inferential internalism and foundationalism are true, then unless we are to embrace a fairly radical skepticism, we must find some proposition of the form E makes probable P that we can justifiably believe without inference. Since most foundationalists will concede that there are at least some propositions of the form E *entails* P that one can know without inference, the closer we can make our analysis of making probable resemble our analysis of entailment, the more plausible will be the claim that we can know without inference propositions of the form E makes probable P.<sup>8</sup> As I implied earlier, I think that *if* there is a solution to skepticism it involves the ability to know a priori epistemic probability claims. But that is a very big *if* indeed, and the reason I think skepticism looms so ominously on the horizon is that it is difficult to convince oneself that one is acquainted with the relevant probability connections.

## **Another Sort of Domestic Skepticism**

Suppose one cannot convince oneself either that there are relations of epistemic probability holding between propositions or that we have access to such relations. Is there further discussion we might have with the skeptic? Jonathan and I agree that skepticism about the knowledge that Descartes sought is not worth fighting. I'm not sure whether or not we agree that skepticism that presupposes inferential internalism is interesting. In any event, *these* issues concern the potentially exotic nature of *conceptual* requirements for knowledge and justified belief. As I understand Jonathan, however, there is yet another way of distinguishing between exotic and domestic skepticism. The domestic skeptic is prepared to *give* us the legitimacy of certain basic sorts of inference and will argue that even with these argument forms at our disposal we will be unable to generate



the conclusions of common sense from the available evidence. Now the inferential internalist discussed above will allow the use of an argument form only if you can know that the premises make probable the conclusion. You can have inferences from memory, for example, only if you know that your seeming to remember that you had some experience makes likely that you had the experience. You can legitimately employ inferences of enumerative induction only insofar as you can discover the probability connection between premises and conclusions of inductive inference. And the same would be true of the reasoning to the best explanation that Jonathan sees as our best hope of getting justification for beliefs about our physical environment.<sup>9</sup> Before we can even assess the plausibility of the claim that the premises of arguments to the best explanation make probable their conclusions, we would need, of course, a very clear understanding of the *structure* of such arguments.

Whether or not there exists a rich enough array of legitimate and accessible non-deductive reasoning to generate conclusions of common sense, we certainly *can* still ask *conditional* questions of the following form: *assuming that* this or that form of reasoning is legitimate, can we non-arbitrarily choose the hypotheses of common sense over various skeptical hypotheses? Of course, if one includes among the inference forms described in the antecedents of the conditionals Chisholm-style epistemic principles whose formulations are tailor-made to avoid skepticism, the project might seem to lose interest. If, for example, it is conceded that it is epistemically permissible to infer that there is something red and square before one when it appears to one as if there is something red and square, then it will come as no surprise that sense perception will generate many of the conclusions of common sense. If one tries to get reasons for believing in the physical world employing only inferences of memory and enumerative induction, the project will be uphill. If one adds to these sorts of available arguments, reasoning to the best explanation, perhaps one might stand a better chance.

Again, I would emphasize that a reasonable skeptic should have no objection in principle to carrying on discussion concerning the plausibility of relevant *conditionals* asserting what one would be justified in believing were certain inferences legitimate. That same skeptic has every right to insist, however, that one can and ultimately *must* ask questions concerning the legitimacy of the reasoning referred to in the antecedents of the conditionals.

We can also make a distinction between domestic and exotic skeptics based on which principles of reasoning the skeptic is willing to presuppose as unconditionally true. We will, of course, still need some *principled* way of figuring out which presuppositions define the domestic skeptic. Jonathan obviously wants reasoning to the best explanation in the mix, and would probably appreciate an epistemic principle licensing conclusions about the past based on memory. He *wouldn't* insist that the domestic skeptic grant us *sui generis* epistemic principles of perception. One philosopher's domestic species of skeptic, however, is another philosopher's exotic skeptic. Jonathan would be characterized by many (not by me) as being unreasonably parsimonious in the epistemic paths he allows us to travel in an attempt to leave our foundations. One *could* try to single out the relevant presuppositions that define reasonable domestic skeptics by appealing to psychological facts about what inferences people actually employ, but here one must surely worry about the possibility that we find ourselves immersed in an irrational society that employs without hesitation fallacious reasoning.

## Reasoning to the Best Explanation

I haven't addressed the question of whether one could defeat skepticism with the array of argument types Jonathan wants the skeptic to allow us. As Hume (1888, p. 212) so eloquently argued, we don't have a hope of getting justified belief about the external world if all we can rely on is memory and enumerative induction (where the premises of our inductive arguments are restricted to descriptions of correlations between experiences). I have argued elsewhere (1980) that there may be no legitimate form of reasoning to the best explanation understood as an alternative to inductive reasoning. I have further argued (1992) that even if we suppose that reasoning to the best explanation is an independent and legitimate argument form, it is far from clear that we can use it to arrive at the truths of common sense. In particular, it is hard to see how Berkeley's hypothesis that God directly produces in us sensations that come and go in certain predictable ways is inferior to the hypothesis that it is material objects playing that causal role. Berkeley himself claimed that his view was simpler. While he posited just two kinds of things – minds and ideas/sensations – the materialist posits three – minds, ideas/sensations, and material things.<sup>10</sup>

So again, *if* we give ourselves access to past experience through memory, enumerative induction as a way of projecting past correlations found between experiences into the future, and the *prima facie* plausibility of the claim that there are causal explanations for the occurrence of contingent phenomena, and, perhaps, even that we know somehow that simpler, more comprehensive explanations are true more often than complex, less comprehensive explanations, what are the prospects of successfully arguing for the commonsense view of the world over skeptical alternatives? Leaving aside the threat of Berkeley's idealism, a great deal depends on what we take propositions about the physical world to assert. If, for example, we embrace some form of representative realism and view common sense as vindicated only if there exist objects that are in some sense accurately *represented* by sense experience, then in addition to the competitors described by exotic skeptical scenarios, one must worry about the now familiar, but at one time quite disconcerting, fact that science *itself* seems to indicate that the physical world is quite different from anything we took it to be.<sup>11</sup> The solid table is mostly space. Its straight edges, when seen under high magnification, are decidedly crooked. The color that we think is *on* the object disappears under that same magnification. Although we are now used to these discoveries, one might well conclude that science has already undermined the common sense of *representative* realism.

On the other extreme, we might adopt Hylas's last suggestion (in the second of the *Dialogues*) before capitulating completely to Philonous's idealism. Hylas wondered why we couldn't think of matter as that, *whatever* it is, that is causally responsible for sensations coming and going in the familiar ways they do. If encouraged to pursue this idea, he would probably have been willing to offer further *negative*, and also very general *causal*, characterizations of these material things – they are not minds, there are many discrete objects, each with its own causal powers and so on. But on the view, our concept of a physical object involves no characterization of its *intrinsic*, *non-relational* properties. The world of external things is exhausted, both conceptually and epistemically, by their causal powers. On *this* conception of physical objects,

the prospects of avoiding skepticism are far brighter, but only because we require so little of the causes of experience in order for the beliefs of common sense to be vindicated.<sup>12</sup>

Whatever answer we give to these conditional questions, however, I want to stress one last time that we are only postponing the skeptic's final offensive. We will be asked about the truth of the antecedents of these conditionals, and we had better be prepared to indicate what we take their truth makers to be and how we think we can get access to them.

## Notes

- 1 The fact that I refer to this justification as ideal might already be a warning that Jonathan (and others) may characterize its requirements as exotic.
- 2 On Williamson's view evidence is identified with knowledge. A perception of a computer provides a *context*, he thinks, in which one can know that the computer exists. That the computer exists is now part of our evidence and it trivially entails its own truth.
- 3 So, for example, the hypothesis that I am dreaming right now is a competitor to the proposition that there is a computer before me insofar as the truth of that hypothesis (together with the evidence in my possession) makes very unlikely the truth of the proposition that the computer exists. Alternatively, one might restrict competitors to contraries and include the negation of the proposition under skeptical attack as a conjunct of the description of the skeptical scenario.
- 4 If we are *not* foundationalists, we probably won't take the problem of skepticism seriously. The skeptic knows full well that skeptical scenarios are not going to cohere well with what we believe. I know from conversation with him that Jonathan isn't so sure that the skeptical problems disappear from within the framework of a coherence theory.
- 5 The closure principles states that if one is justified in believing P and knows that P entails Q then one has justification for believing Q. The principle of closure has been called into question, but the bottom line is that if a philosopher advances a view that forces us to reject closure, that should be taken as a *reductio* of that philosopher's view.
- 6 Again, for a discussion of reasons to reject global access requirements, weak and strong, for justification see Fumerton (1995, chapters 3 and 4).
- 7 One of the most interesting debates that has clear implications for the internalism/externalism controversy can be found in Keynes (1921) and Russell (1948, part V).
- 8 Notice that the Keynesian approach to understanding probability does not require that for one to be inferentially justified in believing P on the basis of E one must know or be able to formulate general principles of probability. One might be able to see the connection between particular propositions without seeing how to generalize. An analogous point holds of entailment. One can see that P entails Q without being able to see that entailment as an instance of *modus ponens*, *modus tollens* or any other general *kind* of entailment.
- 9 Vogel's (1990) defense of reasoning to the best explanation as the way in which to respond to the skeptic is by far the most careful and well developed in the literature.
- 10 Of course, a great deal depends on how one counts kinds. The mind of God, while a mind, isn't a whole lot like our minds.
- 11 See Eddington (1929) and Russell (1972, lecture IV) for further discussion of the idea that science undermines common sense.
- 12 For a defense of the claim that we do understand claims about the physical world this way, see Fumerton (1985, 2002, chapter 6).

## References

- Brewer, B. (1999) *Perception and Reason*. Oxford: Oxford University Press.
- Chisholm, R. (1977) *Theory of Knowledge*, 2nd edn. Englewood Cliffs, NJ: Prentice-Hall.
- Eddington, A.S. (1929) *The Nature of the Physical World*. New York: Cambridge University Press.
- Feldman, R. and Conee, E. (1998) The generality problem for reliabilism. *Philosophical Studies* 89: 1–29.
- Fumerton, R. (1980) Induction and reasoning to the best explanation. *Philosophy of Science* 47: 589–600.
- Fumerton, R. (1985) *Metaphysical and Epistemological Problems of Perception*. Lincoln: University of Nebraska Press.
- Fumerton, R. (1992) Skepticism and reasoning to the best explanation. *Philosophical Issues* 2: 149–169.
- Fumerton, R. (1995) *Metaepistemology and Skepticism*. Lanham, MD: Rowman & Littlefield.
- Fumerton, R. (2002) *Realism and the Correspondence Theory of Truth*. Lanham, MD: Rowman & Littlefield.
- Fumerton, R. (2004) Achieving epistemic ascent. In *Sosa and His Critics*, ed. J. Greco. Oxford: Blackwell.
- Goldman, A. (1979) What is justified belief? In *Justification and Knowledge*, ed. G. Pappas. Dordrecht: Reidel.
- Goldman, A. (1986) *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.
- Goldman, A. (1988) Strong and weak justification. *Philosophical Perspectives* 2: 51–69.
- Huemer, M. (2002) Fumerton's principle of inferential justification. *Journal of Philosophical Research* 27: 329–340.
- Hume, D. (1978) *A Treatise of Human Nature* (1888), ed. L.A. Selby-Bigge. Oxford: Oxford University Press.
- Keynes, J. (1921) *A Treatise on Probability*. London: Macmillan.
- Russell, B. (1948) *Human Knowledge: Its Scope and Limits*. New York: Simon and Schuster.
- Russell, B. (1972) *Our Knowledge of the External World*. London: Allen & Unwin.
- Vogel, J. (1990) Cartesian skepticism and inference to the best explanation. *Journal of Philosophy* 87: 658–666.
- Williamson, T. (2000) *Knowledge and Its Limits*. Oxford: Oxford University Press.
- Yourgrau, P. (1983) Knowledge and relevant alternatives. *Synthese* 55 (2): 175–190. Special Issue: *Justification and Empirical Knowledge*, ed. R. Audi.

## Chapter Six

# Are Intellectually Virtuous Motives Essential to Knowledge?

Some epistemologists have argued that knowledge is virtuously motivated true belief, where the relevant virtues are identified with refined intellectual traits like open-mindedness, intellectual courage, or love of knowledge. Jason Baehr argues that this view, although inspiring, can't be right. It sets the bar for knowledge too high, and thus rules out many examples of what we would ordinarily classify as knowledge. Linda Zagzebski agrees that not all instances of knowledge require such motivation, but that there is still an important connection between knowledge and virtuous motivations. On Zagzebski's view, we pre-reflectively desire the truth. When we become reflective, we recognize our desire for the truth and become aware that we trust ourselves. Epistemic conscientiousness is the reflective attempt to satisfy our desire for the truth. In some important sense of "knowledge," to know just is to achieve the truth through an exercise of epistemic conscientiousness. Thus, virtuous motivation is required for knowledge in this sense, and paradigmatic cases of knowledge are of this sort.

## Knowledge Need Not Be Virtuously Motivated

*Jason Baehr*

As far back as Socrates and Plato, philosophers have been occupied with trying to identify the essential features of important concepts like justice, love, virtue, beauty, and knowledge. Contemporary epistemologists are no exception. Much ink has been

shed in epistemology in the last 50 years attempting to identify the essential ingredients of knowledge. One such attempt has come from virtue epistemologists, who give a central role to the concept of intellectual virtue in their reflection on knowledge and related dimensions of the cognitive life.<sup>1</sup>

One major group of virtue epistemologists conceives of intellectual virtues on the model of moral virtues, that is, as intellectual *character* traits like reflectiveness, open-mindedness, fair-mindedness, carefulness and thoroughness inquiry, and intellectual courage, intellectual honesty, and intellectual rigor.<sup>2</sup> According to these authors, intellectual virtues are psychologically rooted in something like a “love of truth” or a firm and abiding desire for knowledge. Thus to be open-minded or intellectually courageous or intellectually thorough, say, is to be disposed to engage in cognitive activity characteristic of the virtue in question *out of* such a desire.<sup>3</sup>

As indicated, some proponents of character-based virtue epistemology have sought to defend a virtue-based account of knowledge according to which knowledge is (roughly) virtuously motivated true belief or true belief with its basis in something like a love of truth or knowledge.<sup>4</sup> Our concern in this debate is not with the overall plausibility of such accounts. It is rather with the narrower question of whether knowledge must be virtuously motivated; or, alternatively, whether it is possible for a person to acquire knowledge without possessing intellectually virtuous motives. I shall be arguing that this is possible and thus that knowledge need not be virtuously motivated. The discussion will, I hope, shed some important light on two importantly different dimensions of human cognition. It will also illuminate the prospects of virtue epistemology, an issue I shall return to toward the end of the essay.

## I

Before turning to the main argument, a bit more reflection on the guiding question is in order. First, the question is not whether knowledge *tends* to be virtuously motivated or is *usually* virtuously motivated. Rather, it is whether intellectually virtuous motivation is essential to or required for knowledge – whether it is at least *possible* to acquire knowledge without being virtuously motivated. A second and related observation is that my burden here is relatively light. To prove my case, I need only make it plausible that in a *single* case knowledge is or can be acquired in the absence of virtuous motives. While I intend to do considerably more than this, the fact remains that the greater burden lies with one who claims that knowledge *must* be virtuously motivated. Third, it would be helpful, to be sure, if we had a more refined account of what “virtuous intellectual motivation” amounts to. We have said that to be virtuously motivated is to be motivated by something like a desire for truth or knowledge. But what is the full range of psychological states that might qualify as intellectually virtuous motives? This is no easy matter to sort out, and attempting to do so here would take us well beyond the allotted space. This is not a serious problem, however, for the arguments that follow are consistent with a wide variety of ways of thinking about what exactly intellectually virtuous motives amount to (more on this below).

## II

Let “ $K \rightarrow VM$ ” be the thesis that knowledge must be virtuously motivated. In the present section, I identify two sets of considerations which suggest that this thesis is too restrictive, the second being more comprehensive and formidable than the first.

The first problem with  $K \rightarrow VM$  is that it implausibly limits the class of knowers to those (a) who are personally *praiseworthy* or *admirable* and (b) who possess a capacity for rational *reflection* or *control*. It is widely accepted and very plausible that a love of truth and related motives are personally admirable or praiseworthy; indeed, many virtue epistemologists treat such motives as a fundamental virtue-making quality.<sup>5</sup> A similar point holds for the claim that the possession of intellectually virtuous motives requires a capacity for rational reflection or control. This requirement stems from the fact that the motives in question instantiate a kind of genuine intellectual character excellence, which is a type of excellence that demands a certain kind or degree of rational reflection or control.<sup>6</sup> Accordingly, if knowledge requires intellectually virtuous motivation, then one can know only if one is personally praiseworthy and has the relevant cognitive capacity.

This claim is highly counterintuitive. The bar for knowledge simply does not appear to be that high. I will not attempt to work out exactly whom this conception rules in and whom it rules out. I will note, however, that at a minimum it dictates that animals and small children cannot acquire any knowledge. For some this will by itself be sufficient grounds for concluding that  $K \rightarrow VM$  is too demanding.

A second and more urgent problem with  $K \rightarrow VM$  is that it flies in the face of a wide range of specific instances of knowledge. This includes many clear cases of immediate perceptual knowledge: for example, my belief that there is (or seems to be) a computer monitor before me, that my children are making noise in the other room, that my coffee has a pleasant bitter taste, and so on. Surely these beliefs amount to knowledge. And yet it does not seem plausible to think that their formation involves the manifestation of any intellectually virtuous motives. Moreover, while immediate perceptual knowledge poses the most obvious problem for  $K \rightarrow VM$ , troubling cases of introspective, memorial, and even a priori knowledge can also be identified. My belief that my left ankle is sore, that I am anxious about meeting a deadline, that I ate a banana for breakfast, that I slept in my own bed last night, that seven plus five equals twelve, and that there are no four-sided triangles, are all clear instances of knowledge, and yet it is implausible to think that in forming or maintaining these beliefs I manifest anything that can reasonably be thought of as intellectually virtuous motives.

The foregoing considerations, as brief and straightforward as they are, provide strong *prima facie* support for the conclusion that knowledge without virtuous motivation is possible, and indeed, that it is actual and even quite pervasive. I will, then, allot most of the remaining space to a consideration of possible objections and replies.

## III

There are at least four main lines of objection that might be raised against the foregoing arguments and in support of  $K \rightarrow VM$ . I shall explain and respond to each one in turn.

It might be objected, first, that the cases in question do not amount to genuine knowledge. It is true that some historical philosophers (e.g., Plato and Aristotle) held

views of knowledge according to which many of the beliefs countenanced in the previous section would not amount to genuine knowledge. Thus a defender of  $K \rightarrow VM$  might appeal to these views in an effort to get around the argument in question.

This is not a very promising objection. First, very few philosophers today would be prepared to embrace the Platonic or Aristotelian epistemology and metaphysics that underwrite the objection.<sup>7</sup> Second, it is not at all clear that these views really support  $K \rightarrow VM$ . That is, it is far from obvious that even if these views were correct, knowledge would be impossible in the absence of intellectually virtuous motives.<sup>8</sup> Third, it is important to be clear that the cases in question not only satisfy modern and contemporary criteria for knowledge but in fact are among the *clearest instances* of knowledge understood in a broadly modern or contemporary way. Like Descartes's *cogito* and Moore's conviction that he has hands, they are widely embraced as dialectical starting points. To the modern and contemporary mind, if any of our beliefs amount to knowledge, these beliefs do.<sup>9</sup> I shall, then, assume that the real challenge for a defender of  $K \rightarrow VM$  is to show, not that the beliefs discussed in the previous section fail to qualify as knowledge, but rather that these beliefs involve intellectually virtuous motives.

A second objection contends that in the cases at issue, the knowing subject is virtuously motivated because she is operating *as an intellectually virtuous person would* under similar conditions.<sup>10</sup> There is no doubt that persons marked by qualities like intellectual carefulness, intellectual thoroughness, intellectual honesty, and intellectual rigor form basic sensory and related beliefs in the very spontaneous and unreflective way alluded to above. But not everything an intellectually virtuous person does is *characteristic* of intellectual virtue. When an intellectually virtuous person goes to sleep, checks the mail, eats lunch, or goes for a bike ride, she does not typically do so *as an intellectually virtuous person*. For the activities in question typically do not *implicate* or draw upon the relevant character virtues. The problem with the objection is that the same appears to be true of the cognitive activity that generates the relevant sensory and related beliefs. An intellectually virtuous person does not form these beliefs *qua* or *as an intellectually virtuous person*. In forming these beliefs, the person's volitional or motivational character is not engaged; or, to the extent that it is engaged, the corresponding psychological activity lacks the kind of admirable or praiseworthy quality characteristic of intellectual virtue.

A third and somewhat more promising objection claims that in the cases in question intellectually virtuous motives are present *but only at a low level*.<sup>11</sup> I do not wish to deny that in some of the cases at issue the belief might have a motivational or volitional aspect to it. For instance, my belief that seven plus five equals twelve might be the result of my *wondering* about the sum in question and *attending* to the relevant logical or mathematical relations. Similarly, basic sensory belief formation is sometimes motivated by a kind of interest in or curiosity about the states of affairs in question.

One problem is that even in cases of this sort, it is doubtful that the motivation is really intellectually virtuous. After all, animals and small children apparently can wonder, attend to, and be interested in or curious about a wide range of things and proceed to form certain beliefs as result. While we might be prepared to regard these beliefs as knowledge, we would not be willing to bestow upon their subjects the kind of praise or admiration we give to those with virtuous intellectual character.

A more serious problem, however, is that there remains a vast range of putative instances of knowledge in which the ascription of even a low-level psychological motivation seems extremely implausible. Suppose I am sitting around a seminar table with a group of



graduate students and one of the students suddenly drops his laptop on the hardwood floor; or imagine that as we are talking the lights in the room suddenly go out.<sup>12</sup> Here I will immediately, without any sort of attention or curiosity whatsoever, form a corresponding sensory belief (e.g., a belief to the effect that a loud noise has just occurred or that the lighting in the room has changed). This belief is likely to qualify as knowledge; and yet it seems obviously false that it might be the product of any virtuous motives.<sup>13</sup>

This point can be brought into further relief by a consideration of cases in which a person acquires knowledge despite possessing intellectually *vicious* motives. Imagine a scientist S whose claim to fame is a certain heretofore empirically very impressive and successful theory T. While working in the lab, S encounters data from one of his highly reliable instruments that threaten to refute T. S might initially be incredulous. He might intensely desire that the data be false or misleading. He might even attempt to disbelieve his own eyes or to deceive himself concerning the data's devastating impact on T. Nevertheless, the data might be so clear and indisputable as to make an acceptance of them unavoidable. While it is not difficult to think that S might acquire knowledge of the relevant information, such knowledge surely need not be the product of intellectually virtuous motivation. Instead it might be the result of the more or less natural, automatic, or mechanical operation of S's cognitive endowment.<sup>14</sup>

This case and cases like it underscore an important distinction between two different dimensions of human cognition. The suggested picture is one according to which the operation of human cognitive faculties is in a certain respect *independent* of agency, volition, or psychological motivation. At a certain level these faculties operate naturally or automatically. This is not to deny that the rudimentary cognitive operations in question can be harnessed or cultivated by cognitive agents. On the contrary, on one very plausible way of understanding intellectual character and virtues, these are things that emerge from an agent's engagement with and cultivation of the more mechanistic side of her cognitive nature.<sup>15</sup> What we have seen, however, is that the divide between the mechanistic and volitional dimensions of human cognition fails to correspond to the divide between unknowing and knowing, and more specifically, that untutored or rudimentary cognitive processes are capable of generating knowledge.

Before turning to consider a final objection, it is worth observing that even if the relevant cases were always to involve a low-level cognitive motivation (even motivation that could reasonably be considered intellectually virtuous), while this might vindicate  $K \rightarrow VM$ , the thesis itself would be of minimal epistemological significance. We noted above that the significance of  $K \rightarrow VM$  is tied closely to the viability of a virtue-based account of knowledge: if  $K \rightarrow VM$  is true, this would seem to bode very well for such an account and thus also for the corresponding variety of virtue epistemology. However, for reasons I have discussed elsewhere and do not have the space to elaborate on here, a virtue-based account of knowledge will be viable only if intellectually virtuous motives must play a *significant* role in the formation or maintenance of any known belief, and, more specifically, only if knowledge requires getting to the truth *on account of* or in a way that is *creditable* to the motives in question.<sup>16</sup> The problem is that even if intellectually virtuous motives were to play the kind of minimal, low-level, or background role we have been considering, they clearly would not play this much stronger and more salient role. Therefore, even if it were true that knowledge is always virtuously motivated in the relevant low-level way, the epistemological significance of this fact would be questionable at best.

A fourth and more modest objection to the foregoing case against  $K \rightarrow VM$  is that while virtuous motivation may not be required for knowledge *simpliciter*, it is required for a certain *high-grade variety* of knowledge. I have no interest in ruling out this possibility here. I will point out, however, that for it to be made plausible, considerably more would need to be said by way of support. Specifically, it would need to be shown that there exists a high-grade *pre-theoretical* or *theory-neutral* concept of knowledge according to which the type of knowledge in question requires the possession of intellectually virtuous motives. This is because, in the absence of such a concept, the resulting account of “high-grade knowledge” is bound to appear stipulative and artificial, thereby casting doubt on its overall value or significance.<sup>17</sup> It is true that some epistemologists have gestured at a distinction between high-grade and low-grade knowledge. Ernest Sosa, for instance, regularly appeals to a distinction between “animal” and “reflective” knowledge.<sup>18</sup> What remains unclear, however, is whether it is possible to acquire the relevant kind of high-grade or reflective knowledge without possessing intellectual virtuous motives. Elsewhere I have offered reasons for thinking that it may be.<sup>19</sup> If this is correct, the distinction in question does not favor even a chastened version of  $K \rightarrow VM$ .

We have considered a wide range of reasons for doubting that knowledge requires intellectually virtuous motivation. Despite this negative conclusion, it is important not to drive too deep a wedge between the concept of knowledge and that of intellectual virtue or intellectually virtuous motivation. For the fact remains that much of the knowledge that we as human beings care about most, unlike knowledge of our immediate surroundings or of what we had for breakfast, is *difficult* to come by and indeed makes significant demands on our intellectual *character* or on us as cognitive *agents*. Thus we must bear in mind that in many particular (and important) cases, acquiring knowledge *will* require the possession of intellectually virtuous motives.<sup>20</sup> This does not vindicate the claim that knowledge per se requires virtuous motivation. Nor does it support the attempt to offer a virtue-based analysis of knowledge.

## IV

I turn in this final section to consider the implications of the foregoing case against  $K \rightarrow VM$  for the prospects of virtue epistemology. The argument implies, first, that virtue-based analyses of knowledge are in trouble. For, again, these are attempts to specify the necessary (and sufficient) conditions for knowledge, and we have seen that intellectually virtuous motivation is not required for knowledge. Moreover, while I cannot elaborate on the point here, this in turn suggests, contra the aspirations of some virtue epistemologists, that the concept of intellectual virtue or virtuous motivation does not merit a central and fundamental role within traditional epistemology. In short, this is because traditional epistemological questions and debates (e.g., skepticism, rationalism vs. empiricism, foundationalism vs. coherentism, internalism vs. externalism, and so on) are mainly limited to questions and debates about the *essential* or *defining* features of knowledge or epistemic justification (features, again, which do not include virtuous intellectual motivation). The upshot is that approaches to epistemology aimed at having a revolutionary or transformative effect on traditional epistemology are unlikely to succeed.<sup>21</sup>

But this does not spell doom for virtue epistemology as a whole, for it leaves open at least two additional theoretical approaches. First, it allows for the possibility that the

concept of intellectual virtue or intellectually virtuous motives might play a *peripheral* or *auxiliary* role in connection with one or more traditional epistemological issues or debates. Second, it does nothing to undermine the rather plausible and attractive possibility that intellectual virtues and their role in the intellectual life can *in their own right* be fruitfully explored by epistemologists. In this way it leaves entirely untouched the prospects of what I have elsewhere referred to as “autonomous” virtue epistemology.<sup>22</sup>

## Notes

- 1 For overviews of the field, see Battaly (2008), Greco and Turri (2009), or Baehr (2004).
- 2 These are so-called “virtue responsibilists” or proponents of “character-based” virtue epistemology. The other main approach to virtue epistemology is known as “virtue reliabilism” or “faculty-based” virtue epistemology. Defenders of this approach tend to conceive of intellectual virtues as reliable cognitive *faculties* like memory, vision, and introspection. For more on this distinction see Baehr (2006b).
- 3 See, for example, Zagzebski (1996), Montmarquet (1993), Roberts and Wood (2007), and Baehr (2011).
- 4 The main figure here is Zagzebski (1996, 1998). But see also Fairweather (2001) and Napier (2009).
- 5 See works cited in note 3 above.
- 6 See Zagzebski (1996) and Baehr (2011, chapter 2) and Baehr (forthcoming). See also Watson (2004, pp. 1–2). I leave open the exact nature of the reflection in question.
- 7 This includes, for instance, Plato’s doctrine of the forms and Aristotle’s theory of abstraction.
- 8 See Baehr (2006a) for more on this point.
- 9 I develop this point in more detail in (2011, p. 40) and (2006a).
- 10 This argument is put forth, if not ultimately defended, by Zagzebski in (1996, pp. 279–280).
- 11 Zagzebski defends a reply of this sort in (1996, p. 280). A similar claim is defended in (Napier, 2009, chapter 3).
- 12 I discuss similar cases in my (2006a) and (2011, pp. 43–44).
- 13 This is not to deny that some beliefs might be formed in a way that is at once spontaneous and intellectually virtuous, for example, such that they are a product of a cultivated doxastic habit. See my (2011, chapter 4) and (2006b) for more on such cases. The beliefs we are concerned with at present, by contrast, are as natural, mechanical, and untutored as they come.
- 14 Of course, a person might acquiesce to the relevant evidence out of a countervailing sense of epistemic duty, say; however, this is not how I am imagining the sort of case at issue (nor do we need to appeal to any such sense in order to think of the relevant beliefs as knowledge).
- 15 See my (2011, chapter 2) for more on this point.
- 16 See my (2006a) and (2011, pp. 42–43). This is essential to dealing with the Gettier problem and related issues.
- 17 For more on this point, see Baehr (2008).
- 18 See his (2007).
- 19 See my (2006a) or (2008).
- 20 For more on this, see my (2006b) or (2011, chapter 4).
- 21 Elsewhere I have referred to these as instances of “conservative” virtue epistemology, since they appeal to the concept of intellectual virtue to deal with problems and questions within the epistemological tradition. For more on this and related points, see my (2008) or (2011, chapters 3 and 10).
- 22 See my (2011, chapter 10) for a more extensive development of these points.

## References

- Baehr, J. (2004) Virtue epistemology. In *Internet Encyclopedia of Philosophy*, ed. J. Feser. [www.iep.utm.edu/virtueep/](http://www.iep.utm.edu/virtueep/) (last accessed March 9, 2013).
- Baehr, J. (2006a) Character in epistemology. *Philosophical Studies* 128: 479–514.
- Baehr, J. (2006b) Character, reliability, and virtue epistemology. *Philosophical Quarterly* 56: 193–212.
- Baehr, J. (2008) Four varieties of character-based virtue epistemology. *Southern Journal of Philosophy* 46: 469–502.
- Baehr, J. (2011) *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology*. Oxford: Oxford University Press.
- Baehr, J. (forthcoming) The cognitive demands of intellectual virtue. In *Knowledge Virtue and Action*, ed. D. Schweikard and T. Henning. London: Routledge.
- Battaly, H. (2008) Virtue epistemology. *Philosophy Compass* 3: 639–663.
- Fairweather, A. (2001) Epistemic motivation. In *Virtue Epistemology: Essays on Epistemic Virtue and Responsibility*, ed. A. Fairweather and L. Zagzebski. New York: Oxford University Press.
- Greco, J. and Turri, J. (2009) Virtue epistemology. In *Stanford Encyclopedia of Philosophy*, ed. E. Zalta. <http://plato.stanford.edu/entries/epistemology-virtue/> (last accessed March 9, 2013).
- Montmarquet, J. (1993) *Epistemic Virtue and Doxastic Responsibility*. Lanham, MD: Rowman & Littlefield.
- Napier, S. (2009) *Virtue Epistemology: Motivation and Knowledge*. New York: Continuum.
- Roberts, R. and Wood, J. (2007). *Intellectual Virtues: An Essay in Regulative Epistemology*. Oxford: Oxford University Press.
- Sosa, E. (2007) *A Virtue Epistemology: Apt Belief and Reflective Knowledge*. Oxford: Oxford University Press.
- Watson, G. (2004) *Agency and Answerability*. Oxford: Oxford University Press.
- Zagzebski, L. (1996) *Virtues of the Mind*. Cambridge: Cambridge University Press.
- Zagzebski, L. (1998) What is knowledge? *The Blackwell Guide to Epistemology*. Malden, MA: Blackwell.

# Knowledge and the Motive for Truth

Linda Zagzebski

## 1 The Motive for Truth in Our Epistemic Lives

I assume that a self-conscious being is both conscious of the world and conscious of itself being conscious of the world. Because we are self-conscious, we reflect upon our own conscious states, not because we are especially interested in ourselves, but because we think that in doing so we can monitor and improve the connection between those states and their objects in the world. The material upon which we reflect is what we find in our pre-reflective consciousness – our memories, pre-reflective beliefs, and emotions. It also includes trust in the natural attunement of our faculties to reality. Trust is as much a part of our basic endowment as our faculties of perception and reasoning. Our pre-reflective trust is one of the things upon which we reflect when we attempt to monitor the relation between our conscious states and the world.

When we reflect, we realize that we have no non-circular way to tell that our faculties have anything to do with the way the world is, so either we turn our pre-reflective trust into reflective trust, or we become skeptics. My view is that the right response to epistemic circularity is to trust reflectively. The point of reflection is to increase the trustworthiness of our faculties, but we can only do that by using those same faculties in an especially careful and directed way. Reflection makes the connection between our faculties and the world more accurate by increasing the coherence of the outputs of those faculties. If we were living in an Evil Genius world, being conscientious would not increase the trustworthiness of our faculties. In fact, if we were right in our pre-reflective state only a small percentage of the time – say 10 percent – then careful reflection on the outputs of our faculties could not increase their reliability sufficiently to make them trustworthy. The best we can do is to be reflective, but since most of what we reflect about is, or depends upon, what is unreflective, being reflective is only helpful if we can generally trust our unreflective selves. We must think, then, that our perceptual experiences are generally veridical and our pre-reflective beliefs are generally true. If they were not, then when we are conscientious in the use of our faculties, it would not be reasonable to think that the outcome is truth.

Reflection brings fully to consciousness the use of our faculties and their connection with their ends. The natural end of belief is truth. To believe is to think something is true. When we reflect, we think about the way in which we acquired a belief, or we think about its support from other beliefs and experiences, or the beliefs and experiences of others whom we trust upon reflection. We think that such reflection will lead us to do a better job of having true beliefs. What I mean by *epistemic conscientiousness* is doing reflectively what we do unreflectively. What is added is an awareness of our epistemic end – the truth, and the attempt to reach that end as well as we can. The motive for truth does not appear only at the reflective level; it was there all along. But it becomes an object of awareness at the reflective level. The epistemically conscientious person has a reflective motive for truth and reflectively guides her epistemic behavior by trying her best to reach the end of that motive.

Aquinas says we always act “under the aspect of good” (*De Veritate* q. 22, a.1), and I would say that, similarly, we always believe under the aspect of true. We are not usually aware of the way we think that something we are doing aims at the good, but when we reflect about our acts, we bring into consciousness the end of good and the way in which what we are doing does or does not reach that end. Similarly, when we reflect about our beliefs, we bring into consciousness the natural end of the true and the way in which what we are doing cognitively does or does not succeed in reaching that end. The phenomenon of epistemic circularity reveals that even though our epistemic end is independent of our consciousness, the only way we can tell that we have reached our epistemic end is by the reflective use of our faculties consciously aimed at truth.

Reflection shows us that other people are often better at getting truth than we are ourselves. The human community has developed norms of reasoning by reflection upon the connection between what we do cognitively and what we later, upon reflection, judge as successful. These norms have been codified into rules. The human community has also identified intellectual traits that we reflectively judge to make us more successful at getting truth or reaching other epistemic ends, such as understanding. We call these intellectual virtues. The intellectual virtues are qualities of reflective agents in their attempts to reach their epistemic ends. These traits require basic trust in our faculties, and they would not be virtuous unless our faculties were basically trustworthy.<sup>1</sup>

## 2 What Is Special about Knowledge?

For millennia reflective humans have attempted to distinguish success in getting truth from a higher-level state of getting truth. I do not know whether it is natural to desire a state higher than true belief, but such a desire certainly appears at the reflective level. Ever since Plato it has been thought that there is a state of getting the truth in a particularly good way, a way that is good enough to be worth considerable effort to obtain it. But the fact that we identify and analyze knowledge at the reflective level does not mean that knowledge only exists at the reflective level. Although knowledge is worth effort, it does not necessarily require effort. It might not even require reflection. If we look at the history of philosophy, we see a division on this issue. What Plato called “*episteme*” and Aquinas called “*scientia*” was a state that demanded considerable reflection and cognitive effort. What most contemporary philosophers call “knowledge” does not. So contemporary epistemologists typically treat simple, true perceptual beliefs in ordinary conditions as knowledge, whereas typical ancient and medieval philosophers did not. I suspect that there is no determinate answer to the question whether Plato and Aquinas differ from contemporary philosophers on the analysis of the same epistemic state, or whether ancient and medieval philosophers were simply talking about a different epistemic state than the one that has received the most attention in contemporary epistemology. But I also think that the answer to this question is not very important. If there are a number of good kinds of distinguishable states of believing the truth, it is worth calling attention to them, whether or not they all fall under the class of states of knowing.

I said above that there is a pre-reflective desire for truth as well as a reflectively conscious desire for truth. It is not natural for us to be reflectively conscious all the time, and it is not natural for us to be reflectively guided by the motive for truth all the time. Even when we are unreflective, some ways of getting the truth are better than others, and it is a good idea to call attention to the difference. Some virtue epistemologists have identified knowledge with using our faculties in a way that makes the success of getting the truth credited to the agent rather than to luck or to some external cause, and on this view, unreflective true beliefs can qualify as knowledge (see, for instance, Riggs, 1998; Greco, 2003; Sosa, 2003). My view is that getting to the truth through the reflective use of our faculties is a higher-level epistemic state than getting to the truth through the unreflective use of our faculties, but I agree that the latter is better than getting to the truth through luck or some way that is not due to the use of our faculties. So there are at least three levels of getting the truth, and there are at least two places at which we can draw the line between knowledge and a lesser state. Ernest Sosa (2007) has drawn the line in both places and calls one “reflective knowledge” and the other “animal knowledge.” Put simply, reflective knowledge is reflective true belief, animal knowledge is unreflective true belief that is credited to one’s perceptual and epistemic faculties, and mere true belief that is not credited to one’s faculties is not knowledge.

I have no objections to drawing the distinction between knowledge and lesser states this way as long as reflective knowledge does not disappear from epistemological discussion. Given what I have said so far, I think we can conclude that being epistemically reflective is a good thing. My view is that it not only makes us more likely to reach our epistemic end, it is essential to self-governance. When we believe reflectively, we own our beliefs in a way that parallels our ownership of the acts we perform reflectively. We not only make it more likely that our states will bear the right relation to an external world, we also make it more likely that we will have a unified self. I think, then, that the

value of believing reflectively is undeniable, but as I have said, I cannot see anything that forces us to connect that state with knowledge and I doubt that there is a single state that legitimately falls under the heading “knowledge.”

There are many different good ways to believe. We can be good by satisfying internalist norms. We can be good by having the proper external relation to the ground of truth. We can be good by having intellectual virtues. We can be good because we are fortunate enough to have a properly functioning endowment in a benign world. The *particular* way our belief states are good in a state of knowledge is partly a matter of theoretical decision, but our decision is guided by certain theoretical desiderata. I suggest that we want our definition to capture our most important epistemic values, we want it to have roughly the same extension as other contemporary definitions, and we want to avoid classic objections to previous definitions such as Gettier problems. I would also add a practical aim: We want to find a definition that is not only theoretically explanatory, but is also practically useful, one that has a direct connection with the things we can do ourselves to achieve knowledge.

### 3 Knowledge as a State Acquired by a Conscientious Agent

I propose that knowledge is the epistemically conscientious attainment of truth. It is a state in which we get to the truth by governing our epistemic lives well. Epistemic conscientiousness is the desire for truth brought to reflective awareness, accompanied by using one’s faculties as well as one can to satisfy that desire. In my short book, *On Epistemology* (2008, p. 127), I proposed the following definition:

Knowledge is belief in which the believer gets to the truth because she acts in an epistemically conscientious way.<sup>2</sup>

Epistemic conscientiousness in my sense requires caring about truth, but not necessarily caring about truth for its own sake (2008, p. 126).<sup>3</sup> The intellectual virtues are traits of intellectual character that an epistemically conscientious person would attempt to develop and would admire in others. These qualities do not necessarily require intellectual discipline. As I have said, a conscientious person has general epistemic self-trust. The phenomenon of epistemic circularity means that either a reflective person must have self-trust or succumb to skepticism. Some of the virtues restrain self-trust (e.g., attentiveness, open-mindedness, intellectual humility, intellectual fairness). Other virtues enhance self-trust (e.g., intellectual courage, perseverance, firmness). But none of these traits would be virtues if basic self-trust were not the stance of the conscientious person. That is because the reflective person is aware that she is wasting her time in being attentive, open to the views of others, courageous, persevering, and so forth, unless her faculties are basically trustworthy.

I think that this definition satisfies a number of desiderata in a definition of knowledge. It avoids Gettier problems; it identifies a feature of knowing that makes it better than mere true believing; it has the practical advantage of connecting norms of reasoning and the intellectual virtues with the good-making features of knowledge.<sup>4</sup> In *On Epistemology* I said that I would not maintain that satisfaction of the conditions given in the definition is necessary for knowledge, only that it is sufficient. I have already mentioned that the issue of whether reflection is necessary for knowledge has changed in the history of investigation of knowledge, so the connection between reflection and the definition

above needs to be examined. There is also the issue of counter-motives, or motives opposed to the motive for truth, and whether they affect the conditions for knowing. So there are at least two types of cases in which the conditions given in my definition appear to be unnecessary for knowledge. These are cases in which the believer is (a) unreflective, or (b) reflective but vicious in reaching the truth.

The first kind of case is one that I call “easy knowledge.” It is a situation in which a person forms a true belief on the basis of simple perception or memory or testimony, and does so automatically, with no reflection. In standard cases of these kinds, many philosophers claim that the person knows. For instance, I wake up in the morning, see sunlight, and believe that the sun is up, or I recall that today is my sister’s birthday and believe it without any reflection between the act of remembering and the act of believing. Or my neighbor tells me that they are going out of town. I believe what she tells me immediately, without any intermediate steps. In each case, it is tempting to say my belief state is good enough for knowledge, provided the belief is true. I know that the sun is up. I know that today is my sister’s birthday. I know that the neighbors are going out of town.

In *On Epistemology* I argued that easy knowledge is not ruled out by the definition above. An epistemically conscientious person has basic self-trust and is not continuously reflective. She need not engage in the special cognitive discipline of following norms of reasoning that require conscious attention, nor need she exemplify virtuous habits when the discipline of the virtues is not called for, but she would do so in the relevant counterfactual circumstances (2008, p. 128). The epistemically conscientious person monitors her cognitive activity, reflecting when she is aware of cues that call into question either her basic self-trust, her trust in others, or trust in her environment. A conscientious person does reflect upon her beliefs from time to time. She expects her beliefs to survive conscientious self-reflection. But it is not necessary to engage in self-reflection about every belief that would survive such reflection. To do so would be to succumb to a kind of intellectual neurosis. Managing our cognitive lives appropriately does not mean reflecting on each and every belief at the time it is formed. But if what we *would* do in relevant counterfactual circumstances is important, then what we *do* do when those circumstances arise is also important. In this way, what I do on other occasions of forming beliefs is relevant to the issue whether my belief now is an instance of knowledge.

The same response applies to the case of the viciously motivated believer. A knower need not be motivated to get truth for its own sake, and she need not be reflectively aware of the desire for truth every time she gets knowledge. She need not directly engage the motives that are the components of the intellectual virtues on each occasion in which she obtains knowledge, and she need not refer to the norms of reasoning in order to be a conscientious believer. But if knowledge is the result of the conscientious governance of our epistemic lives, the place of the virtues in conscientiousness is important. Most of the virtues aid our self-governance by acting as constraints on or enhancements of what we do naturally and unreflectively. The parallel point applies to moral virtues. It does not count against an agent’s virtue if she does not directly engage the motives of the various virtues every time she acts, but if she is virtuous, she is disposed to do so when the appropriate occasion arises. She may believe upon reflection that many classes of her acts succeed at reaching their ends when she is acting unreflectively, but she is alert to features of a situation that call for reflection and require her to exercise the discipline of the virtues. But if she acted against the virtues, we would think that counts against the fact that she is morally conscientious. Similarly, if she flagrantly violates the norms of reasoning, or acquires a belief in a way that



is close-minded, intellectually arrogant, inattentive, or sloppy, she has done something unconscientious, and consequently she has risked losing the truth. These constraints are vague because there are many virtues and many norms of reasoning, and there are many ways in which an agent can violate them. Some ways are not very serious; others are much more so. The vagueness in the notion of believing conscientiously leads to vagueness in the application of my definition of knowledge. Since there are degrees of good believing, it is probably not surprising that there are degrees of epistemically valuable states of true believing, and in this way there are degrees of knowledge.

Let me end by summarizing my position on the place of motivation in knowledge. We desire truth at the pre-reflective level, and at that level we have trust in our epistemic faculties for the end of getting truth. A reflective person makes the desire for truth conscious and acquires reflective self-trust. What I mean by epistemic conscientiousness is the reflective awareness of the desire for truth, and the reflective attempt to satisfy that desire as well as one can. I suggest that knowledge, in at least one of its senses, is the conscientious satisfaction of the desire for truth. I have argued that this definition does not rule out easy knowledge obtained from perception, memory, or testimony, but it does rule out true belief that is unconscientiously acquired through the exercise of intellectual vice or disregard of the rules of conscientious reasoning.

## Notes

- 1 Much of the argument of the foregoing section comes from my book, *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief* (2012). That book is about managing our psychic lives. It is not about knowledge.
- 2 I say that defining knowledge as reaching the truth because of intellectually virtuous activity roughly coincides with this definition, but I do not investigate the particular cases in which the definitions may come apart.
- 3 In the first chapter of *On Epistemology*, I argue that caring about anything logically commits us to caring about the truth, but it does not commit us to caring about truth for its own sake. The argument of that chapter is largely taken from Zagzebski (2004).
- 4 I discuss the ways a definition can avoid Gettier problems and the value problem in *On Epistemology* (2008, chapter 5), and argue that the definition above is one way to avoid both problems (p. 127).

## References

- Greco, J. (2003) Knowledge as credit for true belief. In *Intellectual Virtue*, ed. M. De Paul and L. Zagzebski. Oxford: Oxford University Press.
- Riggs, W. (1998) What are the “chances” of being justified? *The Monist* 81: 3.
- Sosa, E. (2003) The place of truth in epistemology. In *Intellectual Virtue*, ed. M. De Paul and L. Zagzebski. Oxford: Oxford University Press.
- Sosa, E. (2007) *A Virtue Epistemology: Apt Belief and Reflective Knowledge*, vol. 1. New York: Oxford University Press.
- Zagzebski, L. (2004) Epistemic value and the primacy of what we care about. *Philosophical Papers* 33: 353–376.
- Zagzebski, L. (2008) *On Epistemology*. Belmont, CA: Wadsworth/Broadview Press.
- Zagzebski, L. (2012) *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*. Oxford: Oxford University Press.

# Reply to Zagzebski

Jason Baehr

Zagzebski's treatment of our question is characteristically rich, innovative, and insightful. As in previous work, she resists certain standard assumptions and categories and in doing so moves the conversation forward in interesting ways. Zagzebski and I apparently agree that knowledge does not require intellectually virtuous motives in the strong sense that a belief counts as knowledge only if, in forming or maintaining it, the believer in question manifests intellectually virtuous motives.<sup>1</sup>

## I

Nevertheless, Zagzebski is committed to some kind of conceptual connection between knowledge and virtuous motivation. She argues that knowledge is "the epistemically conscientious attainment of truth" and that epistemic conscientiousness is "the desire for truth brought to reflective awareness, accompanied by using one's faculties as well as one can to satisfy that desire" (p. 143). According to Zagzebski, epistemic conscientiousness stands to intellectually virtuous motivation in the following way: *under certain circumstances*, an epistemically conscientious person will, of necessity, manifest certain intellectually virtuous motives (e.g., when having such motives would contribute to the person's attempt to reach the truth). Her view, then, is as follows: knowledge requires an epistemically conscientious attainment of a true belief; epistemic conscientiousness sometimes requires an exercise of virtuous motives; therefore, knowledge sometimes requires an exercise of virtuous motives.

This proposed connection between knowledge and intellectually virtuous motivation is limited in scope (it does not apply to every instance of knowledge); it is also indirect (it is mediated by the concept of epistemic conscientiousness). Can it withstand scrutiny? It can, only if the suggested connection between knowledge and epistemic conscientiousness is as Zagzebski describes it. If knowledge does not require epistemically conscientious cognitive activity, then we will have lost our reason for thinking that knowledge sometimes requires virtuous motivation.

Some of the more challenging cases in this regard are what Zagzebski refers to as cases of "easy knowledge."<sup>2</sup> These are cases in which a putative knower "forms a true belief on the basis of simple perception or memory or testimony, and does so automatically, with no reflection" (p. 144). I discussed several such cases in my initial essay. Zagzebski maintains that they can indeed be thought of as involving an "epistemically conscientious attainment of truth" (p. 143). Her reasoning is that they involve a kind of cognitive "monitoring," such that the believer in question is alert, at a non-reflective or subconscious level, to cues that might "call into question either her basic self-trust, her trust in others, or her trust in her environment" (p. 144). Were such cues to appear, this person would respond by manifesting the relevant intellectually virtuous motives and norms; but in the absence of such cues, her cognitive activity will be subtle and relatively unremarkable. This is sufficient, Zagzebski maintains, for reaching the truth in an epistemically conscientious way.

I agree with Zagzebski that at least some cases of easy knowledge fit this description, that is, that what might, from the outside or even from an internal conscious perspective, seem like passive or non-volitional cognitive activity in fact involves a kind of subconscious cognitive monitoring that can reasonably be thought of as manifesting epistemic conscientiousness. However, I think there are many other cases of knowledge with respect to which such an assessment is implausible.

The sort of cases I have in mind involve a kind of easy knowledge that might be dubbed “overwhelming knowledge.” This is knowledge that bears down on or overwhelms the knower. Recall, for instance, the case of a scientist confronting counter-evidence to one of his cherished theories. At the first appearance of a conflict between the data and his theory, the scientist lapses into “denial mode,” doing his best to resist the implications of the data (e.g., by thinking about other things or by hastily constructing alternative but manifestly implausible explanations of it). Nonetheless, the force of data is so strong that, despite vicious motives tempting him to draw an alternative conclusion, he is compelled to accept the data at face value and thus to repudiate his theory.

As suggested in my initial essay, it seems entirely possible that in cases like this a person might reach the truth without engaging in any kind of epistemically conscientious cognitive monitoring. Indeed, to the extent that the person’s agency is active, whether at a reflective or non-reflective level, it might be entirely *unconscientious*. While the person does acquiesce to the evidence, this could very well be attributable to the overwhelming force of the evidence together with the natural, brute, or default functioning of the person’s basic cognitive faculties. It need not – and on the surface seems positively not to – be due to any reflective or unreflective use or employment of these faculties on the part of the agent.<sup>3</sup>

A similar point holds with respect to other cases discussed or alluded to in my initial essay: for example, knowledge of a loud noise, a sudden change in lighting, a shooting pain, or the like. Here as well it seems implausible to think that any epistemically conscientious monitoring is occurring. This is not because there are any epistemically vicious motives in the psychological vicinity; it is rather because the formation of the relevant beliefs is so quick, automatic, and mechanistic that there is apparently little room for such monitoring. More importantly, even if the knowers in question were engaging in epistemically conscientious monitoring, this would not be the *reason* they form the beliefs they do or the reason these beliefs are true. Again, the reason would instead appear to be the overwhelming character of the objects of belief together with the default functioning of the agents’ cognitive equipment. It is implausible, then, to think of these cases as involving an “epistemically conscientious attainment of truth.” And yet it seems utterly clear that the beliefs in question amount to knowledge (surely I know I’m in pain when my consciousness is overwhelmed by the throbbing sensation).

How problematic are these apparent counterexamples? Might this be a situation in which it would be worth abandoning certain first-order epistemic judgments in the interest of embracing a theory with other advantages? This is difficult to answer in a decisive way, but I briefly note two reasons for thinking not: first, the cases in question are neither fluky nor artificial; and, second, their epistemic status appears unambiguous (again, it would seem absurd to suggest that I do not know that I have a splitting headache, that a loud noise has just occurred, and so on).

## II

Before turning to open the door to a certain version of the claim that knowledge requires intellectually virtuous motives, I want to draw attention to a certain feature of Zagzebski's account of knowledge. I begin with two observations: (i) On her account, the core concept (beyond truth) is that of epistemic conscientiousness; the concept of virtuous motives is secondary and derivative. (ii) Because the concept of epistemic conscientiousness is intended to capture the primary value of knowledge over and above the value of true belief, this concept bears a substantial *normative* burden: epistemic conscientiousness, as Zagzebski understands it, must be a significant epistemic *good*. My suggestion is that there is tension between (i) and (ii), and specifically, that for epistemic conscientiousness to count as a significant epistemic good, the concept of virtuous motives or intellectual virtues must be given a more prominent – indeed in some sense a primary – position.

Recall that epistemic conscientiousness, on Zagzebski's view, involves an *awareness* of our natural goal of truth as well as an *attempt* to reach that goal. But epistemic conscientiousness thus described might not be an especially epistemically good thing. I might be aware of the truth goal and make an attempt to reach it, but my attempt might be half-hearted, sloppy, careless, narrow-minded, or the like. It might, in other words, be intellectually *vicious*. Indeed, Zagzebski herself at times indicates, reasonably enough, that an epistemically conscientious attempt to reach the truth will be “careful and directed” and that it will involve using one's faculties “as well as one can” (p. 143). What this suggests is that in order for a person's reflective attempt to reach the truth to have epistemic significance or worth – for it to have the kind of value proper to knowledge – it must be marked by intellectually virtuous motives or actions, for example, by the motives and actions characteristic of traits like intellectual carefulness, fairness, honesty, openness, and so on. This suggests that Zagzebski's notion of epistemic conscientiousness involves an implicit appeal to intellectual virtues or virtuous motives – that for the relevant kind of second-order perspective or cognitive self-regulation to make an epistemic contribution, it must involve such motives or actions. If this is right, then we cannot, as Zagzebski seems interested in doing, define epistemic conscientiousness independently of the concept of virtuous motivation.<sup>4</sup> Indeed, it would appear that an account of knowledge centered on the notion of epistemic conscientiousness must, in a certain important sense, be virtue-centric.

## III

In my initial remarks I left open the possibility that intellectually virtuous motives might be necessary for a certain kind of “high-grade” knowledge, but that a defense of this claim faces a particular challenge: namely, that in order to avoid being purely stipulative or artificial, the kind of knowledge in question must have a sufficient and determinate presence in our pre-theoretical thinking about the cognitive life (again, were this not the case, it would be unclear what the account in question is supposed to be an account of). I also expressed some doubt about the possibility of meeting this challenge.

Zagzebski's rich and insightful discussion of our natural epistemic state, of the way that we reflect on and try to enhance or improve this state, and, most importantly, of the kind of *success* that is possible in this realm, seems to me to do a nice job of identifying, in a reasonably pre-theoretical or pre-philosophical way, a significant and

determinate cognitive good. Moreover, I am unaware of any good reason for denying the title of “knowledge” to such a good (provided, of course, that we allow for other types or varieties of knowledge as well). It may be possible, then, to think of the general contours of Zagzebski’s account as picking out a certain high-grade variety of knowledge. Finally, if the point made in the previous section is correct, this account, when properly formulated, will be virtue-centric. In this respect, it may turn out that a certain high-grade variety of knowledge must be virtuously motivated.

## Acknowledgments

Thanks to John Turri and Linda Zagzebski for helpful conversation and feedback on earlier drafts of this material.

## Notes

- 1 See p. 144 (section 3, paragraphs 4 and 5 above, but especially her reply). This represents an interesting (and in my view sensible) departure from Zagzebski (1996).
- 2 This is an apt label; however, it is worth noting that this understanding of “easy knowledge” is different from the one at the center of the “problem of easy knowledge” as dealt with in Cohen (2002) and elsewhere.
- 3 There is a variation on this type of case in which the person acquiesces to the evidence out of a kind of underlying epistemic conscientiousness at odds with his more immediate and epistemically dubious doxastic inclination. However, as I am thinking of such cases, the person in question does not have mixed motives in this sense (nor, I take it, is such a supposition required for it to be plausible that the person acquires knowledge).
- 4 That Zagzebski is interested in doing this is evident in her apparent interest in defining intellectual virtues as traits that an epistemically conscientious person would pursue (p. 143). The threat of circularity here should be obvious.

## References

- Cohen, S. (2002) Basic knowledge and the problem of easy knowledge. *Philosophy and Phenomenological Research* 65: 309–329.
- Zagzebski, L. (1996) *Virtues of the Mind*. Cambridge: Cambridge University Press.

# Reply to Baehr

*Linda Zagzebski*

From Baehr’s essay, I see that there is a large area of agreement between us on the issue of the place of motives in knowledge. Baehr says he does not deny that knowledge tends to be virtuously motivated. What he denies is that intellectually virtuous motivation is

necessary for every instance of knowledge, and I agree with him about that. Neither virtuous motivation nor reflective awareness is necessarily exemplified on each occasion in which the subject knows something. But Baehr denies the connection I defend between knowledge and reflection, and the connection between reflection and virtuous motivation, so my remarks will focus on these connections. There are also cases in which Baehr and I disagree on the extension of knowledge, and I will mention two such classes of cases. These cases and some of Baehr's other remarks lead me to suspect that we have different views on the point of the project of attempting to answer the question, "What is knowledge?", so I will address that briefly in my reply.

In my essay I argued that my proposal that knowledge is conscientiously reflective true belief naturally arises from the pre-theoretical notion of knowledge as the central good epistemic state, and the assumption that human beings are self-reflective. I assume we have a natural, pre-reflective aim to get true beliefs and to avoid false beliefs. But given that we are reflective beings, the basic norm of reflection – do reflectively what we already do unreflectively – tells us to make our pre-reflective aim for truth reflective, and that means to follow norms that reflection has shown to make it more likely that we will be successful at getting the truth. These norms include rules of reasoning and injunctions to acquire certain traits such as open-mindedness, intellectual attentiveness, thoroughness, and so on. My point here is that epistemic norms are not independent of what we think knowledge is. That is because it is not very helpful to be told that knowledge is X if you can't get X. We think of knowledge as something attainable by conscious effort. We think the same thing about the relationship between moral norms and the moral good. I am not denying that there are some unattainable goods, whether moral or epistemic, but the place of the idea of moral good in a moral theory connects it with the norms developed by the theory, and that is because it is not helpful to be told by a theory that the moral good is X if the theory does not also guide us in getting or reaching X. My underlying methodological principle, then, is that goods and norms are not independent. Since the basic epistemic good is knowledge, the norms for achieving it are not independent of the way we understand the nature of knowledge.

This explains both the connection between knowledge and reflectiveness, and the connection between reflectiveness and virtuous motivation. Let me start with the latter. There is already a pre-reflective motive for truth, but human beings are self-conscious, and self-conscious beings reflect about the end of their natural faculties, including the end of getting truth. The reflective person follows the norms of reasoning and attempts to acquire the intellectual virtues because doing so is doing a better job of what she does naturally. Her motive for truth is thus brought into reflective consciousness, and the norms and virtues are just those that a reflectively conscious being who desires truth adopts. A virtuous motive in my view is the pre-reflective motive for truth brought into reflective consciousness.

If we assume that knowledge is the central good epistemic state, and if we also accept that our idea of the good is theoretically connected with norms for achieving it, we can see the connection between knowledge and reflectiveness. The paradigmatic case of knowledge is true belief acquired through a process of conscientious reflection, following the norms endorsed by conscientiously reflective persons.

This does not mean that every instance of knowing is a direct consequence of conscientious reflection. Although "Be reflective" is the primary norm of self-conscious beings, we also think that people can be too reflective. Doing so is a kind of neurosis.

One thing we grasp when we are conscientiously reflective is that we do not need to be reflective all the time in order to succeed in reaching our ends. This applies to both epistemic ends and practical ends. We think that people who consciously attempt to make every act virtuous have something wrong with them. They are less good than people who are virtuous but who do not consciously reflect about the way their every act leads to a virtuous end. I think of reflectiveness and intellectual virtue the same way. A person who consciously attempts to acquire true beliefs and no false beliefs at every moment she is epistemically active is less good epistemically than an intellectually virtuous person who has learned through reflection that she does not have to be aware of her motives and epistemic processes all the time. Epistemic self-trust is rational, and a self-trusting person relies upon her faculties much of the time without exercising cognitive discipline. But a self-reflective person also monitors her cognitive and belief-forming processes, and that is why what she does in counterfactual circumstances bears on the issue of whether she is conscientious in the formation of any particular belief. Conscientiousness does not require continuous reflection, but it includes monitoring of the self, and a conscientious belief survives future conscientious reflection. Knowledge that *p* is true belief conscientiously monitored and endorsed by conscientious reflection even though it is not preceded by conscientious reflection whether *p* on each occasion in which the subject believes *p*. It is because of reflective consciousness that knowledge exists, but reflective consciousness itself leads us not to consciously reflect about beliefs of certain kinds. There are, however, two classes of beliefs that are often considered to be knowledge in contemporary discourse and which my view rules out: (i) beliefs of children too young to engage in conscientious reflection, and (ii) viciously motivated true beliefs. I think that Baehr considers typical beliefs in these categories as knowledge.

This leads me to the methodological issue of the point of the “What is knowledge?” question. I think there is a difference between Baehr and myself on the way we approach that question. I do not think the question is asking for a conceptual analysis. I also think that even though we want to avoid counterexamples, we should not think that the most illuminating answer to the question focuses on giving necessary and sufficient conditions for knowledge. It is an advantage if the account of knowledge extensionally coincides with other major accounts, but it is not necessary that it does so if I am right that the idea of knowledge is embedded in a theory that includes many other elements— the idea of virtue, other epistemic goods, and norms for belief acquisition and maintenance. My position is that the way we understand knowledge is not separable from theory construction and evaluation. I think that it will turn out that our best theoretical decision about the way to understand the project of epistemology will lead us to make some revisions in the conventional way of dividing knowledge from true belief that is not knowledge. But as I said in my essay, I do not think that the way we make that division is very important in itself. What is more important is the theoretical structure we develop. We want a theory that gives us both understanding of the domain of epistemology and practical usefulness in the conduct of our epistemic lives. We want theoretical clarity, comprehensiveness, subtlety, and simplicity, as well as guidance in becoming the kind of person who can achieve the goods identified in the theory. I suggest that the search for such a theory should precede the search for a definition of knowledge.

## Chapter Seven

# Can Knowledge Be Lucky?

The claim “knowledge cannot be lucky” is virtually platitudinous in contemporary epistemology. But there are many ways in which knowledge clearly can be lucky. It can be a matter of luck that you weren’t killed by a runaway bus just a moment ago, in which case you’re lucky to know that you’re alive. It can be a matter of luck that you caught a glimpse of the intruder, in which case you’re lucky to know that Smith was the intruder. In what sense, then, is luck anathema to knowledge? Duncan Pritchard argues that the anti-luck condition on knowledge is best understood as a *safety condition*: if your belief that P constitutes knowledge, then not easily would your belief that P have been false (when formed in the way it was actually formed). That is, in most nearby worlds when you form the belief on the same basis as in the actual world, your belief is true. Pritchard defends his view from a number of new and standard objections. Stephen Hetherington argues that knowledge can be lucky, in the sense that unsafe beliefs can constitute knowledge. Hetherington claims that a safety condition on knowledge seems appealing because of a subtle modal fallacy when considering alternative possibilities.

## Knowledge Cannot Be Lucky

*Duncan Pritchard*

### 1 Knowledge and Luck

It is commonplace in epistemology to hear people say that there cannot be lucky knowledge (call this the *anti-luck intuition*). This claim requires some unpacking, however, since epistemologists have something quite specific in mind when they put forward this thesis. In particular, they are not denying that it might well be a matter of luck that



an agent has knowledge. Imagine an agent who a few moments ago narrowly avoided being killed by a sniper (because, say, she happened to glance downwards at a very fortuitous moment). If so, then she is lucky to be alive and thus lucky to be in a position to know anything. Accordingly, if that fortuitous downward glance happened to reveal to her that she is wearing non-matching socks, and she comes to know this proposition on this basis, then it will be a matter of luck that she has this knowledge.

It can be a matter of luck that one has knowledge in other ways too. For example, it can be a matter of luck that one is in possession of the evidence in virtue of which one has knowledge. Imagine an agent who just happens to be standing outside a colleague's door at just the right moment to overhear what her colleague, who *always* speaks the truth, was authoritatively asserting on the phone (i.e., if she had passed the door a moment earlier or a moment later then she wouldn't have heard a thing). Given that she has clearly heard what her colleague was saying – for example, that so-and-so is wearing non-matching socks – and given the fact that what this colleague asserts can be relied upon, then she is in a position to come to know the relevant proposition. Accordingly, it is a matter of luck that she knows that so-and-so is wearing non-matching socks, but it is no less knowledge for that.

Clearly, then, there are some senses in which knowledge can be lucky, and no epistemologist would want to deny this. So what specific claim are epistemologists putting forward when they say that there cannot be lucky knowledge? Well, the particular claim they have in mind is that when one has knowledge then *it cannot be a matter of luck that one's belief is true*.<sup>1</sup>

Notice first that in the two cases of lucky knowledge just given it precisely *isn't* the case that the agent's belief is only luckily true. Given how the first agent formed her belief that she is wearing non-matching socks – that is, by getting a good look at them at close range – it is not a matter of luck that her belief is true, even though it is a matter of luck that she is in a position to know what she does. The same goes for the second agent. Given how she is forming her belief that so-and-so is wearing non-matching socks – that is, by clearly overhearing a reliable colleague assert that this is the case – it is not a matter of luck that her belief is true, even though it is a matter of luck that she acquires the evidence that she does which supports this knowledge (i.e., the overheard testimony of her colleague).

Consider now some cases in which it is a matter of luck that the agent's belief is true. First off, a lucky guess.<sup>2</sup> Imagine an agent who, purely on a whim, believes that the next person to come into her office will be wearing non-matching socks, and suppose that this belief turns out to be true. Clearly, she didn't know that the next person to come into her office would be wearing non-matching socks, and the natural explanation of why is that, given how she formed her belief (i.e., by guessing), it was just a matter of luck that her belief was true.

Second, consider a standard Gettier-style case. Imagine our agent sees what she takes to be a sniper in the office block across the way, with her rifle trained on her building. Accordingly, she forms the belief that there is a sniper in that office block. Now suppose that this belief is true, but that what our agent is looking at is not a real sniper (one trying to assassinate our non-matching-sock-wearing friend, say), but rather merely a cardboard cutout of a sniper, left over from the promotional materials of a movie (with the real sniper hidden behind). Does our agent have knowledge of what she believes? Intuitively not, and the natural reason for this is

that given how she formed her belief – by looking at the cardboard cutout of a sniper – it was just a matter of luck that her belief was true. Moreover, notice that she lacks knowledge on account of her lucky true belief even though, in contrast to the lucky guesser we just considered, she has a good reason for believing what she does (after all, she did see something which looked like a sniper in the window of the office building).<sup>3</sup>

Now contrast such cases where lucky true belief undermines knowledge with paradigm cases of knowledge. Imagine, for example, that our agent gets to see that her colleague is wearing non-matching socks at close hand and in optimal viewing conditions. Or imagine that our agent is told by a reliable informant that her colleague is wearing non-matching socks today. In both cases we have an agent who thereby comes to acquire knowledge. But notice that in both such cases we also have an agent whose true belief is not a matter of luck. If one believes that someone is wearing non-matching socks because one gets to see – at close hand, and in normal circumstances – that this is the case, then it is not a matter of luck that one's belief is true. And if one believes that someone is wearing non-matching socks because one is told that this is the case by a reliable informant, then it is not a matter of luck that one's belief is true.

There is thus a strong *prima facie* case that knowledge excludes luck in just this sense: if one has knowledge, then it cannot be a matter of luck that one's belief is true. Call this the *anti-luck condition* on knowledge. But how, exactly, should we understand the anti-luck condition? The devil, as always in philosophy, is in the detail.

## 2 Unpacking the Anti-Luck Condition I: Sensitivity

One way of understanding the anti-luck condition is in terms of the requirement that knowledge entails *sensitive* true belief. In order for a true belief to be sensitive, it must be such that, had what the agent believed been false, she wouldn't have believed it.<sup>4</sup> Two remarks are in order regarding how we should read this claim.

The first is to note that this is a *modal* claim, in the sense that in applying it we need to consider what is going on in relevant possible worlds. In particular, in assessing whether a true belief is sensitive we need to consider not only what the agent believes in the actual world (where her belief is true), but also what she believes in that possible world (or worlds) which is similar to the actual world in every other respect except that the target proposition (the proposition truly believed in the actual world) is false.<sup>5</sup> Call this the *closest not- $p$  world* (where  $p$  is the target proposition which is believed by the agent in the actual world).

The second remark modifies the first. For we need to qualify our understanding of the closest not- $p$  worlds by considering only those closest not- $p$  worlds in which the agent forms her belief about the target proposition in the same way as in the actual world. That is, what we are interested in is not merely whether a certain belief is sensitive, but rather whether that belief is sensitive, *given how it is formed*.<sup>6</sup>

We can illustrate these two points by considering how sensitivity, so understood, can deal with cases in which an agent lacks knowledge due to the epistemic luck in play. Take first the case of a lucky guess that was given above, where the agent believes as a result of a guess that the next person to come into her office will be wearing non-

matching socks. Clearly this is an insensitive belief, since in the closest not-*p* world – that is, the world in which the next person to enter the room is not wearing non-matching socks, but everything else, including the agent's basis for the belief (i.e., guessing), is the same – our agent will form a false belief.

Now consider the Gettier-style case we gave in which the agent sees what seems to be a sniper in the office block window and on this basis forms the belief, true as it happens, that there is a sniper in the office block that she is looking at. Again, this belief is insensitive, since in the closest not-*p* world – that is, the world in which there is no sniper in that building, but everything else, including the agent's basis for the belief (i.e., the promotional cardboard cutout of a sniper that she sees in the window), is the same – our agent will form a false belief.

Clear-cut cases of knowledge also seem to involve a sensitive belief. Imagine, for example, that our agent gets to see that her colleague is wearing non-matching socks at close hand and in optimal viewing conditions. In such a case we have an agent who thereby comes to acquire knowledge, and who is also forming a sensitive belief. For given the way that she is forming her belief, in the closest not-*p* world – that is, the world in which her colleague is not wearing non-matching socks, but everything else remains the same (including the agent's perceptual basis for her belief) – the agent will no longer believe that her colleague is wearing non-matching socks for the simple reason that she will see that she isn't.

Sensitivity can also be used to shed light on another problem in epistemology, known as the *lottery problem*. Imagine a fair lottery with astronomical odds – a billion to one, say – which has been drawn, and imagine two counterpart agents who both own a single lottery ticket. Suppose that our first agent hasn't heard the result yet but believes that she has lost because she has reflected on the very long odds involved in her winning; and suppose that our second agent also believes that she has lost, but does so purely because she has just read the results printed in a reliable newspaper.<sup>7</sup> Here's the thing. Intuitively, the first agent doesn't know that she has lost while the second agent does. But this is puzzling, since the odds in favor of the first agent being right are astronomically in her favor. Indeed, even though reliable newspapers are very careful when it comes to printing lottery numbers (for obvious reasons), nonetheless the probability that these results have been misprinted is surely higher than the astronomical probability that a ticket should win this lottery. So how then can it be that the second agent has knowledge but the first agent doesn't?

The thesis that knowledge requires sensitivity offers us a very attractive way of dealing with the lottery problem, for notice that while the second agent's belief is sensitive, the first agent's belief is not. The closest not-*p* world is where the agent concerned is in possession of a winning lottery ticket. Crucially, though, while this will be a world in which the reliable newspaper prints the winning result, it will continue to be a world in which the odds in question overwhelmingly suggest that one has lost. Thus, if one forms one's belief about whether one has lost on the basis of the odds concerned, then one will form a false belief in this world, but if one forms one's belief by consulting a reliable newspaper then one will form a true belief. It is in this sense, claim sensitivity theorists, that the first agent's belief that she has lost the lottery is only luckily true, even though the odds are massively in her favor when compared with the second agent who, by consulting the reliable newspaper, really does know that she has lost.

### 3 Unpacking the Anti-Luck Condition II: Safety

Unfortunately, the claim that knowledge entails sensitivity faces some problems. Rather than rehearse them all I will focus here on the most pressing of these difficulties, which is the problem it has with inductive knowledge.<sup>8</sup> Suppose that our agent has spent the entire day with her colleague, and throughout the day she has been continually reminded of the fact that her colleague is wearing non-matching socks. Now suppose that her colleague leaves the office to get the train home. Does our agent know that her colleague is still wearing non-matching socks? Intuitively, she does, since she has an excellent inductive basis for this (true) belief, which is her observations of her colleague throughout the day and her knowledge that she is unlikely to be changing her socks on the train journey home. But here is the problem: her belief is insensitive. For notice that in the closest not- $p$  world – that is, the far-fetched world in which her colleague has been forced to take off her socks, perhaps because she is the victim of an armed sock thief – our agent will continue to believe that her colleague is wearing non-matching socks regardless (after all, her basis for this belief – namely that she has seen her wearing these socks throughout the day and she has no reason for thinking that she has taken them off – remains constant).<sup>9</sup>

The reason why sensitivity is unable to account for the knowledge present in such cases is that it focuses on the closest not- $p$  world and ignores the wide range of close worlds where the agent believes the target proposition and believes truly (after all, in most nearby possible worlds the agent believes that her colleague is wearing non-matching socks and she is wearing non-matching socks). This is one of the main reasons why many in the literature opt for the thesis that knowledge demands not sensitive belief but rather *safe* belief.

The basic idea behind safety is that one has a true belief which could not have easily been false.<sup>10</sup> This is usually cashed-out as the claim that one has a true belief that  $p$  such that, in close possible worlds, if one continues to form a belief on the same basis as in the actual world, then one's belief continues to be true.

Firstly, notice that this principle can handle inductive knowledge. For while there is undoubtedly a far-fetched possible world where our agent's colleague has removed her socks while on the train, in all close possible worlds she will retain her socks.<sup>11</sup> Accordingly, our agent will form a belief that her colleague is wearing socks on the train which will be true in all relevant possible worlds, and hence her belief will be safe.

Safety can also deal with Gettier-style cases, for these are characteristically cases in which the agent forms (on the same basis as in the actual world) a false belief in the target proposition in a close possible world. Consider the case of the promotional cardboard cutout being mistaken for a real sniper that was described above. Although the belief so formed (that there is a sniper in the relevant office block) is true in the actual world, there will be a nearby possible world where the agent will form the very same belief on the very same basis (i.e., by looking at the cardboard cutout of the sniper), and yet forms a false belief (i.e., the nearby possible world where the cardboard cutout remains, but the sniper has been delayed and is yet to enter the building).

Safety can also handle the lottery problem, though we need to be a little more precise in how we understand the principle in order to see this. The formulation above

talks simply of the agent continuing to form a true belief on the same basis as in the actual world across nearby possible worlds. But this naturally prompts the question of how extensive this range of nearby possible worlds should be, and to what extent, if any, safety is consistent with there being some false beliefs formed within this range of possible worlds.

What is key to answering this question is that safety is capturing an intuition about our tolerance of the risk of error. In the very closest nearby possible worlds we are extremely intolerant when it comes to such epistemic risk, and so would not want to be forming any false beliefs on the target basis. In far-off possible worlds, however, we are extremely tolerant about such epistemic risk, on account of their modal remoteness. In between we have a descending scale of epistemic intolerance, from extreme intolerance to epistemic risk to extreme tolerance. With this point in mind, we need to think of safety as completely excluding false belief in the very closest possible worlds, but becoming increasingly tolerant to such falsity as one moves further away from the actual world.<sup>12</sup>

Safety so construed can handle the lottery problem. Recall that the two agents concerned form their true beliefs that their tickets have lost by, respectively, reflecting on the odds involved and reading the results in a reliable newspaper. What is crucial here is that the remoteness of error is very different in the two cases. All it takes for the first agent to form a false belief is for a few colored balls to fall in a slightly different configuration. In contrast, what is required for the second agent to form a false belief is a range of mishaps at the newspaper office (e.g., the person inputting the results, despite taking lots of care in doing so, somehow makes a mistake, a mistake that is not spotted when the various copyeditors, hired and retained for their conscientiousness in such matters, somehow collectively fail to spot the error). This is why the first agent's belief is unsafe, and hence not knowledge, while the second agent's belief is safe and hence is in the market for knowledge.<sup>13</sup>

One could say an awful lot more about how safety handles particular cases, but given that space is limited I hope that I have said enough to have motivated the thought that safety has advantages over sensitivity, and hence that it is a plausible candidate when it comes to unpacking the intuition that knowledge cannot be lucky. Put simply, the claim that knowledge cannot be lucky, when properly understood, is tantamount to the claim that there cannot be knowledge that *p* which involves an unsafe belief that *p*. Before moving on to some objections to this specific thesis, I want to make two further points.

First, it is often thought that the safety principle has a problem accounting for belief in necessary propositions.<sup>14</sup> One can see how the objection would run. Such propositions are true in all nearby possible worlds, and hence all one needs to do is happen to form a true belief in a necessary proposition in the actual world and – hey presto! – one has a belief which is necessarily safe (since there can necessarily be no nearby possible worlds where one continues to form this belief and believes falsely). Crucially, however, one does not evaluate the safety of a belief by focusing only on nearby possible worlds where the agent continues (on the same basis as in the actual world) to form a belief *in the very same proposition* as in the actual world. Rather, what one is interested in is the truth value of the belief that is formed in nearby possible worlds on the same basis as in the actual world, even when the resulting belief is not of the same proposition.

In order to see this, imagine that someone forms a true belief in a mathematical proposition – that  $2 + 2 = 4$ , say – by flipping a coin. Since there is no possible world where the proposition believed is false, there is thus trivially no nearby possible world in which the agent believes *this specific proposition* and believes falsely. But that *doesn't* mean that the belief is thereby safe, and the reason for this is that there are lots of nearby possible worlds where the agent's actual way of forming her belief – that is, by flipping a coin in order to determine mathematical truths – leads to false belief, such as the possible world where the coin toss prompts her to believe that  $2 + 2 = 5$ . The key point here is that in assessing whether a belief that  $p$  is safe, we are interested in whether the agent forms a belief in the same way in nearby possible worlds and believes falsely, but this is different from being interested in whether the agent forms a belief *that*  $p$  in nearby possible worlds and believes falsely.

The second point I want to make about safety is to emphasize that the thesis in play here is specifically the claim that this is a *necessary* condition on knowledge. There is a good reason for this, which is that aside from the anti-luck intuition we also have a strong intuition that knowledge reflects ability (call this the *ability intuition*), in the sense that when an agent has knowledge then it is to some significant degree the product of the agent's exercise of her cognitive abilities. One can draw this point out by noticing that we can formulate cases where the agent forms a true belief which is clearly not-lucky (i.e., which is clearly safe) and yet which isn't a case of knowledge because the cognitive success in question in no way reflects the agent's cognitive abilities.

Imagine, for example, that an agent was forming beliefs about what color socks the next person she saw would be wearing purely on the basis of flipping a coin. But imagine now that these beliefs are guaranteed to be true since there is, unbeknownst to our agent, a helpful demon out there whose mission it is to ensure that beliefs formed in just this way are true. Consequently, he makes it his business to see that the socks worn by the next person seen by the agent also correspond to what the agent guessed they would be.

Clearly, one cannot gain knowledge by guesswork in this way. Note, however, that the problem here isn't that the agent's beliefs are only luckily true, since given the interference of the helpful demon they are in fact guaranteed to be true, and hence can't help but be safe. Instead, the problem is that the cognitive success on display in no way reflects the agent's cognitive abilities, but rather is simply the result of the interference of the helpful demon. Put simply, although there is a perfect match between belief and fact across the relevant possible worlds (such that the agent always believes truly), there is the wrong direction-of-fit in play for knowledge, in that the facts are changing to fit with what the agent believes rather than the agent's beliefs being responsive to the facts. Given that the problem in play here does not concern knowledge-undermining epistemic luck, and given that the role of safety is to exclude such luck, it should be clear that it is not a failing of safety that it is unable to deal with such cases.<sup>15</sup> Rather, what such cases remind us is that there is more to knowledge than safe true belief. In particular, at the very least what is also required is some sort of ability condition on knowledge.<sup>16</sup> As we will see below, this point is important for our purposes since we need to be sure that a putative counterexample to the necessity of safety for knowledge is not in fact trading on something other than the anti-luck intuition.

## 4 Three Putative Counterexamples

A range of putative counterexamples have been put forward to the idea that knowledge entails safety, and we will close by considering a representative sample to explain why they fail to achieve their intended aim.<sup>17</sup> The first is due to Ram Neta and Guy Rohrbaugh (2004):

WATER: “I am drinking a glass of water which I have just poured from the bottle. Standing next to me is a happy person who has just won the lottery. Had this person lost the lottery, she would have maliciously polluted my water with a tasteless, odorless, colorless toxin. But since she won the lottery, she does no such thing. Nonetheless, she *almost* lost the lottery. Now, I drink the pure, unadulterated water and judge, truly and knowingly, that I am drinking pure, unadulterated water. But the toxin would not have flavored the water, and so had the toxin gone in, I would still have believed falsely that I was drinking pure, unadulterated water. ... Despite the falsity of my belief in the nearby possibility, it seems that, in the actual case, I know that I am drinking pure, unadulterated water.” (Neta and Rohrbaugh, 2004, pp. 399–400)<sup>18</sup>

My initial reaction to such a case is to hold that it is simply not a case of knowledge. Is it really intuitive that the agent in WATER could gain knowledge that what she is drinking is water even despite the clear epistemic unfriendliness of her environment? After all, she could so very easily have been drinking the toxin rather than the water, and *ex hypothesi* she wouldn’t have been able to tell the difference. Indeed, were our agent to discover just how lucky it was that she formed a true belief in this case, then surely she wouldn’t ascribe knowledge to herself. But of course, if this isn’t a case of knowledge then it can’t be a counterexample to the necessity of safety for knowledge.

There is, however, an obvious dialectical drawback to dismissing such a case out of hand. After all, it is clear that others have found this example compelling, and so we are in danger of merely trading opposing intuitions here. Fortunately, I think we can diagnose why someone might hold that the protagonist in WATER has knowledge, even though (so says I, anyway) she doesn’t.

Recall that we noted earlier that when epistemologists say that they want to exclude lucky knowledge they are not meaning to suggest that it can’t be a matter of luck that one is in a position to acquire knowledge. As we noted above, it could be merely a matter of luck that one happens to overhear a conversation – a moment earlier or a moment later and one would have missed it – but given that one does get to properly overhear it, one can nonetheless acquire knowledge of what was said as a result (call this example “OVERHEAR”). Safety captures this point by specifically focusing on the way in which the belief was formed. It may be a matter of luck that one overhears the conversation, but so long as one’s hearing is working fine and the voices are loud enough (and so on) then this will be a way of forming one’s beliefs which is safe, and hence knowledge conducive.

If one thought that WATER was relevantly analogous to OVERHEAR then one might be inclined to ascribe knowledge in this case. There is certainly a surface similarity between the two cases, in that one might think that just as it is a matter of luck that one happens to overhear the conversation in OVERHEAR, so it is a matter of luck that one happens to drink the uncorrupted water in WATER. Crucially, however, there is a

key difference. For although it is a matter of luck that the agent in OVERHEAR overhears what she does, nonetheless she is in a great position to epistemically exploit this opportunity (since her hearing is working fine, and so on). But the same is not true of the agent in WATER. After all, what looks and tastes like water in her environment need not be water. This is why the agent's belief in WATER is unsafe, but the agent's belief in OVERHEAR is safe. I suggest that once we understand the difference between these two types of case then one can see why some epistemologists might be inclined to ascribe knowledge in a case like WATER, even though knowledge is not in fact possessed by the agent concerned.<sup>19</sup>

A more interesting case is offered by Christoph Kelp (2009), which we can express as follows:

DEMON: A demon wants our hero – let's call him "Chris" – to form the belief that the time is 8.22 a.m. when he comes down the stairs first thing in the morning (the demon doesn't care whether the belief is true). Since he is a demon, with lots of special powers, he is able to ensure that Chris believes this proposition (e.g., by manipulating the clock). Now suppose that Chris happens to come downstairs that morning at exactly 8.22 a.m., and so forms the belief that the time is 8.22 a.m. by looking at the accurate clock at the bottom of the stairs. Accordingly, the demon achieves what he wants without having to do anything.<sup>20</sup>

Kelp's claim is that insofar as the demon doesn't intervene then, given how Chris formed his belief, he gains knowledge. But since the demon will ensure that Chris continues to believe that the time is 8.22 a.m. in all nearby possible worlds, even when this is false, Kelp also claims that this belief is nonetheless unsafe.

While I think this example is ingenious, I don't think it works. In particular, I don't at all share the intuition that the agent in DEMON has knowledge. After all, given how Chris formed his belief it is pure luck that this belief happens to be true – had he come downstairs a minute earlier or a minute later then he would have formed a false belief. Indeed, Chris is effectively finding out the time by looking at what is (for him anyway) a stopped clock, since whatever time he comes downstairs the clock will say "8.22 a.m." But one cannot gain knowledge about the time by consulting a stopped clock, even when one happens to form a true belief!

That said, I do think that there is something epistemically laudable about the agent's true belief, in that (given that the demon didn't in fact intervene) it is a cognitive success that is attributable to his cognitive ability and thus his cognitive agency. In this sense, it constitutes a cognitive achievement on the part of the subject, even though it isn't knowledge. Often knowledge and cognitive achievement go hand in hand (which I think may explain Kelp's inclination to ascribe knowledge to the agent in DEMON), but what I think cases like this illustrate quite neatly is that they can come apart. In particular, they will come apart in cases where the luckiness of the cognitive success is entirely due to some feature of the modal environment which is absent in the actual world.<sup>21</sup>

The third and final putative counterexample is due to Ian Church (2010):

VIRUS: Smith is ill and exhibits a unique set of symptoms, S. Given these symptoms, Dr Jones forms the belief that "Smith has Virus X," which she deduces from the true proposition that "Virus X is the only known virus to exhibit S." What is more, Dr Jones does a blood test which verifies that Smith's body contains antibodies for Virus X, further justifying Jones' belief.



Based on the evidence, it is extremely feasible that Smith has Virus X. As it happens, however, Smith's symptoms are in fact due to an unknown virus, Virus Y, which exhibits identical symptoms to Virus X; Smith only exhibits antibodies for Virus X due to an idiosyncratic feature of Smith's particular biochemistry which causes his immune system to maintain high levels of antibodies long past a given infection. Nevertheless, Dr Jones' belief turns out to be true divorced from Smith's symptoms or his blood work, because Smith was infected with Virus X just before meeting with Dr Jones – the infection being so recent that blood work cannot detect it and it is causing no symptoms. (Church, 2010, p. 9)<sup>22</sup>

Notice that this case is different from the other two, in that rather than being a putative case of knowledge where the belief in question is unsafe, it is instead an example of a safe belief which doesn't amount to knowledge. The reason why Church thinks that VIRUS is nonetheless a counterexample to the necessity of safety for knowledge is that he holds that the reason knowledge is lacking in this case is due to the epistemic luck involved. But since it is the job of safety to exclude knowledge-undermining epistemic luck, it follows that this case presents a problem for proponents of safety.

Unfortunately, the case doesn't work. Even if we further add – as Church (2010, p. 10) himself suggests – that the circumstances of the case are such that Smith is virtually guaranteed to catch virus X just before seeing Dr Jones, this is only at best a counterexample to a crude form of safety which focuses only on the subject's continued belief that *p* across the relevant possible worlds. Remember the point made above about how a belief in a necessary proposition can be unsafe, even though there is obviously no nearby possible world where the necessary proposition in question is believed falsely. The same applies here. Even if there can be no nearby possible world in which Dr Jones believes that Smith has virus X and believes falsely – because this proposition is true across all nearby possible worlds – it doesn't follow that the belief is safe, since we also need to consider the other beliefs that Dr Jones forms in nearby possible worlds on the same basis as in the actual world. Once we remember this, it becomes clear that Dr Jones will form false beliefs in nearby possible worlds on the same basis as in the actual world. Consider, for example, the close possible world where all that is different is that Smith doesn't happen to maintain a high level of antibodies in his blood and doesn't clearly exhibit the symptoms for virus X. In such a world Dr Jones would likely form the false belief that Smith *didn't* have virus X, or the false belief that Smith had another virus which he didn't in fact have.

Of course, we can always set a case up in which the agent is guaranteed not just to form a true belief in the actual world, but also to form a true belief on the same basis across all nearby possible worlds. Perhaps the case offered by Church could be reconstructed to do this. But even so, this need be nothing for the defender of the necessity of safety to knowledge to worry about. For as noted above, such an epistemologist does not claim that safe true belief is sufficient for knowledge, and so it is perfectly compatible with the story they tell that there may be cases of safe true belief which aren't thereby cases of knowledge. Moreover, given that the agent is now guaranteed to have a true belief on the relevant basis across all possible worlds, it seems that what is lacking in such a case is nothing to do with the anti-luck intuition. After all, if one has a true belief in these circumstances then it is surely not a matter of *luck* that one's belief is true.

Indeed, the point made above about how we shouldn't expect the safety condition to fully capture our intuitions about the role of cognitive ability in knowledge acquisition

are very salient to such a case. For it seems that with the case so construed, what has gone awry, epistemically, is not that the agent is forming beliefs such that they could so very easily have been false, but rather that her cognitive success is not appropriately related to her cognitive ability. Put another way, it seems that what is epistemically problematic about such beliefs is not that they fail to satisfy the anti-luck intuition about knowledge, but rather that they fail to satisfy the ability intuition.

## Acknowledgments

This paper was written while I was in receipt of a Phillip Leverhulme Prize. My recent thinking about these issues has been informed by conversations with (amongst others) Kelly Becker, Tim Black, Adam Carter, Ian Church, Julien Dutant, Georgi Gardiner, John Greco, Allan Hazlett, Avram Hiller, Jesper Kallestrup, Chris Kelp, Ram Neta, and Ernie Sosa. Special thanks go to Stephen Hetherington, Brent Madison, and John Turri who offered comments on an earlier version.

## Notes

- 1 See Unger (1968) for what is almost certainly the first clear statement of the idea that there is a specific sense in which we want our knowledge to be non-lucky (or “non-accidental,” as he terms it). See Pritchard (2005) for a sustained development of this idea.
- 2 I bracket here the issue of whether a guess can ever be properly considered a belief.
- 3 This is a variation on a famous example given by Chisholm (1977, p. 105). Although most epistemologists grant that agents lack knowledge in Gettier-style cases, there are some detractors. See, especially, Hetherington (1998; 2002, chapter 1). For a very useful recent discussion of Hetherington’s view, see Madison (2010).
- 4 For the key texts in this regard, see Dretske (1970, 1971) and Nozick (1981). For some recent texts which sympathetically explore the sensitivity principle, see Roush (2005), Becker (2007), Black and Murphy (2007), and Black (2008).
- 5 Of course, a change in truth value of any proposition is bound to have further implications, and hence the difference between the actual world and the target possible world is unlikely to be only that the proposition in question is false. But we can set this complication to one side for our purposes.
- 6 For the classic discussion of this constraint on sensitivity, see Nozick (1981, pp. 179ff.). For further discussion, see Pritchard (2008, §3).
- 7 If you like, add the stipulation that this second agent doesn’t even know how long the odds for this lottery are.
- 8 For a more thorough discussion of some of the problems faced by the sensitivity principle, see Pritchard (2008).
- 9 This problem for the sensitivity principle is due to Sosa (1999).
- 10 Versions of safety-type principles have been offered by a number of authors, including Luper (1984; cf. Luper, 2006), Sainsbury (1997), Sosa (1999), Williamson (2000), and Pritchard (2002, 2005, 2007).
- 11 And note that it will do no good to say that the actual world is in fact such that in a great deal of nearby possible worlds the colleague would have removed her socks (perhaps because she often likes to do this while on the train, unbeknownst to our agent), since if that’s the case then there is no intuition that our agent has inductive knowledge which requires rescue.

- 12 Although it would take me too far afield to go into this issue here, there is in fact a strong motivation not just for safety, but also for this specific rendering of safety in an epistemological program that I call *anti-luck epistemology*. Very roughly, this program involves identifying what luck is, and identifying the specific sense in which knowledge excludes luck, in order to gain an insight into the nature of knowledge. For more details, see Pritchard (2005, 2007, 2008, 2009).
- 13 For skepticism about the prospects of a safety-based account of knowledge dealing with the lottery problem, see Greco (2007) and McEvoy (2009). For responses, see (respectively) Pritchard (2007, 2009).
- 14 Or, for that matter, any proposition which is true across all nearby possible worlds, even if not necessarily true.
- 15 In any case, no formulation of safety could exclude such cases. The reason for this is that a modal principle like safety cannot capture the “direction of fit” between belief and fact that is key to the ability intuition on account of how it simply specifies a match between belief and fact across a range of worlds. For more on this point, see Pritchard, Millar, and Haddock (2010, chapter 3) and Pritchard (forthcoming).
- 16 I have argued elsewhere for a view I call *anti-luck virtue epistemology*, according to which knowledge is essentially safe true belief plus a further epistemic condition (an “ability condition”) which handles the ability intuition. See, for example, Pritchard, Millar, and Haddock (2010, chapter 3) and Pritchard (forthcoming).
- 17 Note that these cases are often put forward against particular formulations of the safety principle, but our interest will be whether they work against the specific formulation we offer here.
- 18 See also the very similar counterexample to the necessity of safety for knowledge offered in Hiller and Neta (2007, pp. 310–311).
- 19 The same diagnosis will apply to the structurally similar, though more complex, “Halloween party” case offered by Comesaña (2005, p. 397), which is also meant to be a counterexample to the necessity of safety for knowledge. That said, as Kelp (2009) points out, it isn’t at all obvious that Comesaña’s example even involves an unsafe belief in the first place.
- 20 This is essentially a type of “Frankfurt-style” example in that what is significant is that the demon *would* have intervened rather than that he did intervene. See Frankfurt (1969).
- 21 I develop this point about how knowledge and cognitive achievement come apart in a number of places. See, for example, Pritchard (2009, forthcoming) and Pritchard, Millar, and Haddock (2010, chapter 2).
- 22 This example is a variation on a case originally proposed by Zagzebski (1994, p. 71), albeit to illustrate a different point.

## References

- Becker, K. (2007) *Epistemology Modalized*. London: Routledge.
- Black, T. (2008) Defending a sensitive neo-Moorean invariantism. In *New Waves in Epistemology*, ed. V.F. Hendricks and D.H. Pritchard, 8–27. Basingstoke: Palgrave Macmillan.
- Black, T. and Murphy, P. (2007) In defense of sensitivity. *Synthese* 154: 53–71.
- Chisholm, R. (1977) *Theory of Knowledge*, 2nd edn. Englewood Cliffs, NJ: Prentice-Hall.
- Church, I. (2010) Getting “lucky” with Gettier. *European Journal of Philosophy*. DOI: 10.1111/j.1468-0378.2010.00433.x.
- Comesaña, J. (2005) Unsafe knowledge. *Synthese* 146: 393–402.
- Dretske, F. (1970) Epistemic operators. *Journal of Philosophy* 67: 1007–1023.
- Dretske, F. (1971) Conclusive reasons. *Australasian Journal of Philosophy* 49: 1–22.
- Frankfurt, H. (1969) Alternate possibilities and moral responsibility. *Journal of Philosophy* 46: 829–839.

- Greco, J. (2007) Worries about Pritchard's safety. *Synthese* 158: 299–302.
- Hetherington, S. (1998) Actually knowing. *Philosophical Quarterly* 48: 453–469.
- Hetherington, S. (2002) *Good Knowledge, Bad Knowledge: On Two Dogmas of Epistemology*. Oxford: Oxford University Press.
- Hiller, A. and Neta, R. (2007) Safety and epistemic luck. *Synthese* 158: 303–314.
- Kelp, C. (2009) Knowledge and safety. *Journal of Philosophical Research* 34: 21–31.
- Luper, S. (1984) The epistemic predicament. *Australasian Journal of Philosophy* 62: 26–50.
- Luper, S. (2003) Indiscernability skepticism. *The Sceptics: Contemporary Essays*, ed. S. Luper, 183–202. Aldershot: Ashgate.
- Luper, S. (2006) Dretske on knowledge closure. *Australasian Journal of Philosophy* 84 (3): 379–394.
- Madison, B.J.C. (2010) Combating anti-anti-luck epistemology. *Australasian Journal of Philosophy*. DOI: 10.1080/00048400903524375.
- McEvoy, M. (2009) The lottery puzzle and Pritchard's safety analysis of knowledge. *Journal of Philosophical Research* 34: 10–20.
- Neta, R. and Rohrbaugh, G. (2004) Luminosity and the safety of knowledge. *Pacific Philosophical Quarterly* 85: 396–406.
- Nozick, R. (1981) *Philosophical Explanations*. Oxford: Oxford University Press.
- Pritchard, D.H. (2002) Resurrecting the Moorean response to the sceptic. *International Journal of Philosophical Studies* 10: 283–307.
- Pritchard, D.H. (2005) *Epistemic Luck*. Oxford: Oxford University Press.
- Pritchard, D.H. (2007) Anti-luck epistemology. *Synthese* 158: 277–297.
- Pritchard, D.H. (2008) Sensitivity, safety, and anti-luck epistemology. In *The Oxford Handbook of Scepticism*, ed. J. Greco, 437–455. Oxford: Oxford University Press.
- Pritchard, D.H. (2009) Safety-based epistemology: whither now? *Journal of Philosophical Research* 34: 33–45.
- Pritchard, D.H. (forthcoming) Anti-luck virtue epistemology. *Journal of Philosophy*.
- Pritchard, D.H., Millar, A., and Haddock, A. (2010) *The Nature and Value of Knowledge: Three Investigations*. Oxford: Oxford University Press.
- Roush, S. (2005) *Tracking Truth: Knowledge, Evidence and Science*. Oxford: Oxford University Press.
- Sainsbury, R.M. (1997) Easy possibilities. *Philosophy and Phenomenological Research* 57: 907–919.
- Sosa, E. (1999) How to defeat opposition to Moore. *Philosophical Perspectives* 13: 141–154.
- Unger, P. (1968) An analysis of factual knowledge. *Journal of Philosophy* 65: 157–170.
- Williamson, T. (2000) *Knowledge and its Limits*. Oxford: Oxford University Press.
- Zagzebski, L. (1994) The inescapability of Gettier problems. *Philosophical Quarterly* 44: 65–73.

# Knowledge Can Be Lucky

*Stephen Hetherington*

Maybe it is a bold claim: “There can be lucky knowledge.” Or is it not really so bold? That depends: “Lucky? In what way?” Lucky we are, then, that recent epistemology has sought to help us with this. The epistemological debate over whether lucky knowledge is possible has become mainly a debate as to whether knowledge needs to be *safe*. What does this mean? Let us find out. Let us also see why knowledge need *not* be safe – how, in at least that sense, there can be lucky knowledge.<sup>1</sup>

## 1 Safety's Modal Nature

When Ernest Sosa (1999) asked epistemologists the question, "How must knowledge be modally related to what is known?" his own answer was, "Safely." What did he mean by this? Like others before and since, he wished to understand knowledge as unable to involve a belief's being true by happenstance – by luck, in this sense. Yes, a belief must be true if it is to be knowledge; but it also *cannot too easily be other than* true if it is to be knowledge.

Can this latter condition be understood more precisely? Here is how, at that time, Sosa characterized knowledge's need to be safe (1999, p. 378):

Call a belief by S that *p* "safe" iff: S would not believe that *p* without its being so that *p*. ...  
*Safety* In order to (be said correctly to) constitute knowledge a belief must be safe.

But a more recent characterization by Sosa favors the slightly more complex idea of *basis-relative* safety (2007, p. 26):

What knowledge requires is ... not outright safety but at most basis-relative safety. What is required of one's belief, if it is to constitute knowledge, is at most its having some basis that it would not easily have had unless true, some basis that it would (likely) have had only if true.

Duncan Pritchard has offered a similar choice of formulations. For example, this is what he calls *the restricted safety principle* (2009a, p. 45):

If *S* knows that *p*, then *S*'s true belief that *p* could not have easily been false.

Strictly, he understands this as saying that (2009a, p. 46):

the safety principle, properly understood at any rate, is concerned not with the safety of a belief that *p* (i.e., the proposition actually believed), but rather with the safety of a relevant doxastic output of a belief-forming process (which, while being a belief that *p* in the actual world, could be a belief in a different proposition in a nearby possible world). In this way, possible worlds where this process results in a different doxastic output from that which results in the actual world – such as a belief that *q* rather than that *p* – can be relevant to the safety of the belief. On the face of it at least, this modification of the view enables it to deal with cases involving necessary propositions.<sup>2</sup>

What is the challenge posed for safe-knowledge advocacy by knowledge of necessary truths? No necessarily true *p* is false in any possible worlds, including worlds which would be claimed to help to constitute the safety or otherwise of the belief that *p*.<sup>3</sup> Nevertheless, presumably safe-knowledge advocates will allow that some beliefs with necessarily true contents are not knowledge; and the accompanying explanation of those beliefs not being knowledge, Pritchard is saying, is their not having been formed through a safe *method*.

Details differ, then, across a range of possible formulations.<sup>4</sup> Generically speaking, though, this much is clear:

To require safety within knowing is to expect the belief in question, along with its truth, to be replicated in aptly similar possible worlds – such worlds perhaps being ones where the belief's basis or its method of formation remain present.

In some such way, knowing is to be a modally robust or reliable state or accomplishment. Although it could well be much more besides, it is *at least* this much.

## 2 Safety-Advocacy's Modal Failing

Even that much, however, is too much. Consider some true belief B, formed via process A or on basis A. (I will use "A" to refer, with systematic ambiguity, to either of these.) Here is how a safe-knowledge advocate may well reason about B:

There is a nearby world W where B is present (still arising from A) – yet where B is no longer true. W's existence constitutes its not being modally difficult here, even given A, for B to be false; which is to say that B has not been formed safely in this world. Accordingly, B is not knowledge even in this world.

The safe-knowledge advocate will reach for such reasoning regarding at least some beliefs. But such thinking has a subtle failing. (i) We may agree that it describes a way in which *the belief B* as such (although true here in this world and although arising through A) need not have been true – and thereby need not have been knowledge. (B in the nearby possible world W is false, hence not knowledge in W.) However, all that follows from this is that B conceived of simply *as* a belief need not have been knowledge. B *qua* belief is present in this world; it is also present in W, yet false there (even while arising through A). So, B *qua* belief, assessed within this world, could have failed to be true (even when still arising via A). (ii) Yet how should we proceed to parse that conclusion? Not in a way that need worry us; for it tells us only that B's being *a belief* is not sufficient to make it knowledge. (It tells us this by having provided the result that B *qua* belief has been formed unsafely.) However, *this* should hardly be news to epistemologists. No one among contemporary epistemologists, at any rate, believes that being a belief (with some or other basis, or with some or other causal background) is sufficient for being knowledge.

So, what is required on behalf of safe-knowledge advocacy is an analogous argument – concerning a more epistemologically realistic sufficient candidate for knowledge than *belief* could ever be. (The aim, we should bear in mind, is to demonstrate the potential explanatory value of the concept of safety, for epistemologists trying to understand how knowledge can be absent.) Let us ask, therefore, whether a safe-knowledge advocate can offer an argument at least for *the true belief B*, or for *the true belief B arising through A*, not being knowledge – because it is unsafe. (After all, from section 1, we may have expected the concept of safety *to* be describing a standard of which a mere true belief could well fall short.)

But here is where the safe-knowledge advocate's story begins most conspicuously to fail. For a start, the hypothesized reasoning presented above will not show directly that *the true belief*, say, is unsafe. To say that it does show this would be modally fallacious. To derive such a conclusion from the safe-knowledge advocate's hypothesized

reasoning would be to ignore the following core thesis regarding metaphysical *essentialism*:

For something,  $x$ , to possess a property  $P$  essentially is for  $x$  to possess it in all possible states of affairs where  $x$  exists. (In possible worlds terms:  $x$  is  $P$  essentially in a world  $W$ , if and only if  $x$  is  $P$  in all possible worlds accessible from  $W$  – such as all possible worlds close to  $W$  – where  $x$  exists.)

We are now seeking to ask (on behalf of the safe-knowledge advocate) whether a specific *true belief* as such is safely present. We will not find the answer to that question by consulting possible worlds where the *belief* is all that we have required to be present – the belief as such being what we are reidentifying across these worlds. The belief as such need *only* be the belief – considered apart from whether it is true, so that we can then ask or discover, for each of those worlds in turn, whether it is true there. We would be reidentifying the belief across worlds without its-being-true being a constitutive criterion involved in this process of identifying. Consequently, insofar as  $W$  includes a false  $B$ ,  $W$ 's accessibility reveals at most the unsafety of belief  $B$  as such – that is,  $B$  considered just *as* a belief with that same content, reidentifiable across worlds. Manifestly,  $W$  does not reveal the unsafety of a *true* belief  $B$  – that is,  $B$  *qua* true belief. For  $B$  – by being false within  $W$  – is not *identifiable* within  $W$  as a true belief in the first place. (It is identifiable only as a belief there – before we pose the question of whether it is true there.)

Why does this matter? Remember that we are now asking whether the safe-knowledge advocate can show some true belief as such not to be knowledge (due to its not being formed safely). We see that she cannot accomplish this. “Should we *move away* from asking that question about true beliefs as such?” But it is significant that  $B$  was only ever a candidate for knowledge in this world by *being* true. Epistemologically, our question should never have been, for instance, “Is  $B$  safe?” It *should* have been more like this: “Is  $B$  – true and arising through  $A$  – safe?” And so far we have found that that even if a particular *belief* that  $p$  as such is unsafe, this does not entail that the *true belief* that  $p$  as such is unsafe. (That sounds odd, I concede. Sections 3 and 4 will explain and defuse that sense of oddity.)

That non-entailment, present just where the safe-knowledge advocate would expect there to be no such thing, bespeaks a potential modal fallacy within her reasoning. Specifically, when comparing this world with other possible ones, she runs the risk of not *maintaining a single subject of predication*. Her reidentified subject should be the *true belief*, along with any accompanying epistemic features, such as its basis  $A$  or its causal background  $A$ . Do *these* jointly constitute knowledge (she should be asking)? As her reasoning typically develops in fact, however, the subject is only said to be the *belief*, along with  $A$ , say – not the *true belief* – as she proceeds to describe worlds where the belief is not true. Within this world the belief in question is true. Within  $W$  it is false. So, within  $W$  this world's *true belief* as such is not present – which means (given the nature of essentialism) that  $W$  is not a world where this world's *true belief as such* is being modally evaluated, such as with an epistemological eye upon whether it is safely formed. Again, we must acknowledge, the belief *qua belief*'s being formed unsafely does not entail the belief *qua true belief*'s being formed unsafely. The safe-knowledge advocate's reasoning establishes only the former – and not thereby the

latter – lack of safety. Yet only the latter – not the former – lack of safety would shed epistemological light upon the nature of knowledge.

### 3 Safety-Advocacy's Dilemma

Section 2 has described a modal predication fallacy threatening safe-knowledge advocacy. To some, though, this fallacy might seem to be easily avoidable, along these lines:

*Belief versus true belief.* There is no illicit shift of predication. Throughout, we are univocally evaluating the belief (or perhaps the belief and its basis, or even the belief-forming process). We are never evaluating the *true* belief as such (or the true belief and its basis, or even the true belief plus the process which has generated it). If the belief, say, has not been formed safely, then *it* – the belief – is not knowledge. And so it is not knowledge, even if it is also true.

But such a move exposes the safe-knowledge advocate to a dilemma, one which begins with this choice:

*Either* she retains, as the subject throughout her reasoning, merely the belief (or the belief-forming process, or the belief plus its basis); *or* she allows the subject of her reasoning to be the *true* belief (or the true belief plus the process via which it is formed, or the true belief plus its basis).

Either of those alternatives would avoid section 2's modal predication fallacy – by the simple maneuver of not shifting between them within any reasoning offered on behalf of safe-knowledge advocacy. However, each alternative, considered individually, has a consequence which is fatal to requiring safety within knowledge:

If merely the *belief* (etc.) is the subject of the safe-knowledge advocate's reasoning, then her reasoning cannot be showing that the *true belief* (etc.) is not knowledge because unsafe. Yet the latter is what safe-knowledge advocacy needs to be able to show.

If the *true belief* (etc.) is the subject of the safe-knowledge advocate's reasoning (as it should be), then her reasoning cannot be showing that there is *unsafety* – involving a *lack* of truth for that true belief (etc.) within some possible world – and thereby a failure by that true belief (etc.) to be knowledge here.

Then we may easily combine those two horns of the dilemma:

Once the safe-knowledge advocate retains, throughout her reasoning, the correct *subject* of investigation – namely, the true belief (etc.) – she is unable ever to explain in terms of a lack of *safety* that subject's not being knowledge.

And if safety can never explain how some particular true belief (etc.) *fails* to be knowledge, it cannot ever really explain, even partly, how some true belief (etc.) *succeeds* in being knowledge.



## 4 Diagnosis of the Dilemma's Applicability

Section 3's dilemma presents any safe-knowledge advocate with two untoward consequences. The second flows readily from the conceptual details of what it is for a belief (etc.) to be safe. The first consequence is more likely to be disputed: a safe-knowledge advocate may well continue to say that the belief (etc.) – such as the belief-forming process – is the proper subject of her scrutiny. In contrast, I have urged that the *true* belief (etc.) is the proper subject.<sup>5</sup> So, *why* might the safe-knowledge advocate nonetheless fix her gaze upon beliefs (etc.), not upon true beliefs (etc.)? I have two speculative suggestions regarding this.

(1) *Priority*. The safe-knowledge advocate might say that safety is a condition upon the belief (etc.) that must be satisfied prior to, or independently of, whether the belief is true – in effect, before the question arises of whether the belief is true. For instance, the belief-forming process would need to be safe – prior to, or independently of, whether the resulting belief is true. The motivation for such a suggestion might be that of trying to replicate an inquirer's perspective, not an analyzer's. The epistemic agent could be conceived of as working with a way of potentially forming a belief, wondering whether it would be safe, before committing to a particular belief on the basis of it.

However, such a mode of interpretation would be beside the immediate epistemological point. An analysis reflecting that approach would show the knowledge to be safe *only* in the sense of showing the safety of the belief (etc.), such as the belief-forming process. At best, this would be *indirect* safe-knowledge advocacy. But safe-knowledge advocacy itself – directly so – is what we are discussing, not safe-belief advocacy or safe-belief-plus-its-basis advocacy or even safe-belief-forming-process advocacy. A fortiori, our concern is not with any of these latter three kinds of advocacy insofar as the belief or the belief-plus-its-basis or the belief-forming process is being considered *apart* from its actually being or involving or producing a true belief.

A mark of the irrelevance of safe-belief advocacy (or safe-belief-forming-process advocacy, etc.) is precisely the fallaciousness – revealed in section 2 – of reasoning like this:

The *belief* in question is not knowledge because it is not formed safely. Therefore, the *true* belief in question is not knowledge because it is not formed safely.

Insofar as a belief is considered apart from whether it is actually true (and apart from its actual basis or the actual process of forming it), it might well not be knowledge because it is unsafe. But we have seen why this does not entail that the belief – once it is considered as true (and as having been formed as in fact it has been, such as on its actual basis) – is also not knowledge because it is unsafe. Perhaps that putative entailment sounds to the safe-knowledge advocate as if it *should* obtain. Nevertheless, it does not obtain. Even if a belief *qua* belief can be understood as – because unsafely formed – not knowledge, this does not entail that the same belief *qua true* belief is to be similarly understood. (And maybe there is something apt in that non-entailment. A true belief as such is closer than a belief as such to being knowledge.)

(2) *Conjunction*. Many epistemologists will say that any instance of knowledge is a conjunction of truth and belief (along with whatever else is needed). Those same epistemologists might then insist that if one of those conjuncts is not safely present, neither

is the conjunction; in which case (these epistemologists would conclude), if the belief is not safely formed then knowledge is absent due to a lack of safety.

But that reasoning fails, we have found. Once a belief is characterized as being true, whether it is knowledge is not aptly determined by testing it (etc.) for safety; for even if some true beliefs are not knowledge, no belief can be characterized as a true belief while also being shown not to be safe (and hence while *thereby* being shown not to be knowledge). That is, no belief characterized as true can then – by being shown to be unsafely formed – be shown not to be knowledge. And this is so, even if, by considering the belief *apart* from its being true, we could have explained its being unsafely formed and thereby not knowledge.<sup>6</sup> Truth contributes to something's being knowledge. We already knew so; now we know so more fully.

Why might an epistemologist reach for that conjunction view of knowledge? Perhaps she assumes what we could, in the spirit of John McDowell (1998), describe as a *highest common factor* conception of what is being evaluated for its safety. She will claim that, on such a conception, it is belief (etc.) – not true belief (etc.) – which persists as the subject throughout the evaluation of that subject's safety: true *or* not, the belief (etc.) would thus be the continuing subject of this epistemological scrutiny.

However, that approach fails; and it does so even if we accept the idea of allowing the question of safety's nature and epistemic significance to be one that does indeed arise about whatever is in fact the highest common factor. The reason for the failure, even then, is that whatever is knowledge and whatever is not-knowledge-only-because-it-is-not-safe *can* have true belief – rather than merely belief as such – as a highest common factor. In fact, they *do* share this highest common factor, insofar as (from section 1) the question of safety arises from the outset as a putative way of distinguishing knowledge from true belief (etc.) in particular.

## 5 Knowledge Fallibilism versus Knowledge Infallibilism

Safe-knowledge advocacy is not generally intended to be part purely of an *infallibilist* conception of knowledge. Pritchard, for example, invites us to consider *close* possible worlds, not *all* possible worlds, where the belief in question is formed as it is here. Nonetheless, an epistemologist's motivation for engaging in safe-knowledge advocacy could be quite similar to how infallibilism is sometimes – and ineffectively – motivated. I will not pursue this issue definitively, because epistemologists are yet to agree on how to formulate knowledge fallibilism and knowledge infallibilism.<sup>7</sup> Still, we can make some progress.

Here is a traditional but fatally flawed claim with which people have often sought to motivate respect for knowledge infallibilism:

- (1) If you know, you cannot be mistaken.

And here is a seemingly analogous way – mentioned at the start of section 1 – of trying to motivate respect for the idea for knowledge safety:

- (2) If you know, you cannot easily be mistaken.

The manifest difference between (1) and (2) is the latter's use of "easily" – clearly meant to lead us toward knowledge fallibilism rather than knowledge infallibilism.<sup>8</sup> That is not enough of a difference, however, to save (2)-as-intended-motivation-for-knowledge-safety from the same failings, *mutatis mutandis*, which wholly undermine (1)-as-intended-support-for-infallibilism. We see this as follows.

As epistemologists are well aware, (1) admits of these two readings:<sup>9</sup>

- (1a) It must be that if you know that *p* then it is true that *p*.
- (1b) If you know that *p*, then it must be true that *p*.

(1a) is true yet trivially so, in a way which does not bespeak infallibilism's truth in particular: any case of knowledge – whether *or* not it is infallible – is of a truth; and this must be so. What of (1b)? It is false, we presume: you could know something that, although true, never *had* to be true. (For example, you know that it *is* raining – even though the world might not have included this rain.) Yet (1), when considered precisely, is either (1a) or (1b). Accordingly, (1)'s *prima facie* luster as a truth supporting knowledge infallibilism in particular fades swiftly.

Now let us consider (2). To many, it could well sound like a natural way of motivating a requirement of safety within any epistemological conception of knowledge: "Obviously (2) is true. And it is simply a way of saying that knowledge must be safe (and in this sense not lucky)." But (2) – as I will now explain – is as unhelpful for motivating knowledge safety as (1) is for motivating knowledge infallibilism. (1) fails in that respect; so does (2).

Unlike (1), (2) does permit knowing to include a possibility of being mistaken. Nonetheless, (2) is meant to convey that the possibility of such a mistake is allowed only to be "remote." Correlatively, here are two candidate readings of (2), ones that are analogous to (1a) and (1b):<sup>10</sup>

- (2a) It must (at least in all possibilities that are not remote; maybe even in the remote ones) be that if you know that *p* then it is true that *p*.
- (2b) If you know that *p*, then it must (at least in all possibilities that are not remote; maybe even in the remote ones) be true that *p*.

Like (1a), (2a) is true yet trivially so. (2a) provides no substantive support for a knowledge fallibilism in particular – just as (1a) was no reason to favor knowledge infallibilism. After all, (2a) tells us merely that in at least any world fairly similar to this one (and perhaps even in ones very unlike this one), any case of knowledge is knowledge of a truth, regardless of whether or not the knowledge is infallible. And that makes (2a) true – but without its describing anything beyond knowledge's being of a truth, in at least almost all possible worlds, perhaps in all. This is weaker than a requirement of safety within knowledge.

What now of (2b)? Like (1b), (2b) is substantial yet false. This is so, even if the parenthetical condition "(at least in all possibilities that are not remote; maybe even in the remote ones)" means – as it probably does – something like "true in at least all possible worlds similar to this one."<sup>11</sup> Unless we beg the question at issue (of

whether knowledge has to include safety) by decreeing that any possible world similar to this one includes *p*'s being true, (2b) is readily seen to be false: most of the many ways for a world to be similar to this one will not include the specific *p* being true. Even given this world's including your knowing that *p*, it is not clearly true that every world similar to this one must be similar in ways that are precisely as would be expected given your knowing that *p*. That is because a possible world's overall similarity to this world need not retain *that* element of this world. That particular possible world could be similar to this world in *other* ways, since your knowing that *p* is merely one small element of this world. And there could be many similar worlds like that.

Until its analogy with (1) was highlighted, (2) might have seemed like a good way of motivating safe-knowledge advocacy. Now it should not seem so promising.

## 6 Anti-Luck

Are there other ways instead for a safe-knowledge advocate to motivate her position? Recall the widely accepted view with which this chapter began, that there are kinds of *luck* which preclude knowing – and which are themselves precluded by safety? Wait a moment; did I call that a “view”? If Pritchard is right, it is nothing as arguable, let alone as questionable, as a view. Rather, it is a *platitude* (2007, p. 277; 2009b, p. 33) that knowledge excludes such luck – veritic luck, he calls it. *That* is how obviously true it is, according to Pritchard. In rejecting the demand that knowing be safe, therefore, am I entering deeply unsafe epistemological waters by spurning nothing less than a platitude?

Not at all. (i) I have not merely rejected that demand. I have undermined arguments for it. This should weaken the case for the safe-knowledge advocate's central claim's resting upon something so clearly true as to be a platitude. (ii) What remains available to the safe-knowledge advocate? Further arguments? Possibly; we await them. Intuitions? Philosophy can – no, it must – do better than that; it should find arguments. Bear in mind that, as Juan Comesaña observes (2005, p. 399), the concept of safety is *technical*. It is not “intuitive.” Nor therefore are theses, such as anti-luck claims, which are intended to display its significance. (iii) “But surely,” a safe-knowledge advocate may reply, “the anti-luck claim itself – the *motivation*, no less, behind the epistemological debate about knowledge and safety – is intuitive. *It* is a genuine intuition, of which the safe-knowledge demand is a technical formulation.” Yet even if the anti-luck claim is intuitive, the following point persists: If an epistemologist's best attempt to formulate precisely that intuitive claim ends in futility, she should discard that intuition. She might seek a replacement intuition. Or she may decide to be content with something not so intuitive. Our role as philosophers should include our reaching and evaluating various technical formulations as part of testing our supposed intuitions – in a spirit of openness to the possibility of proceeding to replace those intuitions, either with fresh ones or even with views that are not intuitive at all. The anti-luck intuition is not *so* well entrenched or fundamental as to be beyond the reaches of this form of methodological challenge.<sup>12</sup> Nor, a fortiori, is safe-knowledge advocacy.

## 7 Pritchard's Form of Analysis

I will end by applying some of my reasoning to a representative case from the argument in Pritchard's essay above, "Knowledge Cannot Be Lucky." I will not contest his argument for replacing a requirement of sensitivity with one of safety, or his requirement that the latter be a necessary rather than sufficient condition of knowledge's presence.<sup>13</sup> But it will be instructive to reflect on how he approaches the analysis of specific cases of knowing or not knowing, since that is the core of his methodology.

Pritchard describes someone's flipping a coin, thereby coming to believe that  $2 + 2 = 4$ . Is that a safe method of belief formation? Not at all (infers Pritchard), in spite of the belief's content being necessarily true: there are many nearby possible worlds where that same method generates false beliefs (even though this specific belief is never false when present). And that lack of safety explains why the belief is not knowledge. Too easily could the method by which this belief has been formed *not* have led to a true belief such as this one.

But that assessment does not show why we should accept that this belief that  $2 + 2 = 4$  is not knowledge *because* of its being unsafely formed. Is the belief's having been unsafely formed *why* – formed as it is via coin flipping – it is not knowledge? Apparently Pritchard wants us to reason thus:

- KS (1) The belief in question has been formed unsafely. [From Pritchard's account of safety]  
(2) The belief in question is not knowledge. [Independent assessment, perhaps "intuitively"]  
(3) (1) – this lack of safety – explains (2) – this lack of knowledge. [Also "intuitively"?]  
(4) So, knowledge in this instance needs to have been formed safely. [From (1)–(3)]

And presumably Pritchard is confident that he would repeat such reasoning in any other apt cases. He accepts (4)'s universal generalization:

- (5) Knowledge in *each* instance needs to have been formed safely.

Yet we have found the following reasoning to be no less well supported than KS:

- not-KS (1\*) The *true belief* in question cannot have been formed unsafely. [From this chapter's argument]  
(2\*) The *true belief* in question is not knowledge. [Independent assessment; for argument's sake, the same as Pritchard's of (2)]  
(3\*) A lack of safety does not explain (2\*) – this lack of knowledge. [From (1\*)–(2\*)]  
(4\*) So, knowledge in this instance need not have been formed safely. [From (3\*)]

And, given (4\*), there is no manifest reason not to infer this universal generalization:

- (5\*) Knowledge in *no* instance needs to have been formed safely.

How should we choose between these two patterns of inference, KS versus not-KS? For a start, from section 4, (1) does not entail not-(1\*). The question of the safety or otherwise in how a particular *true belief* that  $2 + 2 = 4$  has been formed is not decomposable into questions, one of which is that of the safety or otherwise in how the correlative *belief* that  $2 + 2 = 4$  has been formed. In that sense and any other, the latter question is not conceptually prior to the former question. We cannot show that a particular true belief that  $p$  has been formed unsafely, merely by showing that the belief that  $p$  has been formed unsafely. That potential epistemological shortcut is not available.

Are KS and not-KS thereby on a par, equally well supported by our intuitions and our theoretical efforts? No, for two reasons. First, (3\*) in not-KS is better supported than (3) in KS. (3\*) follows immediately from (1\*) and (2\*), whereas (3)'s support is neither self-evident nor an immediate consequence of (1) and (2). All else being equal, this makes not-KS internally stronger than KS. Second, in general (all else still being equal) a true belief that  $p$  is closer to being knowledge that  $p$  than is a belief *simpliciter* that  $p$ . So, not-KS is reflecting upon something – a *true belief* – that is more likely (other things being equal) to be knowledge than is KS's subject matter – a *belief*. In that sense, not-KS is attending to more of the crucial data than KS is doing. This makes not-KS externally stronger than KS.

## Acknowledgments

Thanks to Duncan Pritchard and John Turri for their comments on an earlier version of this chapter.

## Notes

- 1 For earlier attempts to argue that, in possibly a different way, there could be lucky knowledge, see Hetherington (1998, 1999). Those attempts occurred prior to the emergence of the debate about *safety* and knowledge. I argued that there can be beliefs that are both *Gettiered* and knowledge, even though many epistemologists have regarded a belief's being Gettiered as entailing its not being knowledge, due to its being only luckily true. For more on this issue (including the supposed links between safety and being Gettiered), see Pritchard (2005) and Hetherington (2011b, chapter 3).
- 2 Pritchard (2009a, pp. 46–47) views that modification – one to which Timothy Williamson (2000, p. 182) also seems sympathetic – as introducing “a new layer of vagueness regarding which possible worlds are relevant when it comes to determining whether a belief is safe” (2009a, pp. 46–47; see also 2009b, p. 34). But such vagueness was likewise present in his earlier main formulation of the principle of safety (2005, p. 163).
- 3 See Roland and Cogburn (forthcoming) for expansion upon why this is problematic for safe-knowledge advocacy.
- 4 For more on how to formulate safety, see DeRose (2004, pp. 29–33).
- 5 Yamada (forthcoming) apparently concurs. Interestingly, he argues on different grounds to mine that knowledge need not be safe, even while accepting that knowledge must not be accidental. See section 6 below on knowledge and luck.
- 6 “In that case, could one person accurately describe a particular belief as ‘an unsafely formed belief and therefore not knowledge’ while a second person *equally accurately* deems the

same belief to be ‘a true belief and therefore knowledge, other things being equal, because – by being a true belief – it is not unsafely formed? That sounds odd. Yet it seems to be a consequence of this chapter’s analysis.” It is not really so odd when precisely understood. The first person could be correct in saying that the belief, *considered just as a belief*, is unsafely formed. But the second person would be correct to note that, once the belief is considered more *fully* in the way she is exemplifying (by classifying it as a true belief in particular, not merely and more generically as a belief), the belief is being described in terms that imply its being safely formed. Insofar as a decisive verdict was being sought, therefore, the first person was *prematurely* classifying the belief as not knowledge, due to her not yet having taken into account whether or not it is true. Once its being true is taken into account (as the second person is doing), the first person’s verdict of “unsafely formed” is correctly reversed – correctly, since in fact the belief *is* true. (For more on how we should assess individual cases, see section 7.)

- 7 For helpful discussion, though, see Fantl and McGrath (2009, chapter 1).
- 8 See, for example, Sosa (1999, pp. 382, 384 n.11).
- 9 The transition from (1) to these alternatives occurs along the following lines. From (1)’s “If you know, you cannot be mistaken,” we move to “If you know, it is impossible that you are mistaken”; from which we derive, “If you know, it must be that you are correct.” From that, we reach the choice between (1a) and (1b).
- 10 Unlike (1a) and (1b), each of (2a) and (2b) includes a lengthy phrase reflecting (2)’s use of “easily.”
- 11 The parenthetical condition’s first clause allows us to set aside – as irrelevant for (2)’s purposes, even if not for (1)’s – remote possible worlds extremely unlike this one. The point of the parenthetical condition’s second clause is to negate any *presumption* of a contrary result then obtaining in those remote possible worlds.
- 12 See Hetherington (2011a) for more on this point.
- 13 For a fuller sense of Pritchard’s views on this, see his (2009a).

## References

- Comesaña, J. (2005) Unsafe knowledge. *Synthese* 146: 395–404.
- DeRose, K. (2004) Safety, sensitivity, and skeptical hypotheses. In *Ernest Sosa and His Critics*, ed. J. Greco, 22–41. Malden, MA: Blackwell.
- Fantl, J. and McGrath, M. (2009) *Knowledge in an Uncertain World*. New York: Oxford University Press.
- Hetherington, S. (1998) Actually knowing. *The Philosophical Quarterly* 48: 453–469.
- Hetherington, S. (1999) Knowing failably. *The Journal of Philosophy* 96: 565–587.
- Hetherington, S. (2011a) Abnormality and Gettier situations: an explanatory proposal. *Ratio* 24 (2): 176–191.
- Hetherington, S. (2011b) *How To Know: A Practicalist Conception of Knowledge*. Malden, MA: Wiley-Blackwell.
- McDowell, J. (1998 [1982]) Criteria, defeasibility, and knowledge. In *Meaning, Knowledge, and Reality*, 369–394. Cambridge, MA: Harvard University Press.
- Pritchard, D. (2005) *Epistemic Luck*. Oxford: Clarendon Press.
- Pritchard, D. (2007) Anti-luck epistemology. *Synthese* 158: 277–297.
- Pritchard, D. (2009a) *Knowledge*. Basingstoke: Palgrave Macmillan.
- Pritchard, D. (2009b) Safety-based epistemology: whither now? *Journal of Philosophical Research* 34: 33–45.
- Roland, J. and Cogburn, J. (forthcoming) Anti-luck epistemologies and necessary truths. *Philosophia*.

- Sosa, E. (1999) How must knowledge be modally related to what is known? *Philosophical Topics* 26: 373–384.
- Sosa, E. (2007) *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Vol. I*. Oxford: Clarendon Press.
- Williamson, T. (2000) *Knowledge and its Limits*. Oxford: Clarendon Press.
- Yamada, M. (forthcoming) Getting it right by accident. *Philosophy and Phenomenological Research*.



## Chapter Eight

# Is There a Priori Knowledge?

For advocates of a priori knowledge, the chief task is to explain how such knowledge comes about. According to Laurence Bonjour, we acquire a priori knowledge of a proposition  $p$  when we grasp that  $p$  is necessarily true. For opponents, the task is twofold. They must give an account of how empirical knowledge of putative a priori propositions is possible, and they must explain why such propositions cannot be known in the way Bonjour claims, namely through rational insight or the recognition of necessity. To do the former, Michael Devitt appeals to the thesis of holism, according to which beliefs face the tribunal of experience not individually but only as a system. Putative a priori propositions are part of our belief system and thus, like everything else we believe, are subject to revision in the light of experience. To accomplish the second task, Devitt argues that, unlike the empirical justification provided by experiences, rational insight as a source of knowledge is utterly obscure. In response, Bonjour claims that the appeal to holism begs the question because it leaves unanswered the question of which aspects of a holistic system are indicative of truth, and that empirical justification, at least when it comes to indirect empirical knowledge, is no less problematic than a priori justification. In the final round of their exchange, Devitt denies and Bonjour asserts (i) that justification through holistic evidence requires knowing that such evidence is truth indicative, (ii) that holistic naturalism is inflicted with circularity, and (iii) that Bonjour's rational insight account of a priori knowledge avoids circularity.

## In Defense of the a Priori

*Laurence Bonjour*

The official subject of this debate is the existence of a priori knowledge. But the main focus of my discussion will in fact be not a priori *knowledge*, but a priori

*justification* – or rather, more specifically, a priori *reasons* for believing something to be true. In approaching the issue in this way, I am assuming both (i) that justification is one of the requirements for knowledge (the only one to which the issue of a priori status is relevant) and (ii) that justification in the relevant sense consists in having a good reason for thinking that the belief in question is true. But having stated these two background assumptions, I will say nothing further in support of them here.

The view I will defend is that a priori reasons, in a sense yet to be clarified, do exist (and in consequence that a version of epistemological rationalism is true). But the idea of an a priori reason (and also the associated rationalist view) has been understood in a number of different ways, and I will not be defending all of the specific claims that have been associated with this sort of position. My aim is to defend what I take to be a relatively minimal version of the idea of an a priori reason (and of rationalism): more or less the most minimal version that is both philosophically interesting and reasonably faithful to the historical dialectic. All this will take some explanation.

### **The Nature of a Priori Reasons**

As I will understand it here, the concept of an a priori reason has two basic elements, one negative and one positive, the negative one initially more obvious, but both in the end equally essential. Negatively, an a priori reason for thinking that a claim is true is one whose rational force or cogency does not derive from *experience*, either directly (as in sense perception) or indirectly (as by inference of any sort – deductive, inductive, or explanatory – whose premises derive their acceptability from experience). That such a reason is in this way independent of experience does *not* mean that someone who has undergone no experience of any sort could be in possession of it, since the possession of an a priori reason requires understanding the claim for which it is a reason, and experience, even experience of some fairly specific sort, might be required for that. Nor does the idea of an a priori reason, when understood in this way, imply either: (i) that experiences of some sort could not also count for or against the claim in question; or (ii) that such experiences could not override, perhaps even more or less conclusively, the a priori reason in question; or still less (iii) that an a priori reason renders the claim certain or infallible. All of these further claims *might* be true in some cases (though not, I believe, in all or even most), but they in no way follow from or are essential to the basic idea of an a priori reason itself.

What then counts, for these purposes, as *experience*? Obviously the paradigm cases are the various sorts of sense experience, including such things as kinesthetic experiences of bodily orientation in addition to those deriving from the five standard senses. But, in opposition to a number of recent discussions, I would argue that *introspective* awareness of one's thoughts, sensations, and other mental states should also count as a variety of experience, and the reasons for belief that such experience provides as empirical rather than a priori. Introspective experience may not depend on clearly identifiable sense organs, but it is still pretty clearly an awareness of temporally located contingent facts that depends on causal relations between those specific facts and the correlative state of awareness; it is thus far more analogous to sense experience than it is to the sort of experiential process, if it should even be called that, that is involved

in the most paradigmatic cases of allegedly a priori reasons. And basically the same thing is true of even the reason for belief in one's own existence that is supplied by the Cartesian *cogito*, since this is based on introspective awareness of the occurrence of specific thoughts and sensations.

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, about which little will be said here, 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.) Though the term "intuition" has often been used to refer to such insights, I will refer to them simply as "a priori insights," thus, I hope, avoiding any confusion with the other uses of the rather slippery term "intuition."

Here it is important to be clear at the outset that insights of this sort are not supposed to be merely brute convictions of truth, on a par with the hunches and fears that may simply strike someone in a psychologically compelling way. On the contrary, a priori insights at least purport to reveal not just *that* the claim is or must be true but also, at some level, *why* this is and indeed must be 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.

One other point about the nature of a priori insights should also 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 is essentially the one pointed out long ago by Lewis Carroll: 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. And once the need for this non-propositional conception of a priori insight is appreciated in the context of deductive inference, it seems to me in fact plausible to extend it to many other cases as well; in particular, it seems plausible to regard the most fundamental insights pertaining to each of the examples listed in the following section as non-propositional in character.<sup>1</sup>

### **The Argument from Examples for the Existence of a Priori Reasons**

Why then should it be thought that reasons having this a priori character genuinely exist? One reason is that there seem to be many, many examples of propositions for which there are clear and obvious reasons of this sort. Here the most obvious examples come from mathematics and logic, but there are others of many widely varying kinds.

For present purposes, a misleadingly short list, reflecting some of the main types, will have to do:

- (1)  $2 + 3 = 5$ .
- (2) All cubes have 12 edges.
- (3) For any propositions P and Q, if it is true that P or Q and it is false that P, then it is true that Q.
- (4) If object A is larger in a specified dimension (length, area, volume, etc.) than object B and B is in turn larger in that same dimension than object C, then A is larger in that dimension than C.
- (5) No surface can be uniformly red and uniformly blue at the same time.

My basic claim is that anyone who understands and thinks carefully about each of these propositions will be able to see or grasp immediately that it must be true, that it is true in any possible world or situation – and that the same thing is also true of indefinitely many further examples of these sorts and others. The central rationalist thesis I am defending is that this sort of seeing or grasping constitutes, other things being equal, a good, indeed overwhelmingly compelling, reason for thinking that the claim in question is true, albeit not a reason that is capable of being stated as a separate proposition. Moreover, while independent experiential reasons might also be found for some or all of these propositions, insights of this basic sort do not depend on experience in any discernible way.

Examples like these, which could be multiplied more or less without limit, provide, I claim, compelling evidence for the existence of a priori reasons (and, given the assumptions enunciated earlier, for a priori justification and knowledge). One who wishes to reject this conclusion (and who does not adopt the quixotic stance of denying that we have good reasons for thinking that any of these propositions are true) is obligated to offer some alternative account of those reasons, one that makes them dependent on experience after all, initial appearances to the contrary. My view is that there is no such specific and detailed account of examples like these that has any real plausibility.

One other point is worth adding, before turning to other arguments in favor of the existence of a priori reasons. What is perhaps most misleading about the list of examples given here is that, being chosen for their obviousness, they are far from being the most philosophically interesting cases of a priori reasons. I believe in fact that there are many more interesting albeit less obvious examples as well: claims about the unlikelihood of complex coincidences of various kinds;<sup>2</sup> certain moral claims; metaphysical claims about matters such as the structure of time and space; and many, many others.

## **Dialectical Arguments for the Existence of A Priori Reasons**

While the foregoing argument from examples for the existence of a priori reasons strikes me as pretty compelling, it is, from a dialectical standpoint, still capable of being resisted. An opponent might deny that we have good reasons for at least some of the propositions in question, dismissing the intuitive impression to the contrary as an illusion of some sort, and might also appeal to some account of how and why our reasons for the rest of them are really at bottom empirical. I find such views extremely implausible, but there is no doubt that they are dialectically tenable as long as it is

only such apparent examples of a priori reasons that are in question. But there are also other arguments of a more dialectical character for the rationalist view, which I want now to consider. These still do not make the rejection of a priori reasons completely impossible to maintain, but they make clear the intolerably high skeptical price of rejecting the existence of such reasons.

I will consider two closely related arguments of this dialectical sort. The first is concerned with the relation between experience and certain of the beliefs which it intuitively seems to justify. On any account of the justificatory force of experience, there will be some beliefs whose justification derives from a direct relation to experience and others whose relation to experience is less direct. The most straightforward version of this picture would be a broadly foundationalist view in which the more directly justified beliefs are justified by the content of experience alone, without the need for any reasoning or any further premises. Despite much recent criticism, I myself do not see how to avoid a view of this general kind, while retaining the view that experience does indeed in some way justify beliefs. But even if this is mistaken and there is some more complicated story to be told concerning the directly justified beliefs, the problem to be described here will still arise about the justification of beliefs for which experience provides justification but *not* in a direct or immediate way.

Where exactly the line between the beliefs that are directly justified by experience and those that are not actually falls is a difficult issue, which need not be resolved here. All that matters for present purposes is that the class of beliefs that are broadly empirical but clearly *not* justified by a direct relation to experience is extremely large and important, something that is so for any conception of the scope of direct experiential justification that has ever been seriously advocated. On any such view, this indirectly justified class of beliefs will include at least: (i) beliefs about the unobserved past; (ii) beliefs about unobserved situations in the present; (iii) beliefs about the future; (iv) beliefs in laws of nature and similar sorts of generalizations; and (v) beliefs about unobservable entities and processes, such as those described by theoretical science. Taken together, beliefs of these various kinds are obviously fundamental to our picture of the world and our place in it.

But how can experience provide justification for beliefs of these kinds, if not directly? The only possible answer to this question, I submit, is that experience can provide a good reason for thinking that a belief in this category is true only if we have a logically prior good reason for believing some conditional proposition having a conjunction of beliefs for which there are direct experiential reasons as antecedent and the further belief we are focusing on as consequent – for only this can establish the connection between experience and something that it does not justify in the more direct way. Here it will make the issue clearer to suppose that the antecedent of our conditional is in fact a conjunction of *all* the propositions for which there are direct experiential reasons, even though most of these will be irrelevant to any particular consequent.

What sort of reason could we have for thinking that a conditional proposition of the indicated sort is true? If all of the things for which there are direct experiential reasons are already contained in the antecedent and if the consequent genuinely goes beyond the content of the antecedent (as only some highly implausible reductionist view could deny for the sorts of claims in question), then experience can offer no direct reason (and no indirect reason without assuming some other conditional of the same sort) for

thinking that such a conditional proposition is true. It follows at once that the justification for a conditional proposition of this sort, if there is any, can only be wholly or partially (via some other such conditional) *a priori* in character. In this way, the blanket rejection of the very existence of *a priori* reasons leads to a deep and pervasive version of skepticism, one in which we have no reason for thinking that any of the various seemingly empirical claims that are not directly justified by experience are true. And this is a result that seems far too extreme to be acceptable.

Note that I have couched the entire argument in terms of reasons for thinking that the various beliefs are true and not in terms of knowledge. Thus it would be possible for a defender of a view that does not appeal to such reasons in its account of knowledge – such as a version of externalism – to hold that we may have knowledge of such matters, while still denying the existence of *a priori* reasons. But the admission that we have no reasons of any sort for thinking that such beliefs are true, even while insisting that we still have knowledge in a sense that does not involve such reasons, still constitutes in itself a very deep and implausible version of skepticism – especially when it is added, as it should be, that we also have in the same way no reasons to think that the requisite conditions for knowledge, whatever they may be, are *themselves* satisfied (since there are no plausible views in which these conditions are ones whose satisfaction could be directly established by experience).

The second dialectical argument, which I have space here only to indicate briefly, is in effect a generalization of the first. It questions whether any view that denies the existence of *a priori* reasons can account in any satisfactory way for *reasoning* itself. Here the fundamental point is that a reasoned or argumentative transition from a claim or group of claims to some further conclusion relies again on there being a good reason for thinking that a conditional claim is true, in this case one having the conjunction of the premises as its antecedent and the conclusion in question as its consequent. That such a conditional is true (or probably true) is in general not the sort of thing that could be directly established by experience, while to say that it is itself arrived at via some further process of reasoning is only to raise the identical issue about that previous step. My suggestion is that if we *never* have *a priori* reasons for thinking that if one claim or set of claims is true, some further claim must be true as well, then there is simply nothing that genuinely cogent reasoning could consist in. In this way, I suggest, the rejection of *a priori* reasons is tantamount to intellectual suicide.

### **A Priori Reasons without a Priori Insight: Moderate Empiricism**

In the space remaining, I will look briefly at two opposing positions. While virtually all serious epistemologists up to the time of Hume and Kant were rationalists in essentially the sense advocated here, the dominant position since that time and especially in the past century has been a version of empiricism, one that concedes the existence of *a priori* reasons of a sort, but claims that when properly understood, such reasons do not have the epistemological and metaphysical significance that is attributed to them by the rationalist. Instead, according to this *moderate empiricist* view, *a priori* reasons, rather than constituting insights into reality, reflect only linguistic or conceptual conventions or are merely matters of definition.

The basic idea of moderate empiricism is to explain a priori reasons in a way that drastically undercuts their significance. For this purpose, the most standard version of moderate empiricism appeals to the concept of *analyticity*, holding both (i) that all propositions for which there are genuine a priori reasons are analytic, and (ii) that an a priori reason for an analytic proposition does not require the sort of insight into the character of reality advocated by the rationalist. The problem for a would-be moderate empiricist is to find a univocal conception of analyticity in relation to which *both* of these two claims can be plausibly defended. In fact, moderate empiricists have put forth not one, but many different and not obviously equivalent conceptions of analyticity, and have tended to shift illegitimately among them depending on which of these two theses they are defending at any particular moment.

When the various conceptions of analyticity have been sorted out, they fall, I suggest, into two main groups. Some conceptions are *reductive* conceptions: they explain some cases of a priori reasons by appeal to other cases, while providing in principle no way to account for the latter cases. Here the most obvious example is the Fregean conception of an analytic proposition as one that is reducible via definitions or synonyms to a proposition of logic (where it is the propositions of logic that remain unaccounted for). Other conceptions of analyticity in effect lose sight of the main epistemological issue altogether by equating analyticity with one of the features that a proposition for which there is an immediate a priori reason undeniably has according to the rationalist account, without realizing that this fails to yield an independent account of the a priori reason. The plainest example of this mistake is the view that identifies an analytic proposition with one that is “true by virtue of meaning”: once reductive accounts are set aside, this turns out to amount to nothing more than the view that one who understands such a proposition can see directly or intuitively that it is true, where this is really just a misleading restatement of the rationalist view, not an alternative to it. In this way, I suggest, the moderate empiricist view turns out under scrutiny to be epistemologically bankrupt.<sup>3</sup>

## **The Rejection of a Priori Reasons: Radical Empiricism**

A more radical alternative is to reject the very existence of any sort of a priori reasons, a view that has been advocated by Quine. There are two main questions that need to be asked about this more radical empiricist view. One is what the arguments for it, and against the existence of a priori reasons, are supposed to be. A second is whether, especially in light of the dialectical reasons in favor of a priori reasons offered above, it is possible for radical empiricism to offer a non-skeptical epistemology.

Quine himself tends to assume that anyone who defends the idea of an a priori reason must be a moderate empiricist, and some of his arguments (in particular the famous “circle of terms” argument in Quine, 1961) really apply only to that view and are thus irrelevant here. When these are set aside, the only very clear argument that remains is one that appeals to the Duhemian thesis that claims about the world cannot be experimentally tested in isolation from each other but only in larger groups. Quine’s extreme version of this thesis is the holistic claim that nothing less than “the whole of science” can be meaningfully confronted with experience. From this he infers that any claim in the total “web of belief,” including those for which there are allegedly a priori reasons, might be “given up” in order to accommodate “recalcitrant experience,” and so, apparently, that such a

priori reasons do not exist after all (see Quine, 1961). But this conclusion simply does not follow, even if the holistic view is accepted. Quine is in effect assuming that the *only* reasons relevant to retaining or giving up a claim in the “web of belief” have to do with accommodating experience, but this is just to beg the question against the existence of independent, a priori reasons for or against such claims. And if this assumption is not made, then the rationalist can freely admit that holistic empirical reasons of this sort may count against a claim for which there is an a priori reason (or the reverse), with the ultimate outcome depending on their relative weight in a particular case – though he will also insist (see below) that the very connections among beliefs that result in the holistic web can only be understood as a priori in character.

The other main issue concerning Quine’s radical empiricism is whether it can offer a genuinely non-skeptical epistemology. While the details of Quine’s view are quite obscure, it is clear that a claim is supposed to be justified in virtue of being an element of a system of beliefs, some of whose members are appropriately related to experience and which as a whole satisfies certain further criteria, such as simplicity, scope, explanatory adequacy, fecundity, and conservatism. Consider then the conditional proposition that if a claim satisfies all of the conditions thus specified, then it is likely to be true, and ask what reason there is for thinking that this conditional proposition is itself true. Clearly such a proposition is not directly justified by experience, and to appeal to its inclusion in such a system of belief would be plainly circular. Thus either there is an a priori reason (whether immediate or resulting from a more extended a priori argument) for thinking that this conditional proposition is true or there is no reason at all. If the latter is the case, then Quine’s view fails to yield genuine justification, while if the former is the case, then his rejection of a priori reasons is mistaken. In this way it can be seen that the idea of an a priori reason is both indispensable for any justification beyond that yielded by direct experience and at least as well understood as the idea of holistic empirical justification, which turns out in fact to depend upon it. (It is worth adding that similar points could also be made about the claim that the various Quinean criteria are themselves satisfied.)

## Notes

- 1 This point did not emerge clearly in my fuller discussion of the a priori in BonJour (1998); for further discussion, see Boghossian (2001), together with my reply to Boghossian in BonJour (2001).
- 2 Which are the basis, in my view, for the justification of induction (see BonJour, 1998, chapter 7).
- 3 See BonJour (1998, chapter 2), for a much fuller discussion of this view and its problems.

## References

- Boghossian, P. (2001) Inference and insight. *Philosophy and Phenomenological Research* 63: 633–640.
- BonJour, L. (1998) *In Defense of Pure Reason: A Rationalist Account of a Priori Justification*. Cambridge: Cambridge University Press.
- BonJour, L. (2001) Replies. *Philosophy and Phenomenological Research* 63: 673–698.
- Quine, W.V. (1961) Two dogmas of empiricism. *From a Logical Point of View*, 2nd edn. Cambridge, MA: Harvard University Press.



# There Is No a Priori

Michael Devitt

## 1 Introduction

It is overwhelmingly plausible that *some* knowledge is empirical, “justified by experience.” The attractive thesis of naturalism is that *all* knowledge is; there is only one way of knowing.<sup>1</sup> But this naturalism seems to be refuted by intuitions about a range of troublesome examples drawn from mathematics, logic, and philosophy. Thus, how could experience have anything to do with justifying the belief that  $5 + 7 = 12$ ? Furthermore, it does not seem *possible* that such knowledge could be revised in the same sort of way that “All swans are white” was by the sighting of black swans in Australia. It seems that the troublesome knowledge *must* be justified in some other way, justified a priori.

So we have a motivation for abandoning naturalism and accepting the thesis that some knowledge is a priori. Yet there is a consideration against this thesis: the whole idea of the a priori seems deeply obscure. *What is it* for a belief to be justified a priori? What is the nature of this non-empirical method of justification? Without satisfactory answers the a priori is left mysterious.

In light of this, a naturalistic critic of the a priori faces two tasks: to undermine the motivation by showing that the troublesome knowledge could be empirical after all; and to demonstrate the obscurity of the a priori. Success in the second task would show that an a priori explanation of the troublesome knowledge, indeed of anything, was very unpromising. Success in the first task would show that an empirical explanation was available. So we would have a nice abduction for naturalism: the *best* explanation of that knowledge is that it is empirical.

But, first, a preliminary point. Our concern is with the *justification* of beliefs, not with their *source*. Experience is clearly not the source of many mental states: they are innate. Perhaps some of these are justified beliefs (although I doubt it). If so, the naturalist insists, beliefs of that sort were somehow justified by the experiences (broadly construed) of our distant ancestors and we have inherited that justification via natural selection.

Drawing on earlier works (Devitt, 1996, 1997, 1998, 2002), I shall attempt the two tasks. I shall conclude by considering Laurence Bonjour’s (1998, 2001a, 2001b) thorough and vigorous defense of the a priori.

## 2 Motivation

**The naturalistic alternative.** Our aim is to provide an alternative naturalistic account of the troublesome examples of allegedly a priori knowledge. With the help of Quine (1961, 1966, 1969, 1975), and before him Duhem (1954), I think that we can do this.

The key to the naturalistic alternative is breaking free of the naive picture of justification suggested by the swan example. We must view justification in a more holistic way: beliefs, even whole theories, face the tribunal of experience not alone, but in the company of auxiliary theories, background assumptions, and the like. Much evidence for this “Duhem–Quine thesis” has been produced by the movement in philosophy of science inspired by Kuhn (1962). In light of this, we have no reason to believe that whereas scientific laws, which are uncontroversially empirical, are confirmed in the holistic empirical way, the laws of logic and mathematics are not; no reason to believe that there is a principled basis for drawing a line between what can be known this way and what cannot; no reason to believe that there is, in Quine’s vivid metaphor, a seam in the web of belief.

Quine is fond of an image taken from Otto Neurath. He likens our web of belief to a boat that we continually rebuild while staying afloat on it. We can rebuild any part of the boat – by replacement or addition – but in so doing we must take a stand on the rest of the boat for the moment. So we cannot rebuild it all at once. Similarly, we can revise any part of our knowledge – by replacement or addition – but in so doing we must accept the rest for the time being. So we cannot revise it all at once. And just as we should start rebuilding the boat by standing on the firmest parts, so also should we start rebuilding our web. So we normally take the propositions of logic and mathematics for granted. Still, each of these propositions is in principle revisable in the face of experience: taking a stand on other such propositions, and much else besides, we might contemplate dropping the proposition.

Given this naturalistic alternative, we have no need to turn to an a priori explanation of our knowledge of mathematics, logic, and the like. The original intuitions were really that this knowledge is not justified in some *direct* empirical way. Those intuitions are preserved. Yet we can still see the knowledge as empirical: it is justified empirically in an *indirect* holistic way.

I shall develop this account by answering objections.

**Objection 1:** “You are surely not suggesting that these few hand-waving remarks about the empirical nature of mathematics come close to solving the epistemological problem of mathematics.”

No, I am not. But there are two reasons why this is beside the point. First, as Georges Rey (1998) is fond of pointing out, we are not close to solving the epistemological problem of *anything*. Since we do not have a serious theory that covers even the easiest examples of empirical knowledge – examples where experience plays its most direct role – the fact that we do not have one that covers the really difficult examples from mathematics hardly reflects on the claim that these are empirical too. We all agree that there *is* an empirical way of knowing. Beyond that, the present aim needs only the claim that the empirical way is holistic. *We have no reason to believe that a serious theory would show that, whereas empirical scientific laws are confirmed in the holistic empirical way, the laws of mathematics are not.*

Second, there is a special reason for not expecting the epistemological problem of mathematics to be anywhere near solved: the *metaphysical* problem of mathematics – what mathematics is *about* – remains so intractable. How could we solve the epistemological problem when we remain in such darkness about the metaphysical one? *We no*

*longer have any reason to think that, if we solved the metaphysical problem, the epistemological problem would not be open to an empirical solution.*

**Objection 2:** “We need to explain our knowledge of *necessities*; for example, that *necessarily*  $5 + 7 = 12$ , that *necessarily* all bachelors are unmarried. Yet all we can know from experience is how things are – how they are in the actual world – not how they *must* be – how they are in all possible worlds.”

But why should we accept that necessities can only be known a priori? Prima facie, *some* necessities are known empirically; for example, that water is necessarily  $H_2O$  and that Hesperus is necessarily Phosphorus. Indeed, science seems to be discovering necessities all the time. Now, one might respond that what science discovered was only that water *is*  $H_2O$ , not that it *necessarily* is; the necessity is not an empirical discovery. But, again, why should we accept this? Certainly, we do not simply *observe* the necessity of water being  $H_2O$ . But we do not simply observe most scientific facts: we discover them with the help of a lot of theory. And that, according to the naturalist, is how we discover necessities. More needs to be said of course, but to say it we would need to take a stand on the *metaphysical* problem of necessity. That problem is another difficult one. There is no reason to believe, however, that if we solved it we would not be able to explain our knowledge of necessities empirically. The situation for that knowledge is analogous to that for our knowledge of mathematics.

**Objection 3:** This objection concerns logic. It arises out of the dominant theme of Bonjour’s defense of the a priori: that “the rejection of any sort of *a priori* justification leads inexorably to a severe skepticism” and to the undermining of “reasoning or argument in general” (Bonjour, 2001a, pp. 625–626). Bonjour’s discussion suggests the following objection. “On your Quinean alternative, experience justifies beliefs in the interior of the web via links with beliefs at the periphery, via links with beliefs ‘close to experience.’ But these justifications depend on the links themselves being justified: clearly a belief is not justified by other beliefs unless those others give it *genuine support*. The objection to your alternative is then: the justification of these links has to be *a priori*; it could not come from experience.” Indeed, Bonjour (2001b, p. 679) claims, “if there is no *a priori* insight ... no prediction will follow any more than any other ... any ... sort of connection between the parts of the system will become essentially arbitrary.” “[T]he rejection of all *a priori* justification is tantamount to intellectual suicide” (Bonjour, 2001a, p. 626). In brief, the objection is that logic must be seen as a priori because we need logic to get evidence for or against anything.

Many would agree with this objection. I have three responses. But first I will give what I hope is some fairly uncontroversial background.

The links that hold the web of belief together reflect a set of rules that are part of “an evidential system” (Field, 1996, 1998). As a result of nature and nurture each person embodies such a system which governs the way she arrives at her beliefs about the world. A system must include dispositions to respond selectively to perceptual experiences and to infer according to certain rules. A likely example of a rule is *modus ponens*. So, a person embodying an evidential system *S* containing this rule is disposed to infer according to the pattern:

If  $p$  then  $q$ ,  
 $p$ ,  
So,  $q$ .

Now, the objection is surely right in claiming that for a person using  $S$  to have justified beliefs, its rules have to be good ones. This is not to say that she must know the epistemological theory,

$T$ :  $S$  is a good evidential system,

for her beliefs to be justified. So it is not to say that she must know

$MP$ : *Modus ponens* is a valid inference

for her beliefs to be justified by *modus ponens* arguments. Indeed, as Lewis Carroll made clear a century ago, the demand for this sort of extra premise in an argument leads to a regress (see Boghossian, 2001, for a discussion). And it is just as well that our person is not required to have this epistemological knowledge for, if she is an ordinary member of the folk, she is unlikely to have given such matters much thought. Still, it is certainly appropriate to give them thought. So it is appropriate for the person to stand back from her arguments and ask some epistemological questions. What are the rules of  $S$ ? This is a question in *descriptive* epistemology. Are the rules of  $S$  good? How do we know  $T$  and  $MP$ ? These are questions in *normative* epistemology. Any answers to such questions will be further beliefs, additions to her web. I take the point of the objection to be that normative answers must be obtained by a priori insight.

Because *modus ponens* is a deductive rule, it is a rather misleading example of a system's rules. Given any theory and a body of evidence that it entails, we can easily construct rival theories that entail the same evidence. So we need more than deductive rules to choose between these theories and avoid skepticism: we need non-deductive "ampliative" rules. And these are rules that we don't have much insight into, whether a priori or not. It is largely because of this ignorance that, as already noted, we lack a serious epistemological theory. We can, of course, wave our hands and talk of enumerative induction, abduction, simplicity, and the like, but we are unable to characterize these in the sort of detail that would come close to capturing the rules that must constitute our actual evidential systems; for example, we are unable to specify when an explanation is good, let alone the best, or when we should take the belief that all observed  $F$ s have been  $G$  to justify the belief that all  $F$ s are  $G$ . Aside from that, some of these vague rules are controversial; for example, scientific realists love abduction, Bas van Fraassen does not. In sum, when we move beyond deduction, we have few if any specific and uncontroversial rules to be insightful about. The non-skeptics among us will share the very general insight that, *whatever the rules of our evidential system may be*, those rules are for the most part pretty good. So, if  $S$  is that largely unknown system, we believe  $T$ .

**Response 1.** Is the objection claiming that  $T$  is known a priori? If so, the claim hardly seems tempting. It seems more plausible to view our general insight that  $T$  is true as supported by the empirical success of  $S$ , whatever  $S$  may be. Similarly, someone afloat on a boat may not know the methods by which it was built but, noting its seaworthiness, infers that the methods, whatever they were, are good. In sum, when we focus on the

largely unknown ampliative parts of *S*, our confidence in *S* seems as empirical as anything. To that extent, *T* does not even appear to be supported by a priori insight.

"But what about the specific deductive rules that we do have insight into, rules like *modus ponens*? Even if our overall confidence in *S* is empirical, our confidence in these deductive parts is a priori. We know *MP* a priori at least."

**Response 2.** But why must we see the support for the deductive rules as different in principle from that for the ampliative rules? They are all rules of *S*, they are all needed to avoid skepticism, and we can see them all as supported by the overall empirical success of *S*. Then the justification of the deductive parts of *S* is no different in principle from that of the ampliative parts. Similarly, all parts of *S* are empirically *revisable*. Thus, suppose that experience leads us to abandon *T* in favor of *T'*, a theory that recommends an evidential system *S'* built around a non-classical logic. Then clearly we should use *S'* instead of *S*. In this way our logical practices are themselves open to rational revision in the light of experience. These practices are far from "arbitrary": they are recommended by an experience-based epistemology.

Still, many will feel that I have not yet got to the heart of the objection. "On the one hand, you talk of *T* being supported by the empirical success of *S*. Yet that alleged support must come via *S* itself. So, the attempt to support *T* is circular. On the other hand, you talk of the possibility of experience leading us to abandon *T* in favor of *T'*. Yet experience must be brought to bear on *T* by using *S* and so could not show that *T* is false and hence that we ought not to use *S*. The attempt to refute *T* is self-defeating."

**Response 3.** In considering the circularity charge we need to follow Braithwaite (1953, pp. 274–278) in distinguishing "premise-circularity" from "rule-circularity." An argument is premise-circular if it aims to establish a conclusion that *is assumed as a premise in that very argument*. Premise-circularity is clearly reprehensible. But my argument for *T* is not guilty of it because it does not use *T* as a premise. An argument is rule-circular if it aims to establish a conclusion that *asserts the goodness of the rules used in that very argument*. My argument tries to establish *T* which asserts the goodness of *S*, the system used in that argument to establish *T*. So the argument is certainly rule-circular. This is worrying initially but is there a good reason to think that it is in fact reprehensible? I agree with those who have argued that there is not (Van Cleve, 1984; Papineau, 1993; Psillos, 1998). Guided by the Neurath image, we accept the non-epistemological part of our web for the moment and seek to justify the epistemological part, *T*. And that justification is governed by just the same rules that govern the justification of anything, the rules of *S*.

The self-defeat charge is also worrying initially. Yet there are reasons for thinking that we can indeed show an evidential system to be defective using that very system.

First, it seems undeniable that our evidential systems have changed. (i) A good deal of the impressive scientific progress over the past three centuries has been in improved methodologies: we have learnt a vast amount not only about the world but also about how to learn about the world. As a result, much education of the young scientist is in these methodologies: think of physics and psychology, for example. (ii) Educated folk have tried to adjust their thinking in light of evidence that we normally tend toward certain sorts of irrationality; for example, counter-induction, and ignoring base rates in thinking about probabilities. (iii) Even our deductive practices have been affected by the rise of modern logic.

Next, the process of making any of these system changes must have been governed by some evidential system, the one that was then current. So, that system was used to

establish an epistemological thesis that led to the system's replacement. These examples give us good reason to think that an evidential system could be used rationally to undermine itself. Accepting the non-epistemological part of our web and governed by *S* as usual, we find *T* wanting and so replace it and the system *S* that it recommends.

Despite this response, worries about circularity and self-defeat may persist. It helps to remove them to note that if the worries were appropriate, analogous ones would be just as appropriate *if *T* were justified by a priori insight*. For, if *T* were thus justified, a priori insights would be part of our evidential system *S*. We could then generate a circularity worry. The argument for *T*, which asserts the goodness of *S*, uses part of *S* to establish *T*. And we could generate a self-defeat worry. The apriorist must allow that we could abandon *T* in favor of *T'* on the basis of a *a priori* insight, part of *S*. So *S* is used to establish *T'*, which leads to its own replacement. If these circularity and self-defeat charges are unworrying for the apriorist, the analogous ones are surely so for the naturalist.

Faced with the circularity and self-defeat charges we could conclude that our evidential system *is* unjustified. But this throws the baby out with the bathwater. Although the charges are worrying in the beginning, I have argued that they should not be in the end. In any case naturalism and apriorism are on an equal footing in dealing with them.

**Objection 4.** "Suppose that it really is the case that any belief can be confirmed or disconfirmed by experience in the Duhem–Quine way. This does not show that agreement with experience is the *only* consideration relevant to the belief's rational acceptance and rejection. Hence it does not show that there is no a priori justification. By supposing that it does show this you beg the question."

The objection misses the main point of the naturalistic alternative. That point is not to show that there is no a priori knowledge but to remove the motivation for thinking that there must be. Everyone agrees that there is an empirical way of knowing. The Duhem–Quine thesis, supported by the history of science, is that this way of knowing is holistic. I have argued that our troublesome knowledge of mathematics, logic, and the like can be accommodated within this holistic empirical picture. We are far short of a detailed epistemology for this knowledge, of course, but we are far short of a detailed epistemology for any knowledge. Now, if I am right about all this, we have clearly removed the theoretical need to seek another, a priori, way of knowing. This is certainly *part* of the case against the a priori, but it cannot stand alone. The rest of the case is that the whole idea of the a priori is deeply obscure.

BonJour and many others will think that this empirical justification of the troublesome knowledge is inadequate. They will demand a justification that is stronger and that can only be met by appeal to the a priori. I think that this demand might be rational if there were any grounds for optimism about the a priori. But, I shall now argue, there are no such grounds, only grounds for pessimism. If this is right, the demand is not rational.

### 3 Obscurity

The aim in this section is to show that the whole idea of the a priori is too obscure for it to feature in a good explanation of our knowledge of anything. If this is right, we have a nice abduction: the *best* explanation of all knowledge is an empirical one.

We are presented with a range of examples of alleged a priori knowledge. But what are we to make of the allegation? What is the nature of a priori knowledge? We have the characterization: it is knowledge “*not* derived from experience” and so *not* justified in the empirical way. But what we need if we are to take the a priori way seriously is a *positive* characterization, not just a negative one. We need to describe a process for justifying a belief that is different from the empirical way and that we have some reason for thinking is actual. We need some idea of what a priori knowledge *is* not just what it *isn't*.

Why? After all, I have been emphasizing how little we know about *empirical* justification. So why pick on the a priori? The answer is that there are two crucial differences in the epistemic status of the two alleged methods of justification. First, the existence of the empirical method is not in question: everyone believes in it. In contrast, the existence of the a priori way is very much in question. Second, even though we do not have a serious theory of the empirical way, we do have an intuitively clear and appealing general idea of this way, of “learning from experience.” It starts from the metaphysical assumption that the worldly fact that *p* would make the belief that *p* true. The empirical idea then is that experiences of the sort that would be produced by that fact are essentially involved in the justification of the belief. In contrast, we do not have the beginnings of an idea of what the a priori way might be; we lack not just a serious theory but *any idea at all*.

The difficulty in giving a positive characterization of a priori knowledge is well demonstrated by the failure of traditional attempts based on analyticity. Let the example of alleged a priori knowledge be our belief that all bachelors are unmarried. According to the tradition, the content of the concept <bachelor> “includes” that of <unmarried>, thus making the belief analytic. This seemed promising for an account of a priori knowledge because it was thought that, simply in virtue of *having* a concept, a person was in possession of a “tacit theory” *about* the concept; in virtue of having <bachelor>, a person tacitly knew that its content included that of <bachelor>. So a person’s conceptual competence gave her privileged “Cartesian” access to facts about concepts. The required non-empirical process of justification was thought to be one that exploited this access, a reflective process of inspecting the contents of concepts to yield knowledge of the relations between them, which in turn yielded such knowledge as that all bachelors are unmarried. This alleged process is that of “conceptual analysis.”

Even if we grant that we have this Cartesian access to conceptual facts, the account fails. These facts would not justify the proposition that all bachelors are unmarried *unless the proposition that all unmarrieds are unmarried were justified*. But where does the justification for this proposition come from? It does no good to say, rightly, that the proposition is a *logical* truth, for what justifies logical truths? No satisfactory non-empirical account has ever been given of how they can be justified. Without such an account we have not described a non-empirical way of knowing.

In any case, we should not grant the Cartesian view that competence gives privileged access to contents, despite its great popularity. I urge a much more modest view of competence according to which it is an *ability or skill* that need not involve any tacit theory, any semantic propositional knowledge; it is knowledge-how not knowledge-that (Devitt, 1996). Why then should we believe the immodest Cartesian view, particularly since it is almost entirely unargued?

The content of a person's thought is constituted by relational properties of some sort: "internal" ones involving inferential relations among thoughts and "external" ones involving certain direct causal relations to the world. Take one of those relations. Why suppose that, simply in virtue of her thought having that relation, reflection must lead her to *believe that* it does? Even if reflection does, why suppose that, simply in virtue of that relation partly constituting the content of her thought, reflection must lead her to *believe that* it does? Most important of all, even if reflection did lead to these beliefs, why suppose that, simply in virtue of her competence, this process of belief formation *justifies* the beliefs and thus turns them into *knowledge*? The supposition seems to be gratuitous. We need a plausible explanation of this allegedly non-empirical process of justification.

#### 4 Bonjour's Rationalism

I turn finally to Bonjour's wonderfully forthright approach to the a priori. First, he has no more faith in attempted explanations in terms of analyticity than I have and gives an excellent critique of their failings (Bonjour, 1998, chapter 2). Indeed, he is rather contemptuous of these attempts to make the a priori palatable to the modern mind. Bonjour is an unabashed old-fashioned rationalist (apart from embracing the fallibility of a priori claims). He rests a priori justification on "rational insight": "*a priori* justification occurs when the mind directly or intuitively sees or grasps or apprehends ... a necessary fact about the nature or structure of reality" (Bonjour, 1998, pp. 15–16). So, our problem of explaining the a priori becomes that of explaining rational insight. Where is the *justification* to be found in this quasi-perceptual process of apprehending a necessary fact?

Bonjour (1998, p. 107) is only too well aware that most philosophers find this rationalism extremely mysterious. In response, he offers the beginnings of an explanation based on the unpopular thesis that a thought's content is an intrinsic property of the thought (1998, pp. 180–186). In my view (Devitt, 1990, 1996, 2001), this thesis thoroughly deserves its unpopularity. Aside from that, the explanation based on it is very obscure, as commentators have pointed out (Boghossian, 2001; Rey, 2001). But we need not dwell on this explanation because Bonjour himself does not claim much for it. Indeed, he accepts that "we do not presently have anything close" to an adequate explanation of rational insight (Bonjour, 2001b, p. 674). That seems to leave rationalism in trouble. Not according to Bonjour (1998, p. 31): "the supposed mystery pertaining to rationalism ... has been ... greatly exaggerated"; allegations that rationalism is "objectionably mysterious, perhaps even somehow occult ... are very hard to take seriously" (1998, pp. 107–108); "the capacity for rational insight, though fundamental and irreducible, is in no way puzzling or especially in need of further explanation" (1998, p. 16).

What is the source of this extraordinary confidence in an unexplained and apparently mysterious capacity? It comes partly, of course, from the earlier-noted view that to deny the a priori is to commit "intellectual suicide." But it comes also from "the intuitive or phenomenological appearances" of rational insight (1998, p. 107): Bonjour thinks that these appearances, when examining examples of alleged a priori knowledge, provide a *prima facie* case for rationalism that is "extremely obvious and compelling" (1998, p. 99).

So, Bonjour thinks that there just *has* to be rational insight even if we can't explain it. In contrast, I think, for the reasons set out in section 2, that there does not have to



be, and the apparent hopelessness of explaining rational insight shows that there isn't any. I shall end with a few more remarks about that hopelessness.

First, a word on the phenomenology. Bonjour denies that there is any mystery in "our cognitive experience" (1998, p. 108) when we have "direct insight into the necessary character of reality" (1998, p. 107). He may be right. But the mystery lies in the claim that *this experience is an a priori insight*. Nothing in the phenomenology supports *that* or, indeed, *any* view of what justifies the insight. In particular, it does not show that the insight is not justified in a holistic empirical way. This theoretical issue is way beyond anything in the phenomenology.

Turn next to that theoretical issue. A human mind/brain forms beliefs about the external world. In virtue of what is any belief justified and hence likely to be true? We have a rough idea of where to find an empirical answer. We look at the way in which the beliefs are related to the experiences that the world causes. Justified beliefs are appropriately sensitive, via experience, to the way the world is. Many instruments – thermometers, voltmeters, and so on – are similarly sensitive to the world. Of course, the mind/brain differs from these instruments: beliefs are much more complex than the "information states" of instruments and their sensitivity to the world is mediated, in a holistic way, by many others. Still, the mind/brain is similar enough to the instruments to make empirical justification quite unmysterious, despite the sad lack of details.

The contrast with a priori justification is stark. What sort of link *could* there be between the mind/brain and the external world, other than via experience, that would make states of the mind/brain likely to be true about the world? What non-experiential link to reality could support insights into its necessary character? There is a high correlation between the logical facts of the world and our beliefs about those facts which can only be explained by supposing that there are connections between those beliefs and facts. If those connections are not via experience, they do indeed seem occult.

At this point, it remains a mystery what it would be for something to be known a priori. Any attempt to remove this mystery must find a path between the Scylla of describing something that is not a priori knowledge because its justification is empirical and the Charybdis of describing something that is not knowledge at all because it has no justification.<sup>3</sup> The evidence suggests that there is no such path. Hankering after a priori knowledge is hankering after the unattainable.

The nice abduction is established: our knowledge of mathematics, logic, and the like cannot be explained a priori; an empirical explanation of it is the best.

## Acknowledgments

Thanks for comments from John Bigelow, Thomas Bontly, Radu Dudau, Hartry Field, Gilbert Harman, Frank Jackson, Hilary Kornblith, David Papineau, Stathis Psillos, Elliott Sober, Kim Sterelny, and Stephen Stich.

## Notes

- 1 Bonjour (1998) calls this *epistemological* naturalism "radical empiricism." It should not be confused with *metaphysical* naturalism, a reductive doctrine like physicalism.

- 2 Although Quine's influence on the views I will present is large and obvious, I am not concerned to argue that these views are precisely his nor to defend everything he has to say on the a priori and related topics.
- 3 I argue (1998), in effect, that Rey's attempt (1998) to give a reliabilist account of the a priori falls victim to Charybdis.

## References

- Boghossian, P. (2001) Inference and insight. *Philosophy and Phenomenological Research* 63: 633–640.
- BonJour, L. (1998) *In Defense of Pure Reason: A Rationalist Account of a Priori Justification*. Cambridge: Cambridge University Press.
- BonJour, L. (2001a) *Precis of In Defense of Pure Reason*. *Philosophy and Phenomenological Research* 63: 625–631.
- BonJour, L. (2001b) Replies. *Philosophy and Phenomenological Research* 63: 673–698.
- Braithwaite, R.B. (1953) *Scientific Explanation*. Cambridge: Cambridge University Press.
- Devitt, M. (1990) Meanings just ain't in the head. In *Meaning and Method: Essays in Honor of Hilary Putnam*, ed. G. Boolos. New York: Cambridge University Press.
- Devitt, M. (1996) *Coming to Our Senses: A Naturalistic Program for Semantic Localism*. New York: Cambridge University Press.
- Devitt, M. (1997) *Realism and Truth*, 2nd edn with a new afterword. Princeton, NJ: Princeton University Press.
- Devitt, M. (1998) Naturalism and the a priori. *Philosophical Studies* 92: 45–65.
- Devitt, M. (2001) A shocking idea about meaning. *Revue Internationale de Philosophie* 208: 449–472.
- Devitt, M. (2002) Underdetermination and realism. In *Realism and Relativism: Philosophical Issues, Volume 12*, ed. E. Villanueva and E. Sosa. Oxford: Blackwell.
- Duhem, P. (1954) *The Aim and Structure of Physical Theory*, trans. P. Wiener. Princeton, NJ: Princeton University Press (originally published 1906).
- Field, H. (1996) The a priority of logic. *Proceedings of the Aristotelian Society* 96: 1–21.
- Field, H. (1998) Epistemological nonfactualism and the a priority of logic. *Philosophical Studies* 92: 1–24.
- Kuhn, T.S. (1962) *The Structure of Scientific Revolutions*. Chicago: Chicago University Press.
- Papineau, D. (1993) *Philosophical Naturalism*. Oxford: Blackwell.
- Psillos, S. (1998) *Scientific Realism: How Science Tracks Truth*. New York: Routledge.
- Quine, W.V. (1961) Two dogmas of empiricism. In *From a Logical Point of View*, 2nd edn. Cambridge, MA: Harvard University Press.
- Quine, W.V. (1966) The scope and language of science. In *The Ways of Paradox and Other Essays*. New York: Random House.
- Quine, W.V. (1969) Epistemology naturalized. In *Ontological Relativity and Other Essays*. New York: Random House.
- Quine, W.V. (1975) The nature of natural knowledge. In *Mind and Language*, ed. S. Guttenplan. Oxford: Clarendon Press.
- Rey, G. (1998) A naturalistic a priori. *Philosophical Studies* 92: 25–43.
- Rey, G. (2001) Digging deeper for the a priori. *Philosophy and Phenomenological Research* 63: 649–656.
- Van Cleve, J. (1984) Reliability, justification, and the problem of induction. In *Midwest Studies in Philosophy, Volume IX: Causation and Causal Theories*, ed. P.A. French, T.E. Uehling Jr, and H.K. Wettstein. Minneapolis: University of Minnesota Press.

# Reply to Devitt

Laurence Bonjour

Professor Devitt's case against the *a priori* involves two main points: (i) that all genuine knowledge, including "troublesome examples" like the ones cited in my initial essay, can be accounted for entirely in empirical terms, at least mainly via an appeal to what he refers to as the "holistic empirical way" of knowing, so that an appeal to the *a priori* is unnecessary and unmotivated; and (ii) that the idea of *a priori* knowledge or justification or reasons is "deeply obscure," too obscure to provide a satisfactory explanation of anything, even if there were anything further that needed to be explained. I will respond to these in turn.

(i) Devitt asserts in several places that the existence of empirical knowledge is not in question. This is indeed obviously true if by "empirical knowledge" is meant knowledge whose justification depends at least *partially* on experience. But if it means knowledge whose justification derives *entirely* from experience, with no need for any *a priori* element, then it is far from clear that there is very much such knowledge and not entirely beyond question that there is any at all. In particular, the justification of scientific laws is *not* "uncontroversially empirical" in this latter sense, since inductive reasoning or something like it is also required. Thus, contrary to what Devitt sometimes seems to suggest, the central issue as regards his point (i) is not just whether the "holistic empirical" view can account for the "troublesome examples", but whether it can account for the justification of at the very least large portions of our apparent knowledge in an *entirely* empirical way.

Here the fundamental question is what holds together the various elements of the holistic system to which he appeals (including "auxiliary theories, background assumptions, and the like") so that the whole system (and not just its "edges") connects with experience in such a way as to yield a good reason for thinking that some particular belief or theory embedded in it is true (where these connections, as Devitt notes, will involve much more than logic in the narrow sense). The connections that are relevant to any particular issue of justification may be thought of piecemeal in terms of various logical and quasi-logical relations between them or they may be summed up into one overall conditional as at the end of my initial essay, but either way some reason is seemingly needed for thinking both that various specific connections hold and that when taken together they are of the right sort for the overall system to confer justification on those of its elements that are not matters of direct experience. And it is hard to see how such reasons can be anything but *a priori*, since they are surely not a matter of direct experience (and cannot without circularity be based on any appeal to the holistic picture itself).

Devitt offers two main responses to this sort of objection, neither of which seems to me to be satisfactory. (His other response, the second in his list, is a response to an intermediate rejoinder and does not bear directly on the main issue.) The first response is that we are unable to characterize in detail the specific rules, over and above formal logic in the narrow sense, that we follow in arriving at our "web of belief," and so cannot plausibly be said to have *a priori* insight into their correctness; and it is most implausible that we have any such insight into the general claim (Devitt's *T*) that our

overall system of rules is a good one. I agree that we have no insight of the latter sort; indeed any claim of this sort could, as far as I can see, be justified only indirectly by appeal to more specific claims about the particular rules involved. But is it really so clear that we have no insight into the correctness of the specific rules that we use, even if we are unable to formulate those rules in precise detail? That when a scientist reasons inductively or abductively, he has no insight into the cogency of his reasoning, but is simply flying blind, doing what (for whatever reason) comes naturally, but with no reason to think that it is likely to take him in the direction of the truth? I myself find this picture most implausible. But the further thing to be said is that if it is right, then the result is simply skepticism, since this lack cannot be remedied in the way that Devitt suggests: by reasoning that the empirical success of our system of rules is best explained by the supposition that it is generally a good one. For that sort of abductive argument – like the one that Devitt uses at the end to arrive at his main conclusion – is equally one that on his view we have no reason to think is cogent.

Devitt's second main response attempts to answer the charge that it is circular to appeal to the very set of holistic rules whose correctness is at issue in establishing that correctness. (Note that, contrary to what he suggests, the abductive argument for *T* just discussed does not in fact do this in any very clear way.) Devitt follows Braithwaite in distinguishing “premise-circularity” from “rule-circularity,” with the point being that while premise-circularity is clearly objectionable, this is not so clearly true of rule-circularity. Though I have no space for an extended discussion, this point seems to me to be clearly and indeed obviously mistaken. If the issue is whether following that set of rules, operating in the way that the “holistic empirical” approach sanctions, gives us any reason to think that our results are true, it is obviously no help at all to be told that the claim that those results are likely to be true (or that the rules are good ones) can be arrived at by employing the very rules whose truth-conduciveness is in doubt. Such an argument may not beg the question in quite the sense that a premise-circular one does, but it is just as unsatisfactory in relation to the question at issue. (Much of this also applies to the issue of self-defeat, though I again have no space to go into that issue in any detail. I think that changes in our system of rules obviously occur, but that such change is less holistic than Devitt would have it and so does not raise issues of circularity: the rule or alleged insight being rejected does not contribute to the reason for its rejection.)

Devitt adds the further remark that the same issue about circularity would apply to the appeal to a priori insight, but this also seems to me mistaken. He has in mind a view according to which the overall claim *T* is justified by a priori insight, and I have already rejected that. Beyond that, the basic point is that a priori insight is atomistic rather than holistic in character, so that neither the issue of circularity nor that of self-defeat applies in any clear way. Alleged a priori claims can be defeated by a combination of other such insights (plus, sometimes, empirical premises), but the main positive case for such a claim rests only on the immediate insight itself.

Thus the basic point of the objection to the holistic empirical view seems to me to remain unscathed: such a view can give no satisfactory account of how the fact that a belief satisfies its requirements constitutes a reason to think that it is true – or indeed of how we can have reason to think that its requirements are indeed satisfied. In this way, such a view leads only to a deep and pervasive skepticism.

(ii) Devitt's second main point is that the idea of a priori justification – that is, of an a priori reason – is “deeply obscure,” where the obscurity in question seems to be at

least primarily that the source of justification has not been adequately explained. Granting that empirical justification is also not very well understood, Devitt nonetheless claims that there are “two crucial differences” between empirical and allegedly *a priori* justification that make the lack of understanding more telling in the latter case. The first is that “the existence of the empirical method is not in question.” But given the argument of my initial contribution and of part (i) of this reply, the existence of *purely* empirical justification for any claim that is not a matter of direct observation is very much in question – indeed seems to be clearly ruled out. And the existence of *partially* empirical justification, I have argued, must depend on a *a priori* justification and so can hardly be taken as a reason for skepticism about it. This conclusion is reinforced by Devitt’s second “crucial difference”: that we have “an intuitively clear and appealing general idea” of how empirical knowledge and justification are possible, namely that experiences produced by the fact that would make the belief true are what justifies it. Plainly this applies only to the (relatively rare) cases in which experiential justification is direct and does nothing to illuminate the more indirect cases. Thus Devitt’s “two crucial differences” do very little to show that there is any special problem with the *a priori*.

Beyond that, what does the alleged “mystery” really amount to? Devitt seems to me to be simply rejecting the idea that merely finding something to be intuitively necessary can ever constitute in itself a reason for thinking that it is true – albeit one that is defeasible by further considerations (which would themselves ultimately depend on further such insights). My contrary suggestion would be that apart from the limited class of claims that are directly justified by experience, there is in the end simply no other form that a reason for thinking that something is true, whether atomistic or holistic in character, could possibly take. To be sure, such insights can be combined in complicated ways to yield more elaborate arguments of various sorts, but in the end the various steps in such arguments, together with any premises or principles that they invoke, can only be justified by appeal to the very same sort of a *a priori* insight – if, that is, they are to be justified at all. No doubt it would be nice to have a fuller, richer account of a *a priori* insight and how it works. But given both intuitively compelling examples and an argument showing such insight to be essential to any but the most minimal cognitive functioning, the absence of such an account does not yield in any clear way a reason for denying its existence.

## Reply to BonJour

*Michael Devitt*

“There Is no *a Priori*” is, in effect, partly a response to the main arguments in Laurence BonJour’s “In Defense of *a Priori* Reasons.” So this reply is largely to his “Reply to Devitt.”

My essay has two criticisms of the *a priori*: (i) we don’t need it because all justification could be empirical; (ii) the whole idea of the *a priori* is deeply obscure.

The central issue over (i) concerns rule-circularity and self-defeat. I shall focus on rule-circularity, as does BonJour. An argument is rule-circular if it aims to establish a

conclusion that asserts the goodness of the rules used in that very argument. I claimed (Response 3) that although rule-circularity is initially worrying it is not in fact reprehensible. I cited some arguments for this claim but did not give any myself (although I did give one for self-defeat, for the view that rules could govern a procedure that supplies a rational basis for their own revision). BonJour rejects my claim as “clearly and indeed obviously mistaken” but also gives no argument. What hangs on this unargued matter?

First, if naturalism needs to rely on a rule-circular argument, my claim had better be right. Now, taking *S* to be the set of rules constituting our actual evidential system, I did accept that my naturalistic argument for the epistemological thesis

*T*: *S* is a good evidential system

was rule-circular. But, interestingly, BonJour’s discussion raises the possibility that this acceptance was too hasty. After all, the metaphor of Neurath’s boat suggests that the epistemological claim that *a certain one* of *S*’s rules, say *R*, is good could be justified by an argument that uses *other* rules of *S* but not *R* itself; thus perhaps one could use inductive and deductive rules to justify abduction. There would be nothing circular about that. So if we could do that for claims about *each* rule of *S* in turn, we could justify *T* without rule-circularity. And the justification would be naturalistically kosher. Still, accomplishing this does seem a very tall order, particularly when one remembers that *S* must contain rules governing the choice between *T* and a rival *T*’ that recommends a different system *S*’. Given our ignorance of *S* we cannot be certain that the naturalist must accept rule-circularity but I think it very likely that she must.

Second, if rationalism *also* needs to rely on a rule-circular argument then BonJour had better hope that I am right about them! I argued that rationalism does indeed rely on rule-circularity. BonJour disagrees. His “basic point is that a priori insight is atomistic rather than holistic in character.” So, we justify the overall claim *T* only indirectly by justifying particular claims about the rules that make up *S* with the result “that neither the issue of circularity nor that of self-defeat apply in any clear way.” BonJour is wrong about this.

*S* is a system of rules for belief-formation. We all agree that *S* includes rules governing responses to perceptual experiences, ampliative rules, and deductive rules. According to the rationalist, *S* also includes a rule yielding a priori insights. Now the challenge posed by the skeptic is to say why any rule, *R*, is good. BonJour responds to this challenge by appealing largely, if not entirely, to a priori insight; the mind directly or intuitively grasps the necessary fact that *R* is good. Whatever its other problems, there need be no circularity about this *provided R is not the rule for a priori insight itself*. Where *R* is that rule, the rule-circularity is obvious. So BonJour’s move to atomism does not avoid rule-circularity.

Criticism (i) aimed to show that all beliefs could be justified empirically, thus removing the motivation for the a priori. Among these beliefs are epistemological ones about the goodness of rules for belief formation. I doubt that all these epistemological beliefs could be justified empirically if rule-circularity is disallowed but there is no reason to think that they could not if rule-circularity is allowed. BonJour is in no position to disallow rule-circularity because his own rationalism depends on it. For, if he had a justification for believing that a priori insight was a good method of belief formation, the justification would be an a priori insight.

In response to (ii), BonJour continues to minimize the obscurity of the a priori, wondering what its “alleged ‘mystery’” really amounts to. It is important to note something he *does not* do: he does not attempt an explanation that might reduce the mystery. We should not be surprised at this failure if I am right that nothing *can* reduce the mystery.

In charging that the a priori is deeply obscure I am, according to BonJour, simply rejecting the idea that merely finding something to be intuitively necessary can ever constitute in itself a reason for thinking that it is true.” But I am not simply rejecting this: I am *demanding an explanation of how it could be so*. How could this intuitive process *justify* something unless the process is empirical? The a priori is mysterious because we do not have even a hint of a satisfactory answer. It seems like magic that a process in someone’s mind can justify her belief in an external worldly fact without that justification arising from some sort of experiential link to that fact.

Those are my main points, but I have one more.

In (i) I took BonJour to be rightly claiming that for a conclusion to be justified by an inference, the inference must be good, but I argued that he was wrongly claiming that our justification of its goodness must be a priori. This disagreement concerned that justification *whoever* provided it. However, BonJour’s (1998) actual requirement for a justified conclusion was that *the very person making the inference* accompany it with an a priori insight into its goodness. Paul Boghossian (2001), following Lewis Carroll, pointed out that this requirement that a proposition *about* the inference accompany the inference leads to an unstoppable regress. BonJour has responded to this point with a very curious move: “it is often and quite possibly always a mistake to construe [a priori insights] as *propositional* in form”; “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.” BonJour’s requirement, thus construed, has a role in his responses to both (i) and (ii).

This construal seems to commit BonJour to an “a priori knowing-how,” something that surely makes no sense. The relations between propositions in an inference are not propositions, of course, but any insights *about* those relations are *essentially* propositional, having contents specified by “that”-clauses (e.g., “that *p* follows from *q*”) like any other propositions.

BonJour’s requirement was mistaken from the beginning. For an inference to justify a person’s conclusion it simply has to *be* good. In an epistemological moment the person may indeed have the insight that the inference *is* good. Still, the justification of her conclusion does not depend on her having this insight. And, as I argued in (i), we should see such insights as empirical anyway.

Conclusion: BonJour’s response to (i) does not undermine my argument that belief in the a priori is unmotivated. And his response to (ii) leaves the a priori as obscure as ever.

## References

- Boghossian, P. (2001) Inference and insight. *Philosophy and Phenomenological Research* 63: 633–640.
- BonJour, L. (1998) *In Defense of Pure Reason: A Rationalist Account of a Priori Justification*. Cambridge: Cambridge University Press.

# Last Rejoinder

*Laurence Bonjour*

I have space only for three very quick points and two slightly more extended ones.

First, Devitt claims that I give no argument against the acceptability of rule-circular justification. But the passage that follows the phrase that he quotes was intended as such an argument and still seems to me to constitute a compelling one.

Second, Devitt tentatively suggests that a view of the sort he is defending might avoid rule-circularity by justifying each rule in terms of others. But if I understand what he is suggesting, such a justification would still be circular in the objectionable way: the justifications of at least some of the rules would ultimately depend, via a sequence of rules, on themselves.

Third, Devitt fails to understand the point about the atomistic character of a priori justification. As the rationalist conceives it, each individual instance of a priori justification depends only on the specific insight that is relevant to it, so that there is simply no need (and no use) for a general rule “yielding a priori insights.”<sup>1</sup>

Fourth, Devitt asks how finding something to be intuitively necessary can constitute a reason for thinking that it is true. If the insight is genuine, then the answer is obvious. There is (obviously) no non-circular way to establish that such insights are genuine, but there is equally no cogent way to argue that they are not or that we could not have such a capacity which does not tacitly appeal to such insights.<sup>2</sup> To reject all such insights is to reject the capacity of human intelligence to have good reasons for believing anything beyond the narrow deliverances of direct experience. Appeal to “rules” into whose truth-conduciveness one has no such insight does nothing to address this issue. And to simply insist, as Devitt does, that any reason for thinking that any non-tautological claim about the world is true must be empirical is to back oneself into a corner from which there is no escape: to repeat, most of the claims that we think we have reasons to accept are not matters of direct experience, and experience alone cannot establish that they are connected to experience in a way that makes them likely to be true.

Fifth, Devitt denies that a person making an inference must have an insight into its correctness for his conclusion to be justified, claiming that all that is required is that the inference “be good.” Perhaps there is some sense of the multifarious term “justification” for which this is correct. But a person who lacks such an insight has no *reason* for thinking that the resulting conclusion is true, and a person who infers in this way generally has no reason for thinking that any of his conclusions are true. (It seems to me obvious that there is a kind of insight into the cogency of such an inference that underlies and justifies the propositional claim that it is cogent: one sees how and why the conclusion follows, not simply that it follows. But I did not intend that this was a form of “knowing how” as that notion has ordinarily been understood.)

## Notes

1 For more discussion, see Bonjour (1998, pp. 142–147).

2 For some elaboration of this point, see Bonjour (1998, pp. 153–156).



## Reference

BonJour, L. (1998) *In Defense of Pure Reason: A Rationalist Account of a Priori Justification*. Cambridge: Cambridge University Press.

## Further Reading

Boghossian, P. and Peacocke, C., eds (2000) *New Essays on the a Priori*. Oxford: Clarendon Press.  
Butchvarov, P. (1970) *The Concept of Knowledge, Part I*. Evanston, IL: Northwestern University Press.

DePaul, M.R. and Ramsey, W., eds (1998) *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*. Lanham, MD: Rowman & Littlefield.

Hanson, P. and Hunter, B., eds (1992) *Return of the a Priori*. Calgary, Alberta: University of Calgary Press.

Kitcher, P. (1980) A priori knowledge. *Philosophical Review* 76: 3–23.

Kornblith, H., ed. (1994) *Naturalizing Epistemology*, 2nd edn. Cambridge MA: MIT Press.

Kornblith, H. (2000) The impurity of reason. *Pacific Philosophical Quarterly* 81: 67–89.

Moser, P., ed. (1987) *A Priori Knowledge*. Oxford: Oxford University Press.

Pap, A. (1958) *Semantics and Necessary Truth*. New Haven, CT: Yale University Press.

Sleigh, R.C., ed. (1972) *Necessary Truth*. Englewood Cliffs, NJ: Prentice Hall.

Sumner, L.W. and Woods, J., eds (1969) *Necessary Truth*. New York: Random House.

## Chapter Nine

# Is There Immediate Justification?

You have *immediate* justification for believing that *p* if, and only if, your justification for believing that *p* does not depend on any justification you have for believing some other proposition, *q*. James Pryor defends the claim that such justification exists. According to him, the best argument in its support is not the regress argument, according to which unacceptable alternatives – justification coming from an infinite regress of reasons, from a circular chain of reasons, or from initial beliefs that are not themselves justified – lend support to the conclusion that there must be immediate justification. Rather, Pryor argues, the best way to defend the existence of immediately justified beliefs is to cite examples of such beliefs, such as “I am tired” or “I have a headache.” Pryor provides additional support for immediate justification by criticizing what he calls the “Master Argument for Coherentism.” In his essay, Juan Comesaña raises a serious problem for the view that there are immediately justified beliefs. He argues that a view like Pryor’s is committed to four principles each of which is, by itself, exceedingly plausible. Unfortunately, the four principles are inconsistent. In response to Comesaña, Pryor singles out and rejects what Comesaña calls the Entailment Principle. Comesaña, in turn, defends the principle in his reply to Pryor.

## There Is Immediate Justification

*James Pryor*

### 1 Justification

I want to talk about a certain epistemic quality that I call “justification,” and inquire whether that quality can ever be had “immediately” or “non-inferentially.” Before we

get into substantive issues, we need first to agree about what epistemic quality it is we'll be talking about, and then we need to clarify what it is to have that quality immediately or non-inferentially.

When I say I call this epistemic quality "justification," you are liable to think, "Oh I know what that is." You may. But experience has taught me that different philosophers use and understand the term "justification" differently, even *before* they start spinning substantive theories about what "justification" amounts to. So we should proceed cautiously. You *may* use the term "justification" to describe the same epistemic quality as I do; or you may use it to describe some different status or quality. You may use some *other* term, or no term at all, to describe the quality I call "justification."

I say that *you have justification to believe P* iff you are in a position where it would be epistemically appropriate for you to believe P, a position where P is epistemically likely for you to be true. I intend this to be a very inclusive epistemic status.<sup>1</sup> Some philosophers say you can know P without "having any justification" for your belief. We can assume that whenever a subject knows P, she will be in a position where it would be epistemically appropriate to believe P. So on my usage, whoever knows P has justification to believe P. (Perhaps she has that justification *because* she knows.) The philosophers who say otherwise are using "having justification" to mean something different, or more specific, than the epistemic status I am using it to mean. The same goes for philosophers who say a belief can be epistemically appropriate, and so play a role in justifying other beliefs, though you do not "have any justification" for it. On my usage, *all it means* to have justification to believe something is that it is appropriate for you to believe it, and to rely on that belief in forming other beliefs. Some philosophers call this epistemic status "entitlement" or "warrant," rather than "justification." For the sake of a shared discussion, though, we need to fix on a single terminology.

If there is some state or condition you are in *in virtue of which* you have justification to believe P, I'll call it a "justification-making condition," or a *justification maker* for short. This is a condition that *makes it* epistemically appropriate (or more appropriate) for you to believe P, rather than disbelieve P or suspend judgment. It is a truth maker for your having justification to believe P. (Firth, 1964, called these conditions "warrant-increasing properties.") We can say that conditions of this sort *justify* you in believing P. They are *justifiers*. (We will encounter a different way to understand talk of "justifiers" in section 6 below.)

In what follows, it will be useful for us to distinguish between *having justification to believe P*, and actually *appropriately holding* a belief in P. To have justification to believe P, it is not important whether you *actually do* believe P (nor, if you do, why you do); there just have to be things that make believing P an appropriate attitude for you *to have*. To *appropriately believe P* more is required. You need to believe P; you need to have justification *to* believe P; and you also need to believe P *on the right grounds*. You need to believe it *for* reasons that make you have justification to believe it; you can't believe it for other, bad reasons, or on a whim.<sup>2</sup> There are further conditions as well: for instance, you need to be taking proper account of any evidence you have that tells against or undercuts your grounds for believing P. I describe another further condition in Pryor (2004). Only when all such conditions are met will your belief in P be appropriately held.

## 2 Immediate Justification

Now that we have a grip on the notion of “justification,” let’s clarify what it means to talk about “immediate justification.”

For some propositions, you have justification to believe them because *other* propositions you have justification to believe epistemically support them. For instance, suppose you look at the gas gauge of your car, and it appears to read “E.” So you have justification to believe:

(Gauge) The gas gauge reads “E.”

That, together with other things you justifiedly believe about your car, gives you justification to believe:

(Gas) Your car is out of gas.

(It is not important for our purposes whether you *actually do* believe (Gauge) or (Gas). Given your evidence, you *ought* to believe them.) In this example, your justification to believe (Gas) *comes* in part from the fact that you have justification to believe (Gauge). That is, having justification to believe the latter is part of *what makes you* have justification to believe the former. The justification you have in this example to believe (Gauge) does *not* in the same way come from your having justification to believe (Gas). (One mark of this is that evidence that undercut your justification to believe (Gauge) would ipso facto undercut your justification to believe (Gas); but not vice versa.) When your justification to believe P comes in part from your having justification to believe other, supporting propositions, I will say that those latter propositions *mediate* your justification to believe P. (This kind of justification is sometimes called “inferential” justification. We will encounter a second way in which justification can be “inferential” later.) When your justification to believe P does *not* come from your justification to believe other propositions, I will call it *immediate*.

Some clarifications.<sup>3</sup> First, the question whether your justification to believe P is mediate or immediate is a question about *what kind of epistemic support* you have for P. It is not a question about *how much* support you have: nothing in our definition requires immediately justified beliefs to be infallible or indefeasible. Nor is it a question about what psychological processes you have undergone. The support you have to believe P can be mediate (or “inferential”) even if you didn’t *arrive at* P by deriving or inferring it from other beliefs.

Second, in order for you to have immediate justification to believe P, it is not required that your justification *comes from nowhere*, that there is nothing that *makes you* so justified. It is only required that what makes you justified doesn’t include having justification for other beliefs.<sup>4</sup> There are various proposals about what can make one have immediate justification. For example, perhaps you are immediately justified in believing you feel tired because *you do* feel tired. Perhaps you are immediately justified in believing that tiredness is a mental state because *you understand* what tiredness is. And so on. It may be

that there is no single correct account. Different propositions may be justified by different kinds of things.

Third, the fact that you have immediate justification to believe P does not entail that no other beliefs are required for you to be able *to form or entertain* the belief that P. Having the concepts involved in the belief that P may require believing certain other propositions; it does not follow that any justification you have to believe P must be mediated by those other propositions.<sup>5</sup>

Fourth, justification is usually defeasible. What a justification maker for P gives you is provisional or *prima facie* justification to believe P; and *that* is what I am saying can be mediate or immediate. Whether it is *all things considered* appropriate for you to believe P will depend on what *other* evidence you possess, and whether it *defeats* the *prima facie* justification you have to believe P.

Fifth, beliefs can be epistemically overdetermined. You can have immediate justification and independent mediate justification to believe the same thing. In some cases, your belief will be grounded on the facts that make you have immediate justification; in other cases it might be grounded on the facts that make you have mediate justification, or on both sets of facts. This shows that we should try to explain the notion of “grounding” in a way that permits beliefs to be both grounded and immediately justified – if that is possible. (In section 7 we will consider an argument that it is not possible.)

### 3 Why Believe in Immediate Justification?

Now that we have achieved some clarity about what immediate justification *is*, let’s ask why we should believe in it.

The most famous argument for immediate justification is called the *Regress Argument*. Really this is not one argument but several; because philosophers do not always have the same regress in mind. Sometimes they have in mind a *dialectical* regress: to defend your belief that P in argument, you need to appeal to other beliefs, but then how do you defend those other beliefs? Sometimes they have in mind a *grounding* regress: your belief in P is grounded in such-and-such other beliefs, but then what grounds those other beliefs? Sometimes they have in mind a *justification-making* regress: what makes you have justification to believe P is, in part, your having justification to believe such-and-such other propositions, but then what makes you justified in believing those other propositions?

Let’s focus on this third, justification-making regress. There are four possible ways for the regress to play out:

- (i) The regress never ends. The justificatory chain goes on forever.
- (ii) What makes you justified in believing P is your having justification to believe other things, and ... what makes you justified in believing some of them is your having justification to believe P. That is, the justificatory chain includes some closed loops.
- (iii) Eventually we get to a proposition you believe inappropriately, *without* having any justification for it. Though this belief is not itself justified, it is somehow able to justify you in believing further propositions.

- (iv) Eventually we get to a proposition you have justification to believe, but that justification does not come from your believing, or having justification to believe, any further propositions.<sup>6</sup>

The foundationalist argues that options (i) and (ii) are untenable; so we have to accept (iii) or (iv). On either of those options, a subject can have justification to believe some propositions, that does not come from her having justification to believe any other propositions.<sup>7</sup>

Though this Regress Argument is the *most famous* argument for immediate justification, I do not think it is the *best* argument. It has the same weakness as any argument by elimination: everything turns on whether the rejected options really are untenable. That is not a matter that can be quickly decided. In addition, the Regress Argument assumes that justificatory relations always have a linear, asymmetric nature; and some epistemologists deny that that is so.

So I do not think the best argument for immediate justification is this Regress Argument. I think the best argument comes from *considering examples*.

Suppose I feel tired, or have a headache. I am justified in believing I feel those ways. And there do not *seem* to be any other propositions that mediate my justification for believing it. What would the other propositions be?

Suppose I raise my arm. I am justified in believing that I'm doing this in order to scare a fly. That is *my reason for* trying to raise my arm. Sometimes my reasons for acting are opaque to me and have to be carefully reconstructed. But not always. In cases like this one, my reasons can be immediately evident to me. There doesn't seem to be *anything else* I am justified in believing, that *makes me* justified in believing my reason for trying to raise my arm is to scare a fly. What would the other beliefs be?

I am imagining my grandmother. The way I am imagining her is sitting in her kitchen. Or at least, I believe it is. And it seems I could be *justified* in that belief. Again, it is hard to see what other propositions might mediate this justification.

I think about a domino and a chessboard. It is obvious to me that the only way to wholly cover two spaces on the board is to place the domino horizontally or vertically. That is something I *could* have derived from geometric premises. But in this case I didn't. I just immediately saw that it was true. In this case, too, my justification does not seem to be mediated by any further propositions.<sup>8</sup>

These and many other examples provide us with good candidates to be immediate justification. What we need to do is see whether such examples stand up to critical reflection.

## 4 The Master Argument for Coherentism

The main argument against immediate justification was historically directed at the Given Theory. That was a theory that offered a specific account of what some immediate justification makers looked like. The precise details of the Given Theory aren't important for our inquiry.<sup>9</sup> What is important is that the Given Theory is just one possible account among many of what gives us immediate justification. Fans of immediate justification are also free to give different accounts. The following map may be helpful:

I make foundationalism a proper subclass of the view that there is immediate justification, because foundationalists also hold additional theses about the structure of your justification. One does not need to accept those additional theses, merely to believe that *some* beliefs are immediately justified. Coherentists deny that it is possible for any beliefs to be immediately justified. They say that justification always comes, *at least in part*, from your justification for other beliefs. I distinguish between pure and impure versions of coherentism. Pure coherentists claim that a belief can only be justified by its relations to other beliefs. Impure coherentists are willing to give some non-beliefs, such as perceptual experiences, a justifying role. They will just deny that those states are able to justify a belief *all by themselves*. They can only do so in cooperation with other justified beliefs. For instance, an impure coherentist might say that when it looks to you as if you have hands, *and* you have justification to believe that your visual experiences are reliable, *those facts together* can make you justified in believing that you have hands.<sup>10</sup>

Now, as I said, the main argument against immediate justification was historically directed at the Given Theory. This argument alleged that in order to be a justifier, you need to have certain characteristics, and that having those characteristics makes you be the sort of thing that itself needs justification. Here is a sample presentation, from BonJour:

The basic idea of givenness ... is to distinguish two aspects of ordinary cognitive states, their capacity to justify other cognitive states and their own need for justification, and then to try to find a kind of state which possesses only the former aspect and not the latter – a state of immediate apprehension or intuition. But we can now see plainly that any such attempt is fundamentally misguided and intrinsically hopeless. For it is clear on reflection that it is one and the same feature of a cognitive state, namely, its assertive or at least representational content, which both enables it to confer justification on other states and also creates the need for it to be itself justified – thus making it impossible in principle to separate these two aspects. (BonJour, 1985, p. 78)

The characteristics BonJour cites are having “assertive” or representational content. (Sometimes it is claimed, in addition, that a justifier has to have *conceptual* content. We will return to that idea later.) In order to have these characteristics, the coherentist argues, a state would itself have to be a belief – or at least be sufficiently like a belief that it in turn needs justifying. So we aren’t going to find any immediate justifiers: conditions that are able to justify though they don’t themselves need justifying. That is the core of the familiar argument against the Given Theory.

Notice that this argument really doesn’t have anything specifically to do with the Given Theory. If it works, it should work against *any* account of immediate justification. In fact, if it works, then the only things which can play *any* justifying role will be other beliefs (or belief-like states). So the argument threatens impure coherence theories no less than it does views that countenance immediate justification. It is not really an argument *against the Given Theory*, then. It is more an argument *for pure coherentism*. I think of this argument as the *Master Argument for coherentism*.<sup>11</sup>

One step in this Master Argument insists that only states with propositional content can be justifiers. Let’s call this:

## The Content Requirement

In order to be a justifier, you need to have propositional content.

Why should we accept that requirement? Well, if a state doesn't have propositional content, then it can't stand in logical relations to beliefs. Davidson once complained:

The relation between a sensation and a belief cannot be logical, since sensations are not beliefs or other propositional attitudes. What then is the relation? The answer is, I think, obvious: the relation is causal. Sensations cause some beliefs and in *this* sense are the basis or ground of those beliefs. But a causal explanation of a belief does not show how or why the belief is justified. (Davidson, 1986, p. 311)

Of course Davidson is right that merely learning that some sensation S caused belief B does not *show* that B is justified, or that S is what makes it justified. However, it would be compatible with that that S *does* make B justified. Davidson wants to rule that possibility out. He thinks that sensation S *cannot* be what justifies B. So he must think that it is *only* by standing in logical relations to a belief that a state can justify the belief. He may be thinking: if a state doesn't stand in logical relations to a belief, then why should it justify that belief *as opposed to others*? For example, let's assume that headaches don't have propositional content. (Some philosophers argue that all mental states have propositional content. For the sake of discussion, we will assume with Davidson that they are wrong.) Why then should a headache justify me in believing *I have a headache*, as opposed to *I don't have a headache* or *I am a headache* or *There are no such things as headaches*? *My headache* itself wouldn't logically support *any* of those propositions; so it is not clear why it should justify some of them but not the others.

There is then some initial plausibility to the idea that in order to play the role of a justifier, a state has to be able to stand in logical relations; which it can only do if it has propositional content. (Objection: What about *a lack of defeating evidence*? That is not naturally thought of as a state with propositional content; but it does seem relevant to your justification. Reply: The Content Requirement should be understood as stating a necessary condition to be a *prima facie* justifier. The role defeating evidence plays is in settling a further question: When does *prima facie* justification get to become all things considered justification?)

The other step in the Master Argument insists that if a state has propositional content, then it will be a belief or epistemically like a belief:

### Only Beliefs

Only beliefs (or other states that are epistemically like beliefs, and also require epistemic justification) have propositional content.

But wait a minute! *Desires* have propositional content, and they are not the sort of thing which *could* be, nor do they *need to be* justified – at least, not *epistemically* justified.<sup>12</sup>

The coherentist will respond: "That's true, desires have propositional content and don't need any epistemic justification. So the Only Beliefs premise as it stands is false. But desires aren't *capable* of justifying beliefs, either! So they're not a counterexample to *the conclusion* we want: that only beliefs can be justifiers. They just force us to be more specific about what features it is that enable a state to be a justifier. It takes more



than just having propositional content. There are some ways of representing the proposition that P that *purport to be saying how the world is*, and other ways that don't. When a state represents that P in the first way, we can say that the state *assertively* represents that P. Desires may represent that P in some sense, but they do not represent it assertively. Neither do states like imagining and entertaining. When you desire that P, or imagine that P, or entertain the thought that P, your mental state does not purport to be saying that P is (already) true. What our Master Argument *should* say is that, to be a justifier, a state needs to have propositional content, *and* it needs to represent that proposition *assertively*. States that don't do that aren't even *purporting* to say how the world is, so how could they play the role of justifiers? And in order to represent a proposition assertively, a state will have to be a belief, or else sufficiently like a belief that it needs justifying too."

Let's revise the Master Argument as the coherentist proposes:

#### The Content Requirement (Revised)

In order to be a justifier, you need to have propositional content, and you need to represent that proposition assertively.

#### Only Beliefs (Revised)

Only beliefs (or other states that are epistemically like beliefs) represent propositions assertively.

∴ Only beliefs (or other states that are epistemically like beliefs) can be justifiers.

I want to mention two quick worries about this argument, and then dwell at length upon a third.

First, even if this argument were sound, it is not clear that it would establish coherentism. The argument says that the only things that could justify are states like belief, that "require" justification. A foundationalist might agree that beliefs in some sense always "require" justification, but argue that they are still sometimes *able to justify* other beliefs even when *they are not* themselves justified. (This was option (iii) in the Regress Argument.) The other beliefs that got justified in this way would count as "immediately justified," as I have defined it.

Second, we might worry whether the coherentist is himself in a position to accept the Master Argument. After all, doesn't the coherentist want facts about *coherence* to play a justifying role? Yet coherence is not itself a belief or a belief-like state. Here I think the coherentist can reply, "Notice that coherence is a property of *the contents* of your beliefs. Any set of beliefs having the same contents would be just as coherent. So it is OK to say that it is always your beliefs that are doing the justifying. It is just that certain sets of beliefs (those whose contents cohere well) justify more than others. Talk about the justifying role of *coherence* is shorthand for talk about which sets of beliefs justify and which don't." This seems to me a plausible line for the coherentist to take.

There is a third worry that the coherentist cannot so easily finesse, however. This worry once again concerns the Only Beliefs premise. The problem is that many philosophers of mind these days think of *experiences* as having propositional content. To say that experiences have propositional content is not to say that experiences are beliefs. It can look to you as if P without your believing that P. Experiences and beliefs just have it in common that they both represent propositions.<sup>13</sup> And both seem to represent

propositions *assertively*; when they represent that P, they do so in a way that purports to say how the world is. So the Master Argument as we have it gives us no reason to exclude experiences from the ranks of justifiers.<sup>14</sup>

Yet, unlike beliefs, experiences aren't the sort of thing which *could be*, nor do they *need to be* justified. Sure, *beliefs about* what experiences you have may need to be justified. But *the experiences themselves* do not. (If someone comes up to you and demands, "*How dare you* have that experience? What gives you the right?" what should you say?) So we see that, contrary to the Only Beliefs premise, states that assertively represent a proposition *won't* always themselves require justification.

Where does this leave things between foundationalists (or fans of immediate justification more generally) and proponents of the Master Argument? A foundationalist *can* just stop here. He can say, "Well, we've seen that experiences are a counterexample to the Only Beliefs premise; so even if we accept the Content Requirement, that poses no obstacle to letting experiences play the role of justifiers. They might play that role in cooperation with other beliefs, as the impure coherentist allows, or they might do it all by themselves. We have seen no good reason yet to think they can't" (see, for example, Martin, 1993; Steup, 2001).

Some foundationalists do stop there. They are happy to accept the Content Requirement. But that is a pretty demanding constraint to put on what can be a justifier. And in fact, if you think about it, the Content Requirement will be well motivated only if an *even more demanding* constraint is. Let me explain.

The coherentist denies that the mere presence of a headache or a desire can justify you in believing anything, because these states don't assertively represent anything. They can't even justify you in believing that you have those states. But then neither should the mere presence of a belief or an experience justify you in believing you have *them*, either. For although beliefs and experiences do have propositional contents that they assertively represent, those contents are playing no role in the justificatory relation here envisaged. It seems just an unfair prejudice to allow states that assertively represent propositions to justify the belief that *you are in* those states, but deny the same ability to other states. We should either allow this justifying relation in every case, or prohibit it in every case.

The coherentist will try to prohibit it. They will say, "Look it's not just that the justifying state needs to *have* a content. The content needs to be in some sense *what does the justifying.*" What that means, I guess, is that the content of the justifying state needs to *imply*, or *inductively* or *abductively support*, or stand in some other suitable evidential relation to the content it justifies. Let's give this idea a name:

#### Premise Principle

The only things that can justify a belief that P are other states that assertively represent propositions, and those propositions have to be ones that *could be used as premises* in an argument for P. They have to stand in some kind of inferential relation to P: they have to imply it or inductively support it or something like that.

*The contents* of your beliefs and experiences will ordinarily not imply that *you are in* those states; so this principle will prevent your beliefs and experiences from directly justifying the belief that you are in them.<sup>15</sup>

I think this Premise Principle is the real intuitive force behind the coherentist's Content Requirement. Recall the Davidson quote from earlier. He was saying that to be

a justifier, you have to stand in logical relations to the beliefs you justify. I think this is what he really had in mind. Here are some other authors also giving voice to the Premise Principle.

We cannot really understand the relations in virtue of which a judgment is warranted except as relations within the space of concepts: relations such as implication or probabilification, which hold between potential exercises of conceptual capacities (McDowell, 1994, p. 7).

A reason for S's believing that P is a fact about that person which makes her believing that thing intelligible from the point of view of rationality. If this is to happen then the selected fact about S must be somehow related to (her) believing that P. And since this relation is to make her believing that P intelligible *from the point of view of rationality*, it is necessarily a relation which obtains in virtue of the correctness of some kind of reasoning. That is to say, successfully giving such a reason makes essential reference to the premise of an inference of some kind, whose conclusion is appropriately related, most likely by identity, to the content of the belief for which the reason is being given. (Brewer, 1999, p. 154; see also pp. 150–151)

The Premise Principle says that all justifying relations between states hold in virtue of “inferential relations” between their contents. We have to be sure we understand this properly.

First, the Premise Principle states a constraint on what can give you justification *to believe* things; it does not concern your actual beliefs or reasoning processes. It can allow that we often form beliefs that are *supported by* “inferential relations” without *engaging in* any inference.

It can also respect the difference that Harman (1986) emphasized between logic and reasoning. Harman said that just because you believe some premises that together imply P, it doesn't follow that it would be reasonable for you to infer P. It might, for example, be more reasonable to give up your belief in some of the premises. The Premise Principle can allow this. It is only trying to explicate the nature of *prima facie* justificatory relations. According to the Premise Principle, those always consist in “inferential relations” between contents. If your beliefs in some premises together imply P, and you have *prima facie* justification for those beliefs, then you have *prima facie* justification to believe P. But it might be unreasonable for you to *infer* P on those grounds, for example, if you also have *other* evidence that tells against P, or that undercuts some of your *prima facie* justification for the premises.

Although there is *a sense* in which the Premise Principle claims all *prima facie* justificatory relations are “inferential,” this is not the same as saying that all justification is “inferential” in the sense of being *mediated*. The coherentists would like to *argue from* the Premise Principle to that conclusion. But the Premise Principle by itself doesn't say it. Without supplementation, the Premise Principle would allow experiences to justify beliefs. For instance, an experience as of your having hands could justify the belief that you have hands. And justification of this sort would count as immediate. It is just that, according to the Premise Principle, the experience is able to justify that belief because of the “inferential relations” its content stands in to the content of the belief. (In this case, the “inferential relation” is straightforward: the experience's content is the same as the content of the belief.) So the Premise Principle

doesn't imply what people usually have in mind when they say that our perceptual justification is "inferential" or mediated. That is a view according to which we are in the first place justified in believing we have certain experiences, and then that justification for beliefs about our experiences is an essential part of what justifies our beliefs about the external world. The Premise Principle doesn't imply that; it would allow *our merely having* experiences with the right sorts of contents to justify beliefs about the external world.

Nor is the Premise Principle *implied by* the view that perceptual justification is mediated. For the latter view says nothing about what justifies our beliefs about our experiences. Perhaps, *contra* the Premise Principle, we are justified in those beliefs merely *by virtue of having* the experiences.

So there is no straightforward relation between the Premise Principle and the question whether perceptual justification is mediated or immediate. The relation between the Premise Principle and foundationalism more generally is also complicated. As I said before, some Foundationalists are happy to accept the coherentist's Content Requirement, and the Premise Principle that motivates it. They just argue that there can be justifiers that satisfy the Premise Principle but aren't beliefs.

Those arguments interact in interesting ways with the question whether experiences have "conceptual content." Some philosophers combine the Premise Principle with the view that you can only have the required type of "inferential relation" when both *relata* have conceptual content. That seems to be the view of Sellars (1963), McDowell (1994, lectures 1 and 3; 1998),<sup>16</sup> and Brewer (1999, chapter 5). The coherentists will go on to argue that experiences don't have conceptual content and so cannot be justifiers. McDowell and Brewer, on the other hand, think that experiences *are* justifiers; so they argue that experiences do have conceptual content after all. Others have argued that experiences can stand in the "inferential relations" the Premise Principle requires even if their content is not conceptual (see e.g., Heck, 2000; Peacocke, 2001).<sup>17</sup> Personally, I am not really sure what "conceptual content" is; so I won't enter into this debate. I just wanted to call attention to the role the Premise Principle plays in it.

We have been looking at the Master Argument for coherentism, and considering whether it succeeds in ruling out the possibility of immediate justification. I said that even if the Master Argument were sound, it might still be possible for unjustified beliefs to do some justifying; and that would be a kind of immediate justification. We have also seen that one can accept the Premise Principle and still say that experiences justify; that will be another kind of immediate justification. So one does not need to reject the Premise Principle, to believe in immediate justification.

Nonetheless, many foundationalists will want to reject the Premise Principle. It doesn't exclude *the very possibility* of immediate justification; but it does impose *quite a demanding constraint* on what can be an immediate justifier. Many foundationalists believe in justification-making facts that violate that constraint. For instance, many foundationalists want to allow *facts about what sensations you are having*, or *facts about what mental activities you are engaging in*, to count as justifiers. Some want to allow *facts about what is required to possess certain concepts* to play a justifying role. Some say *facts about how reliable you are*, or *facts about whether your cognitive faculties are functioning properly*, or *facts about what beliefs are irresistible*, can play a justifying role. And so on. Each of these facts concerns matters that go beyond what

assertive states you are in; so the Premise Principle would exclude them all from being justifiers.

I think, then, that it would be valuable to consider whether the Premise Principle is really well motivated. The rest of this essay will consider some arguments on its behalf.

There is one type of argument that I won't consider. Those are "arguments from the trenches": arguments of the form "Theory So-and-So gives the correct substantive account of justification; and that theory only postulates justifiers of the sort the Premise Principle permits." Assessing any argument of that type would require examining the pros and cons of different theories of justification, and determining whether Theory So-and-So really is an adequate theory. That is well beyond the scope of this essay. Instead, I will look at arguments that try to establish the Premise Principle "from on high," *before* we have decided upon a substantive theory of justification. I am going to argue that *no argument of that sort* succeeds. Hence, I think we have no reason to give the Premise Principle any authority when we are choosing among competing theories of justification. If I am right, that will clear the way for the many foundationalist theories that postulate immediate justifiers that violate the Premise Principle.<sup>18</sup>

## 5 Avoiding Arbitrariness

Our first argument for the Premise Principle is one that I have already mentioned. Consider a state without propositional content, like your headache. Since it has no propositional content, this state can't stand in logical relations to any beliefs. So why should it justify *any one* belief as opposed to others? Why should it justify the belief *I have a headache*, as opposed to *I don't have a headache*, or any other belief? What can the foundationalist say to make the justifying relations he postulates non-arbitrary?

Chisholm provides a nice example of what is being objected to here. His view was that having an experience would give one *prima facie* justification for (or "tend to make evident") certain beliefs (see e.g., Chisholm, 1989). The justifying relations Chisholm postulated struck many philosophers as *ad hoc*. That impression was particularly forceful because of the model of experience Chisholm worked with. When we describe an experience as one of sensing or being-appeared-to *squarely*, that is supposed to be a description of the experience's intrinsic phenomenal quality. We might naturally *take* such an experience to be one in which we are perceiving an external square; but "being-appeared-to squarely" isn't a description of what external objects we seem to be perceiving. It doesn't mean "having an experience *that represents that there is a square*." Chisholm didn't think experiences had representational content. Hence, the justifying relations he postulated seemed to lack any principled motivation. Why should sensing squarely justify beliefs about squares, rather than beliefs about squirrels? Many think Chisholm had no adequate answer to this question; so his position seems arbitrary and unsatisfying.

I certainly agree that epistemologists should give principled, non-arbitrary rationales for the justifying relations they postulate. However, I see no reason to think that they will *have to* appeal to propositional contents to do it. A foundationalist might attribute other kinds of structure to some of his justifiers. On certain theories of events, for example, events have something like a logical structure. The event of my

having a headache has a logical structure akin to the structure of the proposition that I have a headache. Wouldn't these structures be enough to enable the foundationalist to avoid the charge of arbitrariness?<sup>19</sup> This is just one option for a foundationalist to pursue. It is hard to draw any general assessment, until we see how the details work out. But the idea that only the Premise Principle can save us from arbitrariness seems unwarranted to me.

## 6 Evidence and Reasons

Where I have been talking about "justification makers" or "justifiers," some other philosophers will talk about "evidence" or "reasons"; and there are several arguments that the latter notions have to conform to the Premise Principle. This might be thought to show that justification makers *also* have to conform to the Premise Principle.

Recall that a "justification maker" is defined to be whatever makes it epistemic appropriate for you to believe some propositions rather than others. We should be open to the possibility that terms like "evidence" and "reasons" do not express exactly that notion. They may express different, or more specific, notions.

For instance, some philosophers argue that "reasons" and "evidence" have to be the sort of thing that can *probabilify* a hypothesis, and hence, that the hypothesis can have a *probability conditional on*. Hypotheses also have to be able to *explain* our evidence, be *inconsistent with* our evidence, and so on. All of these roles require evidence to have propositional content. So how could states without propositional content justify or be evidence? (See Williamson, 2000, pp. 194–197; Plantinga, 2001, p. 62.)

In response, I say: let's not assume too quickly that "evidence" and "justifier" are perfect synonyms. Consider that we sometimes use the terms "belief" and "desire" to refer to *propositions that* one believes and desires, rather than to *one's states of* believing or desiring them. Similarly, I think, sometimes we use "evidence" to refer to *propositions that are* evident to one, rather than to the states that make them evident. In other words, we use "evidence" to refer not to our justification makers, but rather to the propositions that they (most directly) justify. We call those propositions "our evidence" because they can serve as evidence for further reasoning. This diagnosis would permit things like headaches, that do not *themselves* have propositional content, to be justification makers – so long as *what* they give one justification *to* believe *is* a proposition.

A second argument says that we ordinarily understand "justifications" for a belief to be *arguments* that support the belief. If you have *reasons* for your belief, they should be considerations you could in principle *cite*, or *give*, to someone who doubted or challenged the belief. You can't give someone else a non-propositional state like a headache (at least, not in the relevant sense); you can only give them *premises* and *arguments* that inferentially support your belief. This seems to show that justifications and reasons *are* limited to things permitted by the Premise Principle. (See, for example, McDowell, 1994, pp. 165–166.)

There may be a notion of "a reason" that these remarks properly articulate. We can call it the *dialectical* notion of a reason. I want to emphasize, though, that that notion is *different* from the notion of a justification maker that I have been employing in this paper, and that the Premise Principle is meant to be formulated in terms of.

It is useful here to distinguish two construals of the verb “justify.” On the first construal, “justifying” a belief in P is a matter of proving or showing the belief to be just (or reasonable or credible). This is something that *a person* does, by giving some argument in support of that belief. (Here we can include both arguments whose conclusion is P, and arguments whose conclusion is that your belief in P is epistemically appropriate, or is likely to be true.) By extension, we can also talk about *things* justifying beliefs; in this extended sense, a thing counts as justifying a belief if it is something *you are in a position to use* to prove or show your belief to be just. Such things would be “reasons” in the dialectical sense articulated above. To be explicit, let’s call these things *justification showers*.

There is also a second way to construe the verb “justify,” which sees it as akin to the verbs “beautify” and “electrify.” When a combination of light and color beautifies a room, it is not *proving* that the room is beautiful; rather, it is *making* the room beautiful. Similarly, on this understanding, justifying a belief is a matter of *making* a belief just or reasonable, rather than a matter of *showing* the belief to be just. That is how I understand the notion of a justification maker.<sup>20</sup>

No doubt there are some interesting connections between justification making and justification showing. But they are two different notions; so we should not assume that their extensions will coincide. It *needs argument* to show that nothing is eligible to be a justification maker unless it can also be a justification shower. Until we have such an argument, the fact that justification showers always conform to the Premise Principle should not persuade us that justification makers must do so as well.

Even if we manage to separate the notion of a justification maker from the dialectical notion of “a reason,” I expect proponents of the Premise Principle will still insist there is enough of a connection between justification and “reasons” for constraints on the latter to support some restrictions on the former.

For instance, they can observe that there is a difference between *reasons there are* to believe P – where these include reasons not now available to you – and *reasons you have* to believe P. For example, one reason *there is* to believe you will soon be sick is the fact that you just drank poison. But if you are unaware of that fact, then it is not a reason you have. For something to be a reason you have, for it to justify *you* in believing P, it has to be in some sense *epistemically available* to you. It has to be the sort of thing you could *take as* a reason. When it is not available to you – for example, when you are not in a state that assertively represents it, and so not in a position to appeal to it in arguing for P – then it may be *a* reason to believe P, but it won’t be a reason you have. For anyone with “internalist” sympathies, these reflections should apply to justifiers just as much as they do to reasons.<sup>21</sup>

I think it is right to distinguish between things such that *you would* be justified in believing P, if you were aware of them, and things that *do* justify you in believing P. I think it is also right that if something justifies you, then it has to be in some sense “available” to you. But I think it would be wrong to assume that this kind of “availability” requires you to be in representational states. As I understand the notion of “availability,” it is correlative to the notion of a ground. A justifier is available to you at a given time – it will be something you can “take as” a reason – if it is something that could then *ground* a belief of yours. If the foundationalist can make sense of beliefs being *grounded on* non-representational justifiers like headaches, then he can make sense of those justifiers being sufficiently available to you.

Our next argument for the Premise Principle will question whether the foundationalist *can* make sense of the grounding relation, without appeal to beliefs or other representational states.

## 7 Grounding and Being Guided by Norms

We introduced the notion of a *ground* to distinguish between cases where you believe *P* for good reasons, or on grounds that justify you in believing *P*, and cases where you believe *P* on bad grounds, ones that do not justify that belief. What does it take for your belief to be *grounded on* some fact or condition *C* that you are in? A natural thought is that your belief counts as so grounded iff it is formed (or sustained) in a way that is guided by the epistemic norm “When in *C*, believe *P*.” If that is right, then the best way to understand the grounding relation is by inquiring into what it takes to be guided by such a norm.

I understand an *epistemic norm* to be a claim about how we should be, in epistemic matters. Some norms merely evaluate the quality of a static epistemic situation, for example, *You should not (it is inappropriate to) have inconsistent beliefs*. Others instruct us how to *change or improve* our situation, for example, *If you believe that A is F, then you should believe that B is F too*; or *You should gather as much evidence as possible*. Only some epistemic norms tell us what to believe or to refrain from believing. Those are the norms that we need to consider here. We can take them to be of the form: *When you are in conditions C, you should believe (or refrain from believing) P*. Putting it in the imperative: *When you are in C, believe P*. Norms like these will be correct just in case being in *C* does make it the case that you should believe *P*. In other words, just in case being in *C* is a justification maker for the proposition that *P*.

For any norm, there will be a difference between acting in a way that merely *happens to accord* with the norm, on the one hand, and *being guided by* the norm, or *complying* with it, on the other. You act in accordance with a norm “When in *C*, do  $\phi$ ” just in case you always  $\phi$  when in *C*. You need not be *trying* to follow that norm. You may have  $\phi$ ed for reasons that have nothing to do with *C*. You may even regard being in *C* as decisive reason *to refrain from*  $\phi$ ing – but just never have noticed that you were in *C*. In order to comply with the norm, on the other hand, the fact that you are in *C* *does* in some sense need to guide or be your reason for  $\phi$ ing. We need to know what this relation amounts to, when  $\phi$ ing consists in forming (or sustaining) a belief.

One account of this will portray you as *deliberately following* the norm, in the way that one can deliberately follow a cooking recipe. I mean three things by this. First, on this account your belief will be voluntarily chosen. Forming it will be a genuine action of yours. Second, your belief will be chosen for a reason (a practical reason). In forming the belief that *P*, you will have been guided by a norm “When in *C*, believe *P*” only if the fact (or apparent fact) that you are in *C* is among your reasons for actively forming that belief. Third, this account says that to be acting for the reason that you are in *C*, you have to *represent* that reason to yourself. You have to be in a position to employ the proposition that you are in *C* as a premise in your practical reasoning. The upshot of these three assumptions is that your belief can be guided by the norm “When in *C*, believe *P*” only when you represent to yourself that you are in *C*, and can employ that proposition as a premise in reasoning. This may be thought to lend some support to the Premise Principle.<sup>22</sup>



Some replies. First, even if this account of belief-formation were right, it is not clear that it would really support *the Premise Principle*. It seems only to support the claim that, for C to be a justifier that grounds your belief, you need to be in some state that assertively represents you as being in C. This doesn't imply that *C itself* is a representational state. Still, this account will imply that representational states are present whenever your belief is properly grounded. And that may be of some use to a defender of the Premise Principle. He may urge that it is really the state that *represents you* as being in C, and not the condition C itself, that is doing the justifying.<sup>23</sup>

Second, this account of belief-formation seems to rely on too reflective and deliberate a picture of what it takes to act for a reason, or in compliance with a norm. Consider activities like playing a musical instrument, figuring out why your car or computer isn't working, or making judgments of grammaticality. These are practical skills whose exercise seems to be governed by rules. But we don't think subjects need to think about or deliberately apply any rules when they are performing those activities. They don't even need to be *aware of why* they are acting in the precise ways they do. Many philosophers would regard their actions as guided by rules, for all that. So, *prima facie*, it seems possible to act in a way that is guided by rules without representing to yourself that you are in conditions C so now you should do so and so.<sup>24</sup>

Finally, I think the present account of belief-formation misrepresents *how active* we are with respect to our beliefs. Many justified beliefs aren't formed in the deliberate way it describes because forming those beliefs *isn't an action of ours* in the first place. We do exercise voluntary control over some aspects of our epistemic lives: what evidence we gather, what sources we consult, and so on. But when it comes to our here-and-now doxastic choices, these are usually involuntary and unreflective. Our beliefs usually just *result from* our other epistemic efforts. They just *happen*, in the way that sneezing or digesting happen. Ordinarily we make no intentional choices about what to believe. One *can* choose to believe something, and then seek ways *to get* oneself to believe it – just as we can seek ways *to get* ourselves to sneeze or digest. But that is not the way we usually form justified beliefs.<sup>25</sup>

Some philosophers argue for the strong thesis that it is impossible for justified beliefs *ever* to be formed by deliberate choice. Perhaps that is right. But here I need only the following, much weaker thesis: some beliefs are appropriately held, and so properly grounded, even though they *aren't* formed by deliberate choice. So it can't in general be required for grounding a belief that one have formed the belief in the deliberate way that the account we are considering describes.

Here a proponent of the Premise Principle might say: "True, we don't always *ourselves* deliberately choose our beliefs. But that is how an ideal reasoner would form beliefs. And it is a constraint on any condition that purports to justify our beliefs that it could ground the belief of an ideal reasoner who did choose his beliefs deliberately."

I am not sure that the fully reflective and deliberate reasoner envisaged here is a coherent ideal for us to aim at. Such a reasoner would never form or change beliefs, except by deliberately following an epistemic norm. But deliberately following a norm requires *already* having beliefs about (or at least, some representation of) whether its antecedent conditions are fulfilled. From where is the ideal reasoner supposed to get *those* beliefs? Since she is an ideal reasoner, she would have to have formed *them* by deliberately following norms, too. But then she would need further beliefs, about

whether the antecedent conditions of *those* norms are fulfilled.... There is a real threat that this reasoner would never be able to get started. She would never *be able to* deliberately follow a norm for forming beliefs. So she would be doxastically paralyzed. Or, if she were able to get started, it would only be by virtue of believing an infinite hierarchy of propositions. Neither option gives us a very promising model to aspire to.<sup>26</sup>

If your belief's being grounded on condition C isn't a matter of your *deliberately following* a norm when you form that belief, then what is it a matter of? This is a difficult question. I can't guarantee that when we work out the details, they won't turn out to be inhospitable to views that violate the Premise Principle. But I think we *can* be assured that there is no argument here for the Premise Principle "from on high," no argument that will settle the question before we work out those details.

## Acknowledgments

I would like to thank Tony Brueckner, Rich Feldman, Mark Greenberg, Ram Neta, Christian Piller, Mark Schroeder, Matthias Steup, and audiences at NYU, UNC-Chapel Hill, Princeton, Kansas, Oxford, and York, for conversations and feedback that helped me write this essay.

## Notes

- 1 Though it doesn't include everything. I think it is possible for subjects to believe P inappropriately, and so without justification, though be non-culpable for doing so. (Perhaps the epistemic faults that led to the belief are too subtle and well entrenched for those subjects to recognize.) So there is one kind of positive epistemic status, being *epistemically blameless* in believing P, that does not entail that one has justification to believe P. See Pryor (2001, section 4) for discussion and references.
- 2 The relation of believing something *for* such-and-such a reason is sometimes called "the basing relation." However, I think that terminology encourages too voluntaristic and reflective a picture of the phenomenon; I prefer to call it "the grounding relation" instead. We will talk more about this relation in section 7. For further discussion, see Korcz (1997) and Audi (1993, chapters 3, 7, and 8).
- 3 For more about immediate justification and what it does and does not require, see Alston (1989, chapter 3), Audi (1993, chapter 3), and Pryor (2000).
- 4 We have to be careful here: I expect that what makes you justified in believing one proposition may often make you justified in believing several. Suppose being in state S makes you justified in believing  $P_1, \dots, P_n$ . As I understand talk of "making it true" that you have justification, being in S can make it true that you have justification to believe  $P_2, \dots, P_n$  (let  $S^*$  be the state of having this justification), and also make it true that you have justification to believe  $P_1$ , without its following that being in  $S^*$  is part of what makes you justified in believing  $P_1$ . Hence, your justification to believe  $P_1$  can be immediate, even if what makes you justified in believing  $P_1$  also makes you justified in believing other propositions, too.
- 5 Consider: in order to have the concept of a unicorn I may need to believe (i) that unicorns have hooves, and (ii) that unicorns have horns. Now suppose I acquire evidence that a virus has killed all hoofed creatures. Since I believe unicorns to be hoofed creatures, I form the belief (iii) that no unicorns currently exist. It is clear that (ii) plays no role in justifying this belief. This shows that there can be propositions *you need to believe* in order to have certain

- concepts (you need to believe (ii) in order to have the concept of a unicorn), without those propositions *mediating your justification* for every belief involving the concepts. Now, (iii) is not an immediately justified belief. But it serves to make my point. We can see the same phenomenon with beliefs that *are* good candidates to be immediately justified, like (iv) *If any unicorn exists, it is identical with itself*. (ii) plays no more role in justifying that belief than it plays in justifying (iii).
- 6 On some presentations of the Regress Argument, the regress stops with your last justified belief (or, as I say, with the last proposition you have justification *to* believe). On others, the regress stops with what justifies your last justified belief: the states or conditions that *make you* justified in believing it. These are just different ways of talking; there is no real philosophical disagreement here.
  - 7 Many foundationalists shun option (iii), but structurally it also qualifies as a foundationalist option. For example, a foundationalist might claim that *merely having* a belief that P – even without justification – makes you justified in believing that *you do* have that belief.
  - 8 I owe this example to Tim Maudlin.
  - 9 Briefly, and I as I understand it, that theory invoked a mode of awareness that (i) had a success-grammar, (ii) was a cognitive relation to things or events, rather than to propositions, and (iii) wasn't mediated by your awareness of anything else. This mode of awareness went under various names: direct apprehension, acquaintance, etc. It was usually claimed that what we are so aware of are sense data. However, as I understand the Given Theory, it neither entails, nor is it entailed by, belief in sense data. The Given Theory is undergoing something of a revival these days: see Fales (1996) and DePaul (2001).
  - 10 Davidson and BonJour (when he was a coherentist) are pure coherentists; Cohen (2002) is a recent example of impure coherentism.
  - 11 The argument dates back to a debate between Schlick and Hempel. Schlick said we can sometimes “compare propositions to facts,” and thereby acquire justification for believing those propositions. Hempel argued that the only way to acquire justification for believing a proposition is to “compare” it to things that stand in logical relations to it, namely *other propositions*. See Schlick (1932/3), Hempel (1934/5a), Schlick (1934/5), and Hempel (1934/5b). Wittgenstein voices a view akin to Hempel's in *Philosophical Investigations*, section 486. The Master Argument has been given many times since (for example, Sellars, 1963, sections 3–7; Williams, 1977, chapter 2; BonJour, 1978, section 4, 1985, chapter 4; Davidson, 1986). (In more recent writings, BonJour rejects the Master Argument. See note 18, below.)
  - 12 Some might want to count *instrumental* desires as epistemically justified when the means–end beliefs that motivate them are justified. I wouldn't. But in any case, we can set instrumental desires aside and just consider cases where you desire that P for its own sake. Those desires have propositional content, and *they* don't need to be epistemically justified.
  - 13 One finds this view in Sellars (1963), Hintikka (1969), Dretske (1981, chapter 4 and 6), Evans (1982, chapters 5–7), Peacocke (1983, chapter 1), Searle (1983, chapter 2), Burge (1986), Lewis (1986), and in many places since. It is disputed whether experiences also have *additional* introspectible properties, beyond their propositional content; but there is broad agreement these days among philosophers of mind that they *at least* have propositional content.
  - 14 John Broome suggested that we might also count *intentions* as assertively representing propositions; propositions about how we will act in the future. If so, then the Master Argument would give us no reason to exclude intentions from the ranks of justifiers, either.
  - 15 They may be able to justify those beliefs *indirectly*, if their content is such that you can infer from its truth that you are likely to have certain beliefs or experiences. But the coherentist should have no objection to that.
  - 16 What McDowell (1994) calls “The Myth of the Given” is the view that things without conceptual content can play a justification-making role. Curiously, McDowell (1994, p. 144)

- seems to allow that *the fact that you are having a certain sensation* or impression might, together with justified beliefs about how such sensations are reliably caused, justify you in beliefs about the external world. McDowell thinks this kind of epistemic role is too “indirect” to be fully satisfactory; but he does seem prepared to count it as a justifying relation. This sits ill with his otherwise thoroughgoing commitment to the Premise Principle.
- 17 Two caveats about Peacocke. First, he would put his point like this: experiences can stand in the kinds of *rational relations* cited in the Premise Principle even if their content is non-conceptual. He is reluctant to call those relations “inferential relations” except when both *relata* are conceptual. Second, Peacocke does not himself accept the Premise Principle. He thinks that non-representational states like pains can also justify beliefs (see pp. 254–255).
  - 18 Although many foundationalists’ theories commit them to rejecting the Content Requirement and the Premise Principle, it is rare to find much explicit and sustained discussion of these principles. BonJour is one author who does discuss them. In BonJour (1978, 1985, chapter 4), he endorsed the Master Argument; but he has changed his mind and now rejects the Premise Principle. He thinks there can be “descriptive relations” between a belief and a non-propositional state that make the belief justified when one is in the non-propositional state. See BonJour (2001, pp. 29ff.). Other authors who discuss the Premise Principle are: Millar (1991, chapter 4), Reynolds (1991), and Fales (1996, chapters 5–6). Fales’s terminology can mislead, but in essence his view is this: experiences do not themselves have propositional contents, and so can’t stand in the kind of “inferential relations” required by the Premise Principle. However, their phenomenal qualities do have a proposition-like structure, and we have a way to non-propositionally “apprehend” this structure. Fales thinks that is all that is needed to justify our perceptual beliefs. Millar *does* allow that experiences have propositional contents, but he thinks experiences are individuated by their phenomenal types rather than by their contents. He also thinks it is these phenomenal types that are epistemologically important. He says there are “quasi-inferential” links between phenomenal types and beliefs that experiences of those types make appropriate. These links, not the experiences’ content, explain why experiences justify the beliefs they do.
  - 19 I am indebted to Mark Johnston for discussions of this possibility.
  - 20 The noun “justification” has both a count use and a mass use. The count use (“He has a justification for that belief”) is most naturally read as referring to arguments or justification showers. The mass use (“He has some justification to believe that”; “She has more justification than he does”; etc.) is more naturally read as referring to the presence of justification makers.
  - 21 I am indebted to Mark Schroeder for discussions of this objection. See also Unger (1975, chapter 5). BonJour (1985, chapters 2 and 3) and Brewer (1999, pp. 19, 49, and esp. 163ff.) claim that nothing can be a reason for you unless you are in a position to recognize it *as* a reason. McDowell (1994, pp. 52–53) insists that anything that is going to count as a reason-giving relation “must be able to come under the self-scrutiny of active thinking.” I think this means that your reasons must be available for you to think about and critically assess.
  - 22 See, for example, Brewer (1999, pp. 165–169). Pollock considers, and criticizes, a similar argument on behalf of what he calls “the doxastic assumption” that only beliefs can be justification makers (Pollock, 2001, p. 41; see also Pollock and Cruz, 1999).
  - 23 Brewer (1999, pp. 168–169) takes this line.
  - 24 Other philosophers have also argued against overly deliberate and reflective accounts of being guided by an epistemic norm. See Pollock and Cruz (1999, pp. 124–130 and 136–137), Millar (1991, esp. p. 121), and Reynolds (1991).
  - 25 For some recent discussion of the question whether belief is voluntary, see Alston (1989, chapter 5), Steup (2000), Feldman (2001), Audi (2001), and Ginet (2001).
  - 26 See also Van Cleve (1979), Pollock and Cruz (1999, p. 125), and Pollock (2001, pp. 44–45).

## References

- Alston, W. (1989) *Epistemic Justification*. Ithaca, NY: Cornell University Press.
- Audi, R. (1993) *The Structure of Justification*. Cambridge: Cambridge University Press.
- Audi, R. (2001) Doxastic voluntarism and the ethics of belief. In *Knowledge, Truth and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, ed. M. Steup. Oxford: Oxford University Press.
- BonJour, L. (1978) Can empirical knowledge have a foundation? *American Philosophical Quarterly* 15: 1–13.
- BonJour, L. (1985) *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.
- BonJour, L. (2001) Toward a defense of empirical foundationalism. In *Resurrecting Old-fashioned Foundationalism*, ed. M.R. DePaul. Lanham, MD: Rowman & Littlefield.
- Brewer, B. (1999) *Perception and Reason*. Oxford: Oxford University Press.
- Burge, T. (1986) Cartesian error and the objectivity of perception. In *Subject, Thought and Context*, ed. P. Pettit and J. McDowell. Oxford: Oxford University Press.
- Chisholm, R.M. (1989) *Theory of Knowledge*, 3rd edn. Englewood Cliffs, NJ: Prentice Hall.
- Cohen, S. (2002) Basic knowledge and the problem of easy knowledge. *Philosophy and Phenomenological Research* 65: 309–329.
- Davidson, D. (1986) A coherence theory of truth and knowledge. In *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, ed. E. LePore. Oxford: Blackwell.
- DePaul, M.R. (2001) *Resurrecting Old-fashioned Foundationalism*. Lanham, MD: Rowman & Littlefield.
- Dretske, F. (1981) *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.
- Evans, G. (1982) *The Varieties of Reference*. Oxford: Oxford University Press.
- Fales, E. (1996) *A Defense of the Given*. Lanham, MD: Rowman & Littlefield.
- Feldman, R. (2001) Voluntary belief and epistemic evaluation. In *Knowledge, Truth and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, ed. M. Steup. Oxford: Oxford University Press.
- Firth, R. (1964) Coherence, certainty, and epistemic priority. *Journal of Philosophy* 61: 545–557.
- Ginet, C. (2001) Deciding to believe. In *Knowledge, Truth and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, ed. M. Steup. Oxford: Oxford University Press.
- Harman, G. (1986) *Change in View*. Cambridge, MA: MIT Press.
- Heck, R.G. Jr (2000) Non-conceptual content and the “space of reasons.” *Philosophical Review* 109: 483–523.
- Hempel, C.G. (1934/5a) On the logical positivists’ theory of truth. *Analysis* 2: 49–59.
- Hempel, C.G. (1934/5b) Some remarks on “facts” and propositions. *Analysis* 2: 93–96. Reprinted in *Essential Readings in Logical Positivism* (1981), ed. O. Hanfling. Oxford: Blackwell.
- Hintikka, J. (1969) On the logic of perception. In *Models for Modalities*. Dordrecht: Reidel.
- Korcz, K.A. (1997) Recent work on the Basing relation. *American Philosophical Quarterly* 34: 171–189.
- Lewis, D. (1986) Veridical hallucination and prosthetic vision. In *Philosophical Papers, Volume 2*. New York: Oxford University Press.
- McDowell, J. (1994) *Mind and World*. Cambridge, MA: Harvard University Press.
- McDowell, J. (1998) One strand in the private language argument. In *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press.
- Martin, M. (1993) The rational role of experience. *Proceedings of the Aristotelian Society* 93: 71–88.
- Millar, A. (1991) *Reasons and Experience*. Oxford: Clarendon Press.
- Peacocke, C. (1983) *Sense and Content*. Oxford: Oxford University Press.
- Peacocke, C. (2001) Does perception have a non-conceptual content? *Journal of Philosophy* 98: 239–264.

- Plantinga, A. (2001) Direct acquaintance? In *Resurrecting Old-fashioned Foundationalism*, ed. M.R. DePaul. Lanham, MD: Rowman & Littlefield.
- Pollock, J. (2001) Nondoxastic foundationalism. In *Resurrecting Old-Fashioned Foundationalism*, ed. M.R. DePaul. Lanham, MD: Rowman & Littlefield.
- Pollock, J. and Cruz, J. (1999) *Contemporary Theories of Knowledge*, 2nd edn. Lanham, MD: Rowman & Littlefield.
- Pryor, J. (2000) The skeptic and the dogmatist. *Noûs* 34: 517–549.
- Pryor, J. (2001) Highlights of recent epistemology. *British Journal for the Philosophy of Science* 52: 1–30.
- Pryor, J. (2004) What's wrong with Moore's argument? *Philosophical Issues* 14: 349–378.
- Reynolds, S.L. (1991) Knowing how to believe with justification. *Philosophical Studies* 64: 273–292.
- Schlick, M. (1932/3) Über das Fundament der Erkenntnis. Reprinted as On the foundation of knowledge. In *Essential Readings in Logical Positivism* (1981), ed. O. Hanfling. Oxford: Blackwell.
- Schlick, M. (1934/5) Facts and propositions. *Analysis* 2: 65–70. Reprinted in *Essential Readings in Logical Positivism* (1981), ed. O. Hanfling. Oxford: Blackwell.
- Searle, J. (1983) *Intentionality*. Cambridge: Cambridge University Press.
- Sellars, W. (1963) Empiricism and the philosophy of mind. In *Science, Perception and Reality*. London: Routledge & Kegan Paul.
- Steup, M. (2000) Doxastic voluntarism and epistemic deontology. *Acta Analytica* 15: 25–56.
- Steup, M. (2001) Foundationalism, sense-experiential content, and Sellars's dilemma. <http://matthiassteup.files.wordpress.com/2013/01/sellarsian-dilemma1.pdf> (last accessed March 23, 2013).
- Unger, P. (1975) *Ignorance: A Case for Skepticism*. Oxford: Oxford University Press.
- Van Cleve, J. (1979) Foundationalism, epistemic principles, and the Cartesian circle. *Philosophical Review* 88: 55–91.
- Williams, M. (1977) *Groundless Belief*. Oxford: Blackwell.
- Williamson, T. (2000) *Knowledge and its Limits*. Oxford: Oxford University Press.

# There Is No Immediate Justification

Juan Comesaña

## Introduction

In his paper, Pryor argues for the existence of immediate justification and defends it against what he calls “the master argument for coherentism.” In this paper, I want to present and evaluate a different argument against one kind of immediate justification. My conclusion will be that, although the argument does not conclusively establish that there is no immediate justification in the cases to which it applies, it does show that the existence of that kind of immediate justification is incompatible with a set of widely accepted principles.

## Where Does Immediate Justification Come From?

Pryor defines “immediate justification” as follows:

When your justification to believe P does not come from your justification to believe other propositions, I will call it “immediate.”

The phrase “coming from” sounds innocuously prosaic, but we will see that it must be functioning in the definition of immediate justification as a technical term.

Right after presenting this characterization, Pryor makes five clarifications, two of which will be particularly important for our discussion. First, to put it in a common terminology, Pryor is talking about propositional rather than doxastic justification. Roughly speaking, propositional justification has to do with which propositions you are justified in believing, whether you believe them or not (and if you do believe them, whether you do so for good reasons or not); doxastic justification, on the other hand, has to do with whether your mental state consisting in believing a proposition is justified. We will be concerned with both propositional and doxastic justification. Second, Pryor believes that “justification is usually defeasible.” So, when he is talking about immediate justification, he is talking about *prima facie* justification, which can turn into all things considered justification if it is not defeated.

Pryor thinks that the best defense of the existence of immediate justification comes not from some theoretical argument such as a version of the regress argument, but rather from consideration of cases. He thus gives examples of what he takes to be cases of immediate justification:

- I feel tired, and I am thereby justified in believing that I feel tired.
- I have a headache, and I am thereby justified in believing that I have a headache.
- I raise my arm to scare a fly, and I am thereby justified in believing that I raise my arm to scare a fly.
- I imagine my grandmother sitting in her kitchen, and I am thereby justified in believing that I am imagining my grandmother sitting in her kitchen.
- I think about ways for a domino piece to cover two spaces on a chessboard, and I am thereby justified in believing that the only way to wholly cover two spaces on the board is to place the domino horizontally or vertically.

Pryor thinks that in each of these cases my justification for believing the relevant propositions does not come from my justification for believing any other proposition. A crucial question in what follows is: where does the justification come from? A good way to approach that question is to recall Pryor’s discussion of the master argument for coherentism. The version of the argument that Pryor settles on is the following:

**Premise Principle:** The only things that can justify a belief that P are other states that assertively represent propositions, and those propositions have to be ones that could be used as premises in an argument for P. They have to stand in some kind of inferential role to P: they have to imply it or inductively support it or something like that.

**Only Beliefs:** Only beliefs (or other states that are epistemically like beliefs) represent propositions assertively.

Therefore,

Only beliefs (or other states that are epistemically like beliefs) can be justifiers.

Pryor points out that there are reasons to doubt both of the premises in this argument. Against the Only Beliefs premise, many epistemologists hold that experiences are states that can represent propositions assertively without being epistemically like beliefs (that is to say, without needing – or indeed, allowing – to be justified in order to justify). When I have an experience as of hands in front of me, my experience assertively represents the proposition that there are hands in front of me, and that proposition could obviously be used as an argument for the conclusion that there are hands in front of me (although to call that an argument is to stretch the terminology to uncomfortable levels). Pryor also argues against the Premise Principle. I won't be concerned here with the details of Pryor's argument against the Premise Principle, but I will be concerned with exactly what aspect of the Premise Principle Pryor is rejecting.

Let us now return to Pryor's initial examples of immediately justified beliefs. Remember our question: if the justification in those examples doesn't come from justification for believing anything else, where does it come from? One might be tempted to think that Pryor's answer is the following: my justification for believing that I have a headache is my headache; my justification for believing that I feel tired is my feeling of tiredness, and so on. That is to say, in each of those cases, my justification comes from a state that doesn't assertively represent any content. Note well: the idea here would be to say that my justification comes from the state itself, not from the fact that I am in that state, or from the event of my being in that state: it is my headache that justifies me in believing that I have a headache, not the fact that I have a headache or my having the headache.

I said that this is what one might expect Pryor to say, but it is not what he actually says. What he actually says is the following:

Many foundationalists believe in justification-making facts that violate that constraint. For instance, many foundationalists want to allow facts about what sensations you are having, or facts about what mental activities you are engaging in, to count as justifiers.

According to Pryor here, it is not the states themselves that are the justifiers, but rather the fact that the subject is in such and such a state (or the event of the subject's being in such and such a state – I won't distinguish between these in what follows).<sup>1</sup> According to this view, for instance, facts about the existence of certain experiences are immediate *prima facie* justifiers for belief in the propositions that are the contents of such experiences. Thus, the fact that I am having an experience with the content that there is a hand in front of me is an immediate *prima facie* justifier for the proposition that there is a hand in front of me. It is for this kind of view that I now want to raise a serious problem.



## The Problem

The problem actually arises not only for that kind of view, but for a much larger class. To be susceptible to the problem, a view must satisfy four principles (eventually we will add a fifth), all of which are in fact widely accepted.

The first principle in question I will call “Inductivism”:

**Inductivism:** A justifier J may provide justification for a subject S to believe that P even if J doesn’t entail that P.

Notice that, to satisfy Inductivism non-trivially, a view must hold that, at least sometimes, justifiers can be propositions (or facts or events, to the extent that we can make sense of the idea that facts and events can stand in logical relations to propositions).<sup>2</sup> The view according to which facts about mental states can be justifiers satisfies Inductivism. To be sure, there will be cases where a justifier does entail the proposition that it justifies – for instance, that I have a headache may be where my justification for believing that I have a headache comes from. But there will be cases where the entailment fails, such as the one already mentioned where the fact that I have an experience with the content that there is a hand in front of me justifies me in believing that there is a hand in front of me. Of course, the proposition which is the content of the experience does entail the proposition that I am justified in believing (they are indeed the same proposition); but, according to the view we are now interested in, it is not the content of the experience, but the fact that I have it, which justifies me – and the fact that I have an experience does not entail that it is veridical. So, someone who believes in the existence of this kind of immediate justification will be committed to Inductivism. The converse, of course, doesn’t hold: one may be committed to Inductivism even if one doesn’t believe in the existence of immediate justification at all, let alone this specific kind of immediate justification.

Second, a principle of Closure:

**Closure:** If S has justification for believing that P and competently deduces Q from P, then S has justification for believing that Q.

Closure principles have been at the center of the epistemological discussion for a while now.<sup>3</sup> This is not the place for further substantive discussion about them, but I will make two points. First, I want to register my sympathy with defenders of Closure principles against their opponents. Nozick (1981) was right when he said (something to the effect) that we may quibble about the precise formulation of an exceptionless closure principle, but it is hard to deny that something like Closure must be right. (Nozick famously went on to deny closure, not on the basis of quibbling with the formulation of the principle, but rather on the basis of an analysis of knowledge.) Second, to generate the problem it is not necessary to hold that a principle like Closure holds with full generality. It is sufficient to hold that, in many particular cases, someone who is justified in believing a proposition P by a justifier which doesn’t entail P may competently perform a deduction from P to Q and thereby become justified in believing Q.

To introduce the third principle, consider the following scenario. Suppose that I throw a fair die and hide the result. I then tell you that the die landed either on 1 or on

2. That gives you some evidence to believe that the die landed on 1, of course – you now have more justification than you did before for believing that the die landed on 1. And that the die landed on 1 entails that it didn't land on 2. But think how absurd it would be if you reasoned this way: "I am now more justified than I was before in believing that the die landed on 1, and that the die landed on 1 entails that it didn't land on 2. Therefore, I am now more justified than I was before in believing that the die didn't land on 2." Why would this reasoning be so absurd? One reason (maybe not the only one) is that the proposition that the die landed on 1 is a mere lemma. All the justification you have for believing that proposition comes from your justification for believing that the die landed either on 1 or on 2. Of course, that the die landed on 1 or on 2 does not give you more justification than you had before for believing that it didn't land on 2 – on the contrary, it gives you more justification than you had before for believing that it did land on 2. Now, if you had independent justification for believing that the die landed on 1, then of course you might well be justified in believing that it didn't land on 2. But given that all your justification for believing that the die landed on 1 comes from your justification for believing that it landed either on 1 or on 2, and given that this latter proposition does not justify you in believing that it didn't land on 2, then you are not justified in believing that it didn't land on 2. In slogan form, mere lemmas do not have any justifying power of their own. In principle form:

**Mere Lemmas:** If S's justification for believing that P comes entirely from Q and Q does not justify S in believing R, then P does not justify S in believing R.<sup>4</sup>

Finally, suppose that P entails Q. Could Q then justify you in rejecting P? Hardly so. That Q is true is what would happen if P were true, so how can the truth of Q justify you in rejecting P? Our fourth principle enshrines this idea:

**Entailment:** If P entails Q, then Q cannot justify S in believing not-P.

Both Mere Lemmas and the Entailment principle are to be understood as restricted to contingent propositions – otherwise they would have counterexamples that are irrelevant for our purposes.

As I said, many theories accept these four principles. Moreover, there seem to be very good reasons to accept these four principles. I have given brief defenses of each, and more will be provided below, but I think that they enjoy a great deal of initial plausibility. Unfortunately, they are incompatible with each other. There is a proposition P of a certain kind such that Inductivism and Closure entail you might be justified in believing P, whereas Entailment and Mere Lemmas entail that you can't be justified in believing P.

Let us start by assuming that the fact that you are having an experience with a certain content (let us say, that there is a hand in front of you, and let's call this fact E) immediately justifies you in believing a certain proposition (let us say, that there is indeed a hand in front of you, "H" for short). As noted above, this case satisfies Inductivism non-trivially, for E does not entail H. Now, H entails the proposition that either H or not-E. Notice that this is equivalent to the proposition that it is not the case that E and not-H, and also equivalent to the material conditional that if E, then H. Let us suppose that you notice the entailment and competently deduce the proposition that either H or not-E. Closure now entails that you are justified in believing this proposition – although, as

noted above, nothing as strong as Closure is required to hold that you are indeed now justified in believing it. Now, a natural question to ask at this point is: what justifies you in believing that H or not-E? Or, to put it in Pryor's terminology, where does your justification for believing H or not-E come from? There seem to be only two candidates in the offing: your justifier is either E or H itself (later I reexamine the assumption that these are the only two possible justifiers in this case). Take the latter option first. We are assuming that all your justification for believing H comes from E. The Mere Lemmas principle then entails that H can justify you in believing only things which E itself justifies you in believing. So, if you are justified at all in believing H or not-E, it must be because E itself justifies you. But notice that the negation of H or not-E is (equivalent to) E and not-H, which entails E. Therefore, according to the Entailment principle, E cannot justify you in believing H or not-E. So, according to the four principles under consideration, you are justified and it is not the case that you are justified in believing H or not-E. Therefore, the four principles are mutually inconsistent.<sup>5</sup>

Several variations of this problem have received attention in the recent literature. Something like it was proposed as a problem exclusive to reliabilism by Fumerton (1996) and Vogel (2000), who called it the "bootstrapping problem." Fumerton and Vogel complained that reliabilism allows one to "bootstrap" perceptual knowledge into knowledge that perception is reliable. Something like it is also behind Wright's worries with Moorean approaches to solving the skeptical problem (see, for instance, Wright, 2007). White (2006) also marshals a version of the problem against theories such as Pryor (2000), which hold that experience can give us immediate justification. Cohen (2002) argued that, properly understood, the problem afflicts not only reliabilism, but any theory which admits that we can have knowledge on the basis of a source without knowing that the source is reliable – he called it the "easy knowledge problem." Independently of Cohen, Huemer (2001) argued that the problem is even more general, and it affects any theory which admits that there can be justification and knowledge on the basis of non-entailing evidence – he called it the "problem of defeasible justification." Pryor (forthcoming) argues that White's problem for dogmatism actually applies to any "credulist" theory – any theory according to which there can be what Pryor calls "non-quotidian" undermining. Recently, Sharon and Spectre (forthcoming) present a similar argument, but aimed against Closure.<sup>6</sup> I cannot here present a detailed comparison between the problems presented by these authors and my own. Suffice it to say that I agree with Cohen, Huemer, and Pryor that there is a problem in the vicinity which affects a great variety of epistemological theories. Indeed, I think it is a virtue of my presentation of the problem that it reveals just how widely it applies. I disagree with Wright and Sharon and Spectre that the solution to the problems lies in denying some version of Closure. I also disagree with White, Cohen, and again Wright that the solution lies in claiming that we have a priori knowledge of the proposition in question. I turn now to precisely this issue.

## **The Neo-Rationalist Gambit and Immediate Justification**

A crucial step in the argument that the four principles are mutually inconsistent is the claim that, in the cases under consideration, your justification for believing H or not-E can only come from two sources: either from E, or from H. Some philosophers will

deny this. They will say that you are a priori justified in believing that proposition – that is to say, you are justified but not on the basis of any empirical justifier such as H or E. Cohen (2010), for instance, holds a view of this sort and presents it as a solution to the easy knowledge problem. According to Cohen, given that E justifies H, a subject can assume E, defeasibly derive H, and then close the assumption by concluding that if E then H. The fact that this defeasible suppositional reasoning is easily available to any subject is what gives them a priori justification for believing that H or not-E, and so Cohen will block the argument presented in the previous section precisely at the point where we said that your justifiers for that proposition can only be E or H. Wedgwood (2012) holds a similar position. Wright (2004) and White (2006) also think that you have a priori justification for believing such propositions, but do not hold that what gives you this a priori justification is the kind of defeasible suppositional reasoning that Cohen advocates.

Is this neo-rationalist position (neo-rationalist because it posits a priori justification for contingent propositions) a way of denying the existence of immediate justification? Not without additional assumptions. One can hold on to the claim that it is E alone which provides me with justification for believing H and also claim that I have a priori justification for believing H or not-E. One way of implementing this idea is by saying that it is E by itself which justifies me in believing H, whereas it is this very fact (that E justifies me in believing H), a fact which obtains independently of whether E itself obtains, that justifies me in believing H or not-E. But this position is not very stable. Remember that, according to Pryor, your justification for believing a proposition is immediate if and only if it doesn't "come from" your justification for believing any other proposition. According to the neo-rationalist position, whenever some fact F gives you justification for believing a proposition P, you already have justification for believing the material conditional if F then P. If so, whenever you are justified in believing P on the basis of F you will have available to you a different justification – one which depends not only on F but also on the conditional if F then P. The justification in question will in addition be logically stronger – it will entail that P. This doesn't mean, of course, that it is epistemically stronger, but it is nevertheless remarkable that, according to neo-rationalists, whenever there is immediate justification for a proposition there is also non-immediate justification for it.

Of course, the friend of immediate justification is free to hold that even though you have justification for believing this conditional, your justification for believing the consequent "comes from" just the antecedent. This response, however, only highlights the obscurity in the phrase "comes from." If we were talking about doxastic justification, there would be a natural explanation for the phrase: your doxastic justification for believing a proposition (if indeed you are doxastically justified) comes from whatever you based your belief on. Presumably, there is a fact of the matter about what your mental state of believing a proposition is based on, and we can appeal to this fact in explaining the phrase "comes from." However, we are talking here about propositional justification, and no natural explanation of the phrase is available in this context. What determines whether your justification for believing P comes from (in part) the conditional or not? Perhaps one could try to argue for the claim that it doesn't by saying that you would still be justified in believing P even if you were not justified in believing in the conditional. But remember that, according to the neo-rationalist, justification for the conditional will be available whenever the antecedent justifies the

consequent, and so the counterfactual in question will have an impossible antecedent. Many people believe that this makes the counterfactual trivially true. Alternatively, one could try arguing that what explains why you are justified in believing P in the circumstances is primarily F, and the conditional is just icing on the cake. Some such explanation, in any case, is indeed needed if one is to reconcile the neo-rationalist position with the existence of immediate justification.

Whatever the prospects for a reconciliation of neo-rationalism with the existence of immediate justification, I will next argue that neo-rationalism cannot solve the problem we are interested in.

## Grounding

The neo-rationalist denies that E and H are the only possible justifiers for H or not-E in the cases with which we are concerned. The position allows us to say the following: when you concluded that H or not-E based on your competent deduction from H, you didn't acquire any justification that you didn't have before. You were already justified in believing that proposition, and so the appearance of pulling yourself by your own bootstraps is neutralized. You would be pulling yourself by your own bootstraps if you were to gain justification to believe H or not-E by following that procedure, but you do not gain anything. This way of formulating the neo-rationalist position allows us to see that it is indeed about propositional justification. For instance, you need not engage in the suppositional reasoning that, according to Cohen, gives you justification for believing H or not-E; and you need not avail yourself of the justification that according to Wright is yours "by default." Nevertheless, that justification is there available to you, and does give you propositional justification even if you do not ground your belief in it.

Consider now a subject who doesn't avail himself of that a priori justification. For all we know, this subject doesn't even reflect on the proposition that H or not-E before he acquires evidence E. Let us further suppose that he does indeed come to believe H, and it is indeed E which justifies him in so believing. Now he goes ahead and competently deduces H or not-E from H. Is the subject then justified in believing that proposition? The question, of course, is ambiguous between the following two: (i) does the subject have propositional justification for believing that H or not-E?; (ii) does the subject have doxastic justification for believing H or not-E?

It is crucial, in considering the answers to those questions, to reflect on the relationship between propositional and doxastic justification. Suppose that E gives me propositional justification for believing H. Suppose now that I do in fact believe H on the basis of E, and that nothing out of the ordinary is going on. Then, I am doxastically justified in believing H. Conversely, if I am doxastically justified in believing that H on the basis of E, then E must give me propositional justification for believing P. True, the clause that nothing out of the ordinary is going on may well hide interesting issues, but they are not issues that will affect our argument. I therefore propose the following principle:

**Grounding:** If S believes that Q based on P, then S is doxastically justified in believing that Q if and only if P provides propositional justification for S to believe that Q.

We should also distinguish two different versions of Closure, one applying to propositional and the other to doxastic justification:

**Propositional Closure:** If *S* is doxastically justified in believing that *P* and competently deduces *Q* from *P*, then *S* is doxastically justified in believing that *Q* on the basis of *P*.

**Doxastic Closure:** If *S* is doxastically justified in believing that *P* and competently deduces *Q* from *P*, then *S* has propositional justification for believing that *Q*.

Notice that, as I am understanding the Closure principles, the justification in question in the antecedent can only refer to doxastic justification – one may well notice that *Q* follows from *P*, but one can hardly deduce *Q* from *P* unless one already believes *P*.<sup>7</sup> Principles similar to Doxastic Closure sometimes go under the name of “transmission” principles – but I do not want to commit myself on how closely related the different transmission principles and my Doxastic Closure are to each other.

Let us now revisit our sample case. Given Inductivism and Grounding, you are doxastically justified in believing *H*. Now, according to Doxastic Closure, you are then doxastically justified in believing *H* or not-*E*. You believe *H* or not-*E* on the basis of *H*, and so Grounding entails that *H* must provide you with propositional justification for believing *H* or not-*E*. But, according to Mere Lemmas, that can only happen if *E* provides you with propositional justification for believing *H* or not-*E*, and according to the Entailment principle it cannot. Hence, the problem reappears.<sup>8</sup>

To recap: the original problem arises from the incompatibility of Inductivism, Closure, Mere Lemmas, and Entailment, plus the assumption that only *E* or *H* can justify you in believing *H* or not-*E*. Neo-rationalism challenges this assumption, positing a priori propositional justification for believing that proposition. But close attention to whether doxastic justification is possible in those cases shows that the problem can be reinstated: Inductivism, Doxastic Closure, Mere Lemmas, Entailment, and Grounding are mutually incompatible, and this time neo-rationalism has nothing to say about the problem. The existence of a priori propositional justification for believing a proposition is irrelevant to whether a subject who doesn't avail himself of that justification is doxastically justified in believing that proposition.

## Options

Let me now consider the prospects for denying each of the principles as a way out of the problem. Some philosophers have proposed that we reject Closure. For instance, Dretske (1970) and Nozick (1981) proposed theories of knowledge according to which closure principles for knowledge fail. Under plausible assumptions about the connection between knowledge and justification (plausible although, notably, not shared by Nozick and Dretske themselves), that means that our Closure fails. Most philosophers have taken these failures of closure principles to be powerful arguments against those theories, however. Other philosophers (notably, Wright) have held that although Propositional Closure is true, (something like) Doxastic Closure is false. As we saw, Wright concedes that if you are justified in believing *H* then you are also justified in believing *H* or not-*E*, but thinks that the justification for believing this latter proposition

does not come from your justification for believing  $H$  – it is rather available to you a priori. As far as I can tell, however, this combination of accepting Closure but denying transmission simply ignores the complications that arise when dealing with doxastic justification. According to Wright, you may well be doxastically justified in believing a proposition  $P$ , competently deduce  $Q$  from  $P$ , and yet fail to be doxastically justified in believing  $Q$  (let us suppose that you don't avail yourself of the propositional justification for believing the  $Q$  in question available to you a priori, relying instead only on your impeccable deduction from  $P$ ). Finally, whatever one thinks about Doxastic Closure, we must remember that, to generate the problem, we need not rely on any general Closure principle, but rather it is sufficient to hold that in particular cases a competent deduction from  $H$  to  $H$  or not- $E$  can give you doxastic justification for believing this latter proposition.

Klein (1995) has in effect denied Mere Lemmas. To be completely fair to Klein, he has defended the conditional claim that if we are to hold on to plausible closure principles, then Mere Lemmas has to go (although I think that Klein is clear that he does want to hold on to those closure principles, and so the stronger claim is also true). This also strikes me as a non-starter. To deny Mere Lemmas is to attribute magic powers to mere lemmas; it is to accept that epistemic justification can be created *ex nihilo*; more importantly, it is just to get the cases wrong.

I anticipate that many philosophers friendly to the existence of immediate justification will want to deny the Entailment Principle. Indeed, Pryor (forthcoming) himself has taken this route. To me, the principle is undeniable. Pryor does two things. First, he points out the enormous class of cases for which the principle makes trouble. On this, I agree wholeheartedly. Second, he argues against one particular way of supporting the principle. I have something to say about this.

Entailment can be supported by appeal to some probabilistic considerations. Suppose that you hold that  $E$  can justify  $P$  only if there is some probability function such that the conditional probability of  $P$  given  $E$  (in symbols,  $\Pr(P \mid E)$ ) is higher than the unconditional probability of  $P$  ( $\Pr(P)$ ). One may think this because one buys into a complete Bayesian package, according to which rational credences must be probabilities and must evolve according to the conditionalization rule.<sup>9</sup> But one need not buy into the full Bayesian package to believe that there must be some probability function (not necessarily one which models the subject's credences) according to which  $\Pr(P \mid E) > \Pr(P)$  (even if one doesn't think that rational update of credences operates only by conditionalization) if one is to be justified in believing  $P$  on the basis of  $E$ . Now, it is a theorem of the probability calculus that if  $P$  entails  $E$  then  $\Pr(\text{not-}P \mid E) < \Pr(\text{not-}P)$  (provided that both  $P$  and  $E$  have non-extreme probabilities to begin with). Therefore, the Entailment principle follows from the probabilistic relevance constraint on justification. Attacking this probabilistic relevance, or perhaps some elements of its implementation, thus removes one important source of justification for the Entailment principle.

But one need not buy into a probabilistic account of justification to find the Entailment principle very plausible. Indeed, when I presented the Entailment principle I made a brief argument in its defense which didn't appeal to any probabilistic considerations. To repeat that argument in evidential terms, the idea is that if a certain proposition entails some evidence, then you cannot use that very evidence to reject the proposition. As far as that piece of evidence is concerned, everything is as it should be

when the proposition is true. Therefore, misgivings about probabilistic accounts of justification are not enough to dismiss the Entailment principle, which enjoys independent support.<sup>10</sup>

One could also deny the Grounding principle, which I used to reinstate the problem even for neo-rationalist positions. As it stands, the Grounding principle is certainly false. For instance, it is perfectly possible for me to have propositional justification for believing that I have no higher-order beliefs (beliefs about beliefs), but it is hard to see how I could ever be doxastically justified (in the way required by the Grounding principle) in so believing. Perhaps we should also take into account the manner in which I ground my belief in whatever provides me with propositional justification – for instance, if I believe that  $Q$  on the basis of  $Q$  and if  $P$  then  $Q$  but do it by applying the “rule” that from propositions of the form  $A$  and if  $B$  then  $C$  one can infer that  $C$ , then I don’t have doxastic justification.<sup>11</sup> But the solutions to these problems, though perhaps not easy to state, shouldn’t lead us to think that there is nothing to the Grounding principle. The Grounding principle is true, details aside. The details are important, of course, but not likely to interfere with the use to which we are putting the principle here.

Finally, we should take a careful look at Inductivism. Many philosophers would think it mad to give it up, for giving it up seems to invite skepticism and uphold the indefeasibility of justification. Although this is not the place to do it, I believe that a case can be made that abandoning Inductivism need not lead to either skepticism or indefeasibility. In a nutshell, we can avoid skepticism by holding that evidence that we usually do have does entail the propositions that we are justified in believing – for instance, we can hold that it is part of my evidence that I see that there is a hand in front of me. This position need not lead one to hold a “disjunctivist” view according to which our evidence differs from “good” to “bad” cases – roughly speaking, ordinary vs. skeptical scenarios – provided that one is comfortable with the possibility of false evidence, a possibility which I have argued for on independent grounds.<sup>12</sup> When it comes to defeasibility, we should be careful not to confuse it with non-entailment. To paraphrase Pryor (2000), defeasible justification is justification which can be lost given improvements in one’s epistemic position.<sup>13</sup> It is of course possible for one’s epistemic position to improve and, as a result, for one to lose justification to believe something entailed by one’s previous (and perhaps even present) evidence. Suppose, for instance, that you have a proof of some complicated theorem. You then show it to a very competent mathematician and (mistakenly) he assures you that there is a fallacy in the proof. You then cease to be justified in believing the theorem, even though you did have (and perhaps still have) evidence which entails it. Therefore, giving up Inductivism is irrelevant to holding on to the defeasibility of justification.

The believer in the kind of immediate justification that I have been examining here is committed to Inductivism. But giving up Inductivism need not mean giving up on immediate justification. To be sure, some ways of giving up Inductivism are indeed ways of giving up on immediate justification. For instance, to put it in terms of our example, if one thinks (as some neo-rationalists may do) that justification for believing  $H$  or not- $E$  is always involved in one’s justification for believing  $H$  on the basis of  $E$ , then one is thereby denying that  $E$  gives us immediate justification to believe  $H$ . But one may also think that my justification for believing that there is a hand in front of me is that there is a hand in front of me (although I only have this justification because



I have an experience with that content). This would count as immediate justification according to Pryor's definition. The connection between Inductivism and immediate justification is therefore not straightforward.

## Conclusion

Although, as just said, a commitment to immediate justification does not entail a commitment to Inductivism, many contemporary epistemologists are Inductivists *because* they hold that there is immediate justification which satisfies Inductivism. If, in addition, they also hold, as I have argued they should, that Closure (or better, particular applications of it), No Lemmas, and Entailment are true, then they are in trouble. There are at least two ways out of the trouble. They might give up on the existence of immediate justification altogether. Neo-rationalism is sometimes seen as a solution of this kind to the problem. I have raised doubts, however, both about neo-rationalism as a solution to the problem as well as about its supposed abandonment of immediate justification. The other way out of the problem consists not in giving up immediate justification, but in giving up inductivist immediate justification. This is the kind of solution that I think merits more attention than philosophers have given it so far.

## Acknowledgments

Many thanks to Stewart Cohen, Jim Pryor, Carolina Sartorio, and Matthias Steup for comments on drafts of this paper. Thanks also to audiences at a Conference on Epistemology organized by the Center for the Study of Mind in Nature in Lofoten, Norway, in May of 2012 and a talk at the Sociedad Argentina de Análisis Filosófico, Buenos Aires, June 2012.

## Notes

- 1 For relevant discussion, see Turri (2009).
- 2 What about the view that it is the states themselves which are the justifiers? Does that view satisfy Inductivism? On the face of it, it doesn't, because states cannot enter into logical relations with propositions. Some philosophers may wish to hold that they do, in which case we would have to look into the details of their theories to see whether they satisfy Inductivism or not. Some of the issues here resemble those raised by "Sellars's dilemma" – see Sellars (1956).
- 3 See the essays by Dretske and Hawthorne in Chapter 2 of this book.
- 4 Weisberg (2010) proposes a similar principle, which he calls "No-Feedback."
- 5 Inductivism is a possibility principle: it says that it is possible for there to be a case of inductive justification. Together with Closure, it entails that it is possible for a subject to be justified in believing H or not-E when E is her evidence for H. But (subject to the provisos discussed in the next section) Entailment and No Lemmas entail that there is no such possible case.
- 6 I reply to Sharon and Spectre in Comesaña (forthcoming).

- 7 The reader may take the comment in the text as an indication of how I intend to use “deduce.”
- 8 For a similar argument, see Silins (2005).
- 9 The conditionalization rule has it that if E is all your evidence at a time *t*, then your confidence in any proposition P at *t* must be equal to the conditional probability of P on E.
- 10 For further arguments for Entailment, see my reply to Pryor in the essay below.
- 11 For relevant discussion, see Turri (2010).
- 12 See Comesaña and McGrath (forthcoming).
- 13 See Pryor (2000, p. 517). In that paper, however, Pryor sometimes slips and talks as if there were no difference between defeasible justification and non-entailing justification. For instance, on the very next page he talks of “defeasible justification, justification that does not *guarantee* that our beliefs are correct” (Pryor’s emphasis), and on the same page he talks of “defeasible, [i.e.] ampliative considerations.” Even more explicitly, footnote 11 on p. 543 includes the following: “The fallibilist’s central thesis is precisely that it’s possible to know things on the basis of defeasible evidence, evidence the possession of which is meta-physically compatible with your being wrong.”

## References

- Cohen, S. (2002) Basic knowledge and the problem of easy knowledge. *Philosophy and Phenomenological Research* 65 (2): 309–329.
- Cohen, S. (2010) Bootstrapping, defeasible reasoning and a priori justification. *Philosophical Perspectives* 24 (1): 141–159.
- Comesaña, J. (forthcoming) On Sharon and Spectre’s case against closure. *Philosophical Studies*.
- Comesaña, J. and McGrath, M. (forthcoming) Having false reasons. In *Epistemic Norms*, ed. C. Littlejohn and J. Turri. Oxford: Oxford University Press.
- Dretske, F. (1970) Epistemic operators. *Journal of Philosophy* 67; 1007–1023.
- Fumerton, R. (1996) *Metaepistemology and Skepticism*. Lanham, MD: Rowman & Littlefield.
- Huemer, M. (2001) The problem of defeasible justification. *Erkenntnis* 54: 375–397.
- Klein, P. (1995) Skepticism and closure: why the evil genius argument fails. *Philosophical Topics* 23 (1): 213–236.
- Nozick, R. (1981) *Philosophical Explanations*. Cambridge, MA: Harvard University Press.
- Pryor, J. (2000) The skeptic and the dogmatist. *Noûs* 34: 517–549.
- Pryor, J. (forthcoming) Problems for credulism. In *Seemings and Justification: New Essays on Dogmatism and Phenomenal Conservatism*, ed. C. Tucker. Oxford: Oxford University Press.
- Sellars, W. (1956) Empiricism and the philosophy of mind. *Minnesota Studies in the Philosophy of Science* 1: 253–329.
- Sharon, A. and Spectre, L. (forthcoming) Evidence and the openness of knowledge. *Philosophical Studies*.
- Silins, N. (2005) Transmission failure failure. *Philosophical Studies* 126: 71–102.
- Turri, J. (2009) The Ontology of Epistemic Reasons. *Noûs* 43 (3): 490–512.
- Turri, J. (2010) On the relationship between propositional and doxastic justification. *Philosophy and Phenomenological Research* 80 (2): 312–326.
- Vogel, J. (2000) Reliabilism leveled. *The Journal of Philosophy* 97 (11): 602–623.
- Wedgwood, R. (2012) A priori bootstrapping. In *The A Priori in Philosophy*. Oxford: Oxford University Press.
- Weisberg, J. (2010) Bootstrapping in general. *Philosophy and Phenomenological Research* 81 (3): 525–548.

- Wright, C. (2004) Warrant for nothing (and foundations for free)? *Aristotelian Society Supplementary Volume* 78 (1): 167–212.
- Wright, C. (2007) The perils of dogmatism. In *Themes from G.E. Moore: New Essays in Epistemology*, ed. S. Nuccetelli and G. Seay. Oxford: Oxford University Press.
- White, R. (2006) Problems for Dogmatism. *Philosophical Studies* 131 (3): 525–557.

## Reply to Comesaña

James Pryor

(1) In my essay, I touched briefly on what I thought were the best reasons to think there is any immediate justification at all. But much of my discussion, there and elsewhere, is concerned not with that broad genus, but with two (overlapping) species of it: (i) immediate justification which is also fallible, defeasible, and especially underminable; and (ii) immediate justification where the source needn't have a content that stands in an argumentatively nice relation to the proposition justified. In my essay, I spent the most time arguing, against the classical Davidsonian "Master Argument," that the case for resisting justification of type (ii) is obscure.

The "dogmatist" epistemology of perception that I've defended elsewhere doesn't obviously require there to be immediate justification of type (ii). On many contemporary views of perceptual experience, and in my own initial thinking, experiences do have contents that stand in some kind of argumentatively nice relation to the propositions they justify. But I've come to think it is less important than I did before whether that is so. Moreover, I suspect that having a reasonable story about the ways in which perceptual justification can be undermined will independently require a dogmatist to overcome resistance to justification of type (ii).

Comesaña's essay also does not focus on the possibility of immediate justification quite generally, but rather on a particular species, roughly what I've here labeled type (i). He presents what deserves to be recognized as a second "Master Argument" in the direction of coherentism.<sup>1</sup>

The argument has several premises. Klein and Silins have entertained rejecting the Mere Lemmas principle, but most philosophers will agree with that premise.<sup>2</sup> Closure is more controversial. As Comesaña points out, he doesn't really need to rely on a Closure principle but only on the conditional "If you have justification for these premises, you also have justification for this consequence" being true in some particular cases. Some may argue that, even if we accept Closure, Comesaña's examples still illustrate a failure of "warrant transmission" rather than the falsity of Inductivism. It is not clear to me whether that can be sustained in the dialectical setting Comesaña has presented, and, like him, I will not explore that strategy.

I will focus on the conflict Comesaña highlights between Inductivism and his "Entailment principle." As Comesaña says, this principle "enshrines" the idea that:

That Q is true is what would happen if P were true, so how can the truth of Q justify S in rejecting P?

I have discussed this motivating idea elsewhere.<sup>3</sup> I agree the idea is intuitively compelling. If my friend Sandra tells me (Q) the FBI's silence about captured aliens justifies her in thinking they're hiding something, and hence in rejecting that (P) they don't have any captured aliens, I want to protest: but if they didn't have any, this silence is just what you'd expect! All the worse for Sandra's argument if, as in Comesaña's examples, the hypothesis being rejected entails the observations that allegedly tell against it.

In the end, though, I think this idea is seductively false. Of course Sandra's argument is a bad one; but I think the Entailment principle must be an overgeneralization about why. I don't think it is an easy matter to persuade one of this, so all I can do here is try to motivate some second thoughts about the principle. I'll also suggest that it's not just the Inductivist that the principle threatens: some of the Inductivist's opponents are in no position to embrace Comesaña's argument, either.

(2) Comesaña proposes to ignore the non-contingent, but I can't see how to assess the Entailment principle without looking at aspects of how entailment and justification interact that lie beyond where he directs our attention. Some of the doubts I'll raise require you to attend to the difference between two kinds of case:

- Case 1. You have some background justification B, which entails not-P, but doesn't yet justify you in rejecting P.
- Case 2. You have the same background justification B, and acquire some extra evidence X, and B and X together do justify you in rejecting P.

If you agree that we sometimes lack justification for believing things entailed by our evidence, and that this lack can be remedied, then presumably you'll think of the remedy as working like X does in Case 2.

(3) Now, Comesaña's Entailment principle says:

If P entails Q, then Q cannot justify S in believing not-P.

I presume that whatever underwrites that principle would also underwrite the more general:

If P entails Q (against the unchallenged background of B), then Q (together with B) cannot justify S in believing not-P.

There may be complications here, especially if we seek a version of this principle applicable to degrees of confidence in P, but I will assume that in the kinds of cases Comesaña is considering, these principles stand or fall together.

(4) Take some conspiracy theory, like Sandra's. Imagine the best possible evidence you can against her theory, which falls short of entailing its falsehood. If you are an Inductivist, you probably agree there is such evidence – you needn't also subscribe to more controversial claims about fallible immediate justification. Now, if Sandra has any skill as a conspiracy theorist, you know what will happen. Her original theory will get supplemented with additional convolutions to explain why your evidence is just what we should expect. Imagine she takes a shortcut and just conjoins your evidence to her original theory. So now her new theory entails your evidence.

Now, you only committed to that evidence telling against her original theory; perhaps you hesitate to say it tells against her new theory. So go gather evidence telling against the new theory: about why the combination of her old theory and the new conjunct are improbable and incredible. Let's suppose this new evidence also falls short of entailing her theory is false. So she will conjoin that and present her third theory ...

If you're an Inductivist, then I expect you'll be inclined to think at some stage – if not already at the first – you will be justified in believing one of Sandra's theories is false, even though her theory entails your evidence. (If not, I have some rare grue emeralds you may be interested in ....) Note that Sandra's theories have none of the explanatory virtues of typical skeptical hypotheses. They only manage to entail all your evidence because they enumerate that evidence.

Let me acknowledge one possible source of hesitation. If your grounds for disbelieving Sandra's theory are entailed by her theory, you may hesitate to call those grounds "evidence" against her theory. But many Inductivists will still think it is possible to be justified in disbelieving Sandra's baroque conspiracy. And you will have some grounds for so believing. Such grounds are all that Comesaña means by "evidence" or "justifier." These Inductivists will agree that these grounds make it reasonable for you to reject a theory, whose truth would entail those grounds. In other words, they should think Comesaña's Entailment principle here is false.

Of course, Comesaña agrees this is what Inductivism commits one to (given his other premises). What I am trying to show is that you should feel no shame in that commitment. This is just a natural application of Inductivism, with no appeal to any of its more controversial extensions. There does seem to be some truth in the vicinity of Comesaña's Entailment principle, as my reaction to Sandra's opening argument evidenced. We haven't yet identified what that is, and Comesaña can fairly hold that against us. But despite its initial allure, an Inductivist who thinks it's possible to reasonably reject any of Sandra's conspiracy theories, on some grounds, should feel secure in rejecting the Entailment principle. As I see it, the challenges here are how to reconcile our intuition that these grounds justify us with our reluctance to count them as "evidence"; and to identify what real truth it is that the Entailment principle wrongly generalizes.

(5) One response Comesaña may make is that Sandra's theories must at some point become intrinsically incredible, given our background evidence. Perhaps the original theory was even inconsistent with that background – though if so, let's suppose not in a way that's immediately obvious. Comesaña may object that it is this intrinsic incredibility, rather than any evidence, that justifies you in rejecting the conspiracy theories. Hence, these cases don't undermine the Entailment principle. (Leaving that principle free to – in other cases – undermine Inductivism.)

Even if Sandra's theories were intrinsically incredible from the outset – and even if they were inconsistent with our background evidence – we are supposing this wasn't initially obvious. Nor should it have been obvious. Instead, there was some first evidence X that, together with your background evidence, brought you to the point of being able justifiably reject one of her theories. This is kind of situation envisaged in (2), above. Now, if her theory at that point is P, perhaps you don't think that your background evidence plus X are enough to justify you in rejecting  $P+X$ , though they

were enough to justify you in rejecting P. In the previous section, I said that even if an Inductivist hesitates at this stage, he is likely to think there's some point, some evidence Y, such that your background evidence plus X plus Y are enough to justify you in rejecting  $P+X+Y$ .

Perhaps in some sense, P will have already had to carry the seeds of its own rejection in it, for this to be so. (Or at least, the seeds of its own rejectability given B.) But those seeds needn't be epistemically transparent to us. Later evidence like X and Y can be needed before we're in a position to justifiably harvest them.

(6) Finally, let's turn to the question of whom Comesaña's argument threatens. In his essay, he discusses "neo-rationalists" who oppose a specific form of Inductivism, that takes our perceptual justification to be fallible and immediate. Comesaña discusses whether those Inductivist views are reconcilable with claims the neo-rationalists make. He also argues by appeal to his Grounding principle that the a priori justification the neo-rationalists posit doesn't prevent the kind of difficulty he's pressing against the Inductivists from arising. These issues are complex and I don't agree with everything Comesaña says, but we can't pursue it here. In any event, the kind of neo-rationalists he's envisaging may well be Inductivists elsewhere – for example, about induction – so they should be as eager as their opponents to resist Comesaña's Master Argument against Inductivism. Given the options we're considering, that means resisting the Entailment principle.

But what about a philosopher who is resolutely and thoroughly a rationalist – who rejects any form of Inductivism. Would such a philosopher be in a position to coherently embrace the argument?

Not necessarily. There may be some non-Inductivist views, perhaps the one Comesaña expresses his own sympathy for, that the argument does not threaten. But I think other non-Inductivist views may want also to reject the Entailment principle.

Seeing this requires keeping in mind the contrast described in (2), above. For given the kind of view we're now considering, any counterexample to the Entailment principle must take this form: some evidence Q supports not-P in a non-Inductive way, and hence entails not-P (at least, against your background evidence B). At the same time, P (against B) entails Q. Hence, P must already be inconsistent with B. However, as we've already noted, that doesn't necessarily mean B already justified you in rejecting P. It may be that only B plus Q do that. If we don't forget the difference between entailing and justifying, I expect that some rationalist views will want to diagnose our epistemic position with respect to Sandra's conspiracies in just this way.

More generally, the issue is that even reason-based evidence can be undermined. One can acquire evidence that at least makes it likely that one's reasons are unsupportive, or are not competently in hand. If our rationalists are going to be non-skeptics, they'll probably want our use of reasoning faculties to make some justified headway against skeptical hypotheses that threaten to undermine them in those ways. Like the Inductivist, they may want to do so even when those hypotheses entail whatever it is that the rationalist deems is our evidence. And as we've observed, our having justification to reject the skeptical threats may sometimes require this, even in cases where the threats entail their own falsity.

## Notes

- 1 As Comesaña notes, Michael Huemer has also stated, but not endorsed, something like this argument.
- 2 See Silins (2005, section 3.2).
- 3 See principle “NFI” in my “When Warrant Transmits” (Pryor, 2012).

## References

- Pryor, J. (2012) When warrant transmits. In *Wittgenstein, Epistemology and Mind: Themes from the Philosophy of Crispin Wright*, ed. A. Coliva, pp. 269–303. Oxford: Oxford University Press.
- Silins, N. (2005) Transmission failure failure. *Philosophical Studies* 126: 71–102.

# Reply to Pryor

*Juan Comesaña*

The meat of Pryor’s reply is what he takes to be a counterexample to Entailment. My main objective in this reply is to show that Entailment survives a proper account of Pryor’s case. Before doing that, however, I briefly address some other points.

Pryor thinks that Entailment is related to other principles. He draws our attention to his discussion of principle NFI in Pryor (2012), and also considers a version of Entailment not restricted to contingent propositions. Interesting as they are, I will set those other principles aside, for I do not see how they are related to my argument. Entailment is not obviously related to Pryor’s NFI, which states that if *E* is compatible with *H*, which is an alternative to *H*, then *E* cannot justify believe in *H*. NFI is a direct rejection of Inductivism, Entailment is not. At the end of my essay, I urged on the reader the possibility of rejecting Inductivism as a solution to the puzzle I presented. But this, of course, doesn’t mean that Entailment itself entails a rejection of Inductivism – it does so only in conjunction with Closure and No Lemmas. Pryor shows also, toward the end of his reply, that if Entailment is not restricted to contingent propositions then puzzles arise regarding cases where our (justified) lack of logical omniscience shows up. Indeed, that is precisely why I restricted Entailment to contingent propositions. I do not think, therefore, that the kind of case with which Pryor ends represents an objection to the argument of my essay.<sup>1</sup>

Now, on to what Pryor takes to be a counterexample to Entailment. This is the case of Sandra the conspiracy theorist, who, when faced with evidence against her theory, simply incorporates that evidence as a conjunct of her new theory. Her revised conspiracy theory now entails the evidence that went against her original conspiracy theory. Therefore, according to Entailment, we cannot use that evidence to reject the revised conspiracy theory. But surely we are justified in rejecting even her revised conspiracy theory. Therefore, Pryor argues, we should reject Entailment.

But there is a gap in that argument (Pryor is perfectly aware of this gap). We can all agree that we are justified in rejecting Sandra's revised conspiracy theory. Entailment is consistent with this claim. What we cannot do, if Entailment is true, is reject Sandra's revised conspiracy theory *on the basis of evidence which the theory itself entails*. Still, Pryor may now ask a fair question: if it is not on the basis of the evidence which the theory entails, on *which* basis do we reject Sandra's revised conspiracy theory? In what follows, I first offer arguments for thinking that Entailment gives the correct verdict in Pryor's case, and then answer that fair question.<sup>2</sup>

What were your opinions about Sandra's different conspiracy theories *before* you encountered them and the evidence that allegedly tells against them? Now, perhaps you never thought about those conspiracy theories before Sandra told you about them, so the answer to my question is "Nothing." But I do not really mean "before" in a chronological sense, but rather in a logical sense. I want you to think about your opinions about Sandra's conspiracy theories *bracketing* the relevant evidence that you now have. To illustrate the point, consider the fact that you now have evidence that Pryor thinks that Entailment is false. Even if, before reading Pryor's reply, you had never considered whether Pryor liked Entailment or not, you can now reflect on how credible it was that Pryor would reject Entailment bracketing the evidence that his reply gives you. I want you to do the same with Sandra's conspiracy theories. Because it simplifies exposition, I will talk as if the chronological interpretation of "before" were correct, but nothing hangs on this.

Sandra has two conspiracy theories.<sup>3</sup> Let's call her first conspiracy theory *not-H*, because it denies some widely accepted hypothesis *H*. The revised one is the conjunction of *not-H* with evidence that overwhelmingly tells in favor of *H*. Let's summarize that evidence as saying that everything tells in favor of *H* (short of entailing *H*, something which I take for granted in what follows). So, Sandra's original theory is *not-H* and her revised theory is *not-H but everything tells in favor of H*. You still don't know whether *everything tells in favor of H* is true or not, but you start wondering about possible conspiracy theories. You first consider *not-H*, and rationally have a very low opinion of it (it is a conspiracy theory, after all).<sup>4</sup> But what should your opinion of *not-H but everything tells in favor of H* be? I take it that it is obvious that it should also be very low (but I examine Pryor's alternative take on this later on). After all, if *everything tells in favor of H* is true, then most likely *H* is true. Now, *everything tells in favor of H* doesn't entail that *H*, so you may want to leave some room for the possibility that the evidence is massively misleading and *not-H* is true but *everything tells in favor of H* is also true – but that room can be very, very small. So, even before knowing that *everything tells in favor of H* is true, your opinion of *not-H and everything tells in favor of H* should be very low. Now consider what should happen when you do learn that *everything tells in favor of H* is true. According to Pryor, what happens is that your confidence in *not-H but everything tells in favor of H* goes down enough for you to reject it on the basis of *everything tells in favor of H*. However, several interrelated considerations suggest that, if anything, your confidence in *not-H but everything tells in favor of H* should go up – although maybe just a tiny bit up. Now, if those considerations are convincing, then of course you cannot rationally reject Sandra's revised conspiracy theory on the basis of that evidence, and so the verdict of Entailment for that case is vindicated.

I offer here three of those considerations. First, probabilistic construals of evidence have it that your confidence in a conjunction should go up when you learn that a



conjunct is true. Like Pryor, however, I don't want to put too much stock on probabilistic construals of evidence. But there are non-probabilistic considerations that point in the same direction.

Second, consider the fact that, when you learn that *everything tells in favor of H* is true, you have eliminated one way in which *not-H but everything tells in favor of H* could have been false. How could eliminating one way in which a proposition can be false not give you more confidence in that proposition? True, you have also eliminated one way in which *H and everything tells in favor of H* could have been false, and so your confidence in this proposition also goes up. Indeed, your confidence in this latter proposition should rise much more than your confidence in the previous one (that is why you end up justified in believing *H*, after all). But still, your confidence in a conjunction cannot but be raised (if only a tiny bit) when you find out that a conjunct is true.

Finally, here is what I take to be a decisive argument for the claim that learning that *everything tells in favor of H* is true should make your confidence in Sandra's revised conspiracy theory go up. Everyone should accept that *it is not the case that everything tells in favor of H* is good evidence for rejecting Sandra's revised conspiracy theory. After all, *it is not the case that everything tells in favor of H* entails that *not-H but everything tells in favor of H* is false (and I am supposing here that you know this to be so). But Pryor wants to claim that *everything tells in favor of H* also constitutes good grounds for rejecting *not-H but everything tells in favor of H*. That is, Pryor is committed to saying that a proposition and its negation can both be evidence for the same hypothesis (in this case, that Sandra's revised theory is false). This cannot be right. To illustrate the implausibility of saying that both *E* and *not-E* can provide evidence for the same proposition, suppose that you are wondering whether a hypothesis is true, and know that you are facing one of those situations where both *E* and *not-E* provide evidence for the hypothesis in question. It seems that, absurdly, you should increase your confidence in the hypothesis even before acquiring any evidence. Indeed, if both *E* and *not-E* are evidence for the hypothesis, then why isn't your confidence in the hypothesis higher to begin with? You know that, no matter which of *E* or *not-E* you learn, you will rationally increase your confidence in *H* (and not for tricky reasons, such as because you will forget relevant information). So why not increase it right now and not go to the trouble of inquiring into *E*?

Those considerations can be further generalized. Carolina Sartorio and I have argued, on independent grounds, for the following principle (Comesaña and Sartorio, forthcoming):

Epistemic Difference-Making (EDM): If *E* is evidence for *H*, then *not-E* is not evidence for *H*.

Notice that an obvious instance of EDM is that if *not-E* is evidence for *H*, then *E* is not evidence for *H*. EDM entails Entailment (and, therefore, obviously entails that Entailment gives the right verdict in Pryor's case). For suppose that *H* entails *E*. In that case, of course, *not-E* is evidence for *not-H*. Given EDM, *E* cannot therefore be evidence for *not-H* as well. I assume that to reject a hypothesis on the basis of some evidence the evidence must count against the hypothesis. Therefore, if *H* entails *E*, *E* cannot be used to reject *H* – that is to say, Entailment is true.

Having defended Entailment and the verdict that it gives in Pryor's case, let me go back to the fair question: if it is not the information that *everything tells in favor of H* is true that allows us to reject Sandra's revised conspiracy theory, what exactly does allow us to reject it? Well, it is simply the fact that it is so unlikely to be true. I said that the truth of *everything tells in favor of H* is some evidence for *not-H* but *everything tells in favor of H*, but that leaves it open that it can be extremely flimsy evidence. So, given that your previous confidence in Sandra's revised conspiracy theory was extremely low, and given that it has increased only by a very tiny margin, your confidence in that theory is still extremely low – low enough to justify you in continuing to reject it. As Pryor puts it, it is the “intrinsic incredibility” of Sandra's revised conspiracy theory that allows us to reject it.

Pryor anticipates this reply, and complains as follows:

Even if Sandra's theories were intrinsically incredible from the beginning – and even if they were inconsistent with our background evidence – I'm supposing that this wasn't initially obvious to you. Nor *should* it have been obvious to you. Instead, there was some evidence X that, together with your background evidence, brought you to the point of being able to justifiably reject one of her theories.

I will ignore, as I said at the beginning, Pryor's flirtations with cases where the hypothesis in question are inconsistent with what you know. So, what Pryor is claiming is that even if *not-H* but *everything tells in favor of H* is initially incredible, this shouldn't have been obvious to you, but rather something that you come to realize with the help of *everything tells in favor of H*.<sup>5</sup> I take it that this means that you start out rationally having some high-ish confidence in *not-H* but *everything tells in favor of H*. One immediate puzzle is then the following: what is Pryor admitting when he admits that *not-H* but *everything tells in favor of H* might nevertheless be initially incredible? Incredible to whom? More importantly: if you do rationally have some high-ish confidence in *not-H* but *everything tells in favor of H*, how can *everything tells in favor of H* justify you in believing *H*?

But let us grant for the sake of argument that it is indeed the case that you rationally have some high-ish confidence in *not-H* but *everything tells in favor of H* (remember that *everything tells in favor of H* is just my name for the mass of evidence that you present in favor of *H*). How could this confidence be rationally shaken by acquiring the information that *everything tells in favor of H* is true? There are two possibilities here. One possibility is that when you learn that *everything tells in favor of H* is true you realize that your confidence in *not-H* but *everything tells in favor of H* should have been low all along, even before acquiring the evidence. If so, this is more akin to a conversion than to a rational accommodation of evidence. More importantly, it doesn't go against Entailment. What you would be doing is to retroactively lower your previous confidence in *not-H* but *everything tells in favor of H*, and then raise it a tiny bit, leaving it low enough to justifiedly reject that hypothesis.

The second possibility, which must be what Pryor has in mind, is that learning that *everything tells in favor of H* is true rationally lowers your confidence in *not-H* but *everything tells in favor of H*. This position, however, clashes head to head with the arguments for Entailment that I presented above. If Pryor's verdict about his case were *obvious*, that would constitute a good argument against Entailment nevertheless

(we would then have to scramble to see what to say about the previous arguments for Entailment). Alas, I do not find Pryor's verdict obvious.

I conclude, therefore, that Pryor's case does not after all constitute a counterexample to Entailment.

## Notes

- 1 Pryor mentions Huemer (2001) as someone who states, but doesn't endorse, the same argument that I present in my essay. I greatly admire Huemer's paper, but there are differences between Huemer's argument and my own, and further differences as to what to do with the argument. For instance, Huemer shows only an intermittent grasp of the importance of No Lemmas, flirting with the Kleinian position I criticize in my paper. Huemer also favors a kind of contrastivist rejection of Closure that I don't agree with, and conflates defeasible with inductive justification.
- 2 In line with the proposal at the end of my paper to take a closer look at life without Inductivism, I would urge the reader to consider the possibility that we have evidence that entails that Sandra's theories are false. In that case, Entailment no longer applies (for Sandra's theories can entail our evidence only at the cost of being inconsistent) and we are free to reject Sandra's theories on the basis of that evidence. In what follows, however, I ignore this line of thought and consider what a more traditional answer would look like.
- 3 Pryor has Sandra build ever more revisions in the face of new evidence. I don't think that going beyond two theories matters.
- 4 In what follows I talk in terms of degreed doxastic notions, such as low or high opinions and levels of confidence. This makes exposition easier, but I believe that the points I'm making survive translation to a coarse-grained approach to the doxastic attitudes.
- 5 Again, Pryor seems to think that something hangs on further iterating the process, but I don't see that it does.

## References

- Comesaña, J. and Sartorio, C. (forthcoming) Difference-making in epistemology. *Noûs*.  
Huemer, M. (2001) The problem of defeasible justification. *Erkenntnis* 54: 375–397.  
Pryor, J. (2012) When warrant transmits. In *Wittgenstein, Epistemology and Mind: Themes from the Philosophy of Crispin Wright*, ed. A. Coliva, pp. 269–303. Oxford: Oxford University Press.

## Chapter Ten

# Can Belief Be Justified Through Coherence Alone?

Catherine Elgin and James Van Cleve both answer this question negatively. But whereas Van Cleve advocates a moderate version of foundationalism, Elgin defends a broadly coherentist view. According to her, justification is primarily a matter of explanatory coherence. The justification an individual belief enjoys is derived from the coherence of the overall system. In his essay, Van Cleve argues that, although coherence is indeed a source of justification, it cannot by itself render a belief completely justified. According to Van Cleve, no belief could be justified unless it were possible for some beliefs to acquire complete justification without receiving support from any other beliefs. In their respective responses, Elgin and Van Cleve continue the dispute, focusing on issues such as conjunction closure, corroboration by independent witnesses, empirical generalizations, revisability, and the skeptical threat of being deluded.

## Non-foundationalist Epistemology: Holism, Coherence, and Tenability

*Catherine Z. Elgin*

Much epistemology assumes that cognitive success consists in knowledge, where knowledge is justified or reliable true belief. On this conception, since propositions are the contents of beliefs and the bearers of truth values, they are what is known. If this is right, the sort of justification of interest to epistemology seems to be the justification of individual propositions. A linear model of justification is almost inevitable. To justify a given proposition is either to infer it from already justified propositions or to show how belief in it emerges from reliable belief forming mechanisms. *S* is justified in believing *p* on the basis of *q*, and *q* on the basis of *r*, and so on. Holists contend that this picture is misleading. They maintain that epistemic acceptability is, in the first

instance, acceptability of a fairly comprehensive system of thought, comprised of mutually supportive commitments. The priority in question is epistemological, not historical. There is no contention that people come to believe a theory before coming to believe the various claims that comprise it. The point is that regardless of the order in which they are acquired, claims are justified only when they coalesce to constitute a tenable system of thought. The acceptability of individual sentences, as well as methods and standards, is derivative, stemming from their role in a tenable system.

The challenge for such an epistemology is to explain how systematic interconnections give rise to justification, how the fact that deliverances dovetail with each other affords reason to believe they are true. Some philosophers hold that the coherence of a sufficiently comprehensive constellation of claims makes them true (Blanshard, 1939; Rescher, 1973). This strikes me as implausible, but I will not argue against it here. The position I want to investigate is that coherence is the source of epistemic justification, not the ground of truth. But if truth is independent of what we believe, why should mutual accord among our beliefs be indicative of truth? What is the connection? To avoid begging questions, it is perhaps better to begin by focusing not on the justification of beliefs, but on the justification of deliverances, these being representations that present themselves as candidates for belief. If we are concerned with justification, we should not limit ourselves to assessing the status of what we actually believe, but ask which of the things that could in given circumstances be believed should in those circumstances be believed. Deliverances, as I use the term, include perceptual inputs, fixed or transient beliefs, passing thoughts, and so forth.

Perhaps things will become clearer if we consider a case. Yesterday Meg's Latin book was stolen from her locker. Three students may have witnessed the theft. None of them is very reliable. Anne is given to proving theorems in her head, and tends to be oblivious to her surroundings when preoccupied with a tricky proof. To compensate for her habitual distractedness, she draws plausible inferences about mundane events, and often does not notice whether her opinion is due to observation or to such an inference. Ben frequently forgets to wear his glasses. Like Anne, he draws plausible inferences about events around him, and tends not to remember having done so. Chauncy is simply a liar. Presumably he knows when he is speaking sincerely, but given the fluency and frequency of his lies, nothing he says is trustworthy. Not surprisingly, the social circles of the three students do not intersect; none would deign to speak to the others. When questioned about the theft, Anne and Ben report what they think they saw, but confess that they are not sure what they actually witnessed and what they inferred. Chauncy insists that his report is accurate, but in view of his record, his claim is suspect.

Individually, none of the reports would count for much. Had only one of the witnesses been present, the most we could reasonably conclude would be that the thief might fit the description. But all three reports agree, and agree in alleging that the thief had an unusual appearance: he had spiked green hair. This makes a difference. Even though individually each report is dubious, and the probability of a green haired textbook thief is low, the fact that the three reports provide the same antecedently improbable description inclines us to believe it. Their accord evidently enhances the epistemic standing of the individual reports (Lewis, 1946, p. 346). We seem to have more reason to believe each of them in light of the others than we have to believe them separately. The question is: why? How can multiple statements, none of which is

tenable, conjoin to yield a tenable conclusion? How can their relation to other less than tenable claims enhance their tenability?

Given the unreliability of the witnesses, we might expect them to be wrong about the thief. But we would not expect them to all be wrong in the same way. The fact that they agree needs an explanation. If they were in cahoots, the explanation would be straightforward: they conspired to tell the same tale. But not being on speaking terms, they are probably not co-conspirators. If the description they provided fit a relevant stereotype, then a penchant for plausibility could explain their accord. But green spiked hair is far from any stereotype one might harbor for a textbook thief. So despite Anne's and Ben's propensity to draw inferences based on plausibility, their descriptions of the thief do not seem to result from such an inference. Evidently the best explanation of the agreement is that the reports are true.

It is not just our ability to exclude obvious alternatives that leads us to credit the allegation. A variety of collateral considerations support it. Some bear directly on the content of the claim. Dan dimly recalls seeing an odd looking stranger lurking in the hallway. The custodian thinks he saw a container of hair dye in the trash. Although the tentativeness of these reports makes them less than wholly credible, they are suggestive enough to buttress the eyewitness testimony. Other collateral considerations concern the witnesses and their circumstances. Book thefts are observable events, so there is nothing inherently dubious about a claim to have seen someone steal a book. The light and the sight lines were such that the witnesses could have seen what they report. The witnesses are adept at recognizing furtive adolescent behavior. None was subject to psychological experiments with implanted memories. None was on drugs. And so on. Separately, these factors count for little. Either their credibility is low or their bearing is slight. But they weave together to make a solid case. This suggests that the epistemic tenability of the several reports and the conclusion they sanction derives from their mutual supportiveness.

Although our focus is on the status of the allegation, it is the account as a whole that is or is not acceptable. Many of the relations of justification are reciprocal. The allegation is acceptable only if (at least most of) the rest of the constellation of supporting considerations is. But since the eyewitnesses are unreliable and the contentions of the collateral witnesses are tenuous, the acceptability of the testimony likewise depends on the acceptability of the allegation. The epistemic status of the allegation is inseparable from the status of the rest of the story. Some of the background information may be separately secured, but to a considerable extent, the various components of the story stand or fall together.

The thesis of the sort of epistemological holism that I want to consider is that epistemic justification is primarily a property of a suitably comprehensive, coherent account, when the best explanation of coherence is that the account is at least roughly true. The epistemic justification of individual claims derives from their membership in a justified account. There is no universally accepted criterion of coherence. But at least this is required: the components of a coherent account must be mutually consistent, cotenable and supportive. That is, the components must be reasonable in light of one another. Since both cotenability and supportiveness are matters of degree, coherence is too. So if it can be shown that epistemic justification is a matter of coherence, there remains the question of how coherent an account must be in order for it to be epistemically justified. Before facing that worry, though, other challenges need to be met. At least two worries

immediately arise. The first is that coherence is too demanding an epistemic requirement. The second is that it is not demanding enough.

Even where we take ourselves to be on solid ground, contravening considerations are not uncommon. [Mrs.] Abercrombie, the aging geometry teacher, says that during the relevant period she saw a young man sporting a green hat. A green hat is not green hair, so her report conflicts with the reports of the other witnesses. Ms. Mintz, the hall monitor, insists no one was in the corridor at the time of the alleged theft. Mr. Miller, the classics teacher, disputes the allegation on the grounds that students do not want Latin books enough to steal them. These reports are clearly relevant to and at odds with the account I gave. If we incorporate them into my account, we render it incoherent. But we seem to have no legitimate reason to exclude them. The problem is this: the discussion so far suggests that the credibility of the various claims comprising an account depends on how well they hang together. If so, the failure of other, equally relevant information to cohere threatens to discredit the account.

Although true, this is not so daunting as it appears. The immediate threat of incoherence comes from assuming that we must take seemingly contravening considerations at face value and incorporate them into an account as they stand. But we need do no such thing. Rather, we assess contravening considerations just as we do the rest of our evidence. Recall that we did not take the eyewitness reports at face value. We initially deemed them suspect because our background information indicated that the informants are unreliable. The credibility of the reports increased because of their agreement with one another and the support provided by collateral information. That agreement gave us reason to think that the general unreliability of the witnesses did not affect the standing of these particular reports. Contravening considerations are subject to similar assessments. Mrs. Abercrombie, being near-sighted and woefully out of date, cannot even imagine that a green thatch on someone's head might be his hair. That being so, her characterization of the suspect as wearing a green hat seems close enough to count as supporting rather than undermining the original allegation. Although Ms. Mintz flatly disputes what others have said, there are reasons to doubt that her claim is true. Since the three eyewitnesses saw each other in the corridor during the period when Ms. Mintz denies that anyone was there, her contention is dubious on independent grounds. Since she occasionally goes AWOL to smoke a cigarette, there is reason to suspect that she was absent when the theft occurred. Mr. Miller's argument cannot be so easily discredited. But the book is gone. Meg put it in her locker when she arrived at school. It was not there when she returned. Even if Latin books are not attractive targets for teenage thieves, the book's having been stolen may better explain its absence than any available alternative would. Just as other considerations compensate for the improbability of a green haired thief, other considerations compensate for the improbability of a Latin book thief. In determining the acceptability of a claim, we assess the considerations that afford evidence pertaining to its tenability. This is not always a simple yes/no matter. We may find that although an evidence statement is unacceptable or unsupportive as it stands, with suitable modifications, it would be. And we may find that the modifications themselves are acceptable. Coherence remains crucial. Sometimes it is achieved directly, sometimes by discrediting or disarming threats.

The coherence that affords epistemic justification is not just coherence among object-level deliverances. We have higher-order commitments about what sorts of object-level

deliverances are trustworthy, about how much credibility to accord them, about how they ought to mesh, and about what to do when commitments clash. These higher-order commitments supply reasons to revise or reject some deliverances but not others when conflicts occur. The coherence that constitutes epistemic justification is something we achieve, not something that simply falls out of the relations in which our object-level deliverances happen to stand to one another.

The second worry is that coherence can readily be achieved through epistemically illicit means. A good nineteenth-century novel is highly coherent, but not credible on that account. Even though *Middlemarch* is far more coherent than our regrettably fragmentary and disjointed views about the book theft, the best explanation of its coherence lies in the novelist's craft, not in the truth (or approximate truth) of the story. The coherence of the story affords virtually no reason to think it is true. This is surely right. But rather than taking this objection to completely discredit the contention that coherence conduces to epistemic acceptability, I suggest that it indicates something different: coherence conduces to epistemic acceptability only when the best explanation of the coherence of a constellation of claims is that they are (at least roughly) true.

Although epistemology generally focuses on the beliefs of a single individual, I began with a public case because the otherwise unlikely agreement of independent witnesses clearly shows how the best explanation of the coherence of a given body of claims may be that they are (at least roughly) true. The case of a single individual can be trickier. Sometimes people confabulate. They compose a coherent narrative by ignoring, bracketing, or overlooking factors that detract from the story they seek to construct. The process may be unconscious. Obviously, when a subject is confabulating, the coherence of her beliefs is not explained by their truth. If it is hard to tell whether she is confabulating, it is hard to tell whether coherence confers epistemic standing on her beliefs. But to understand how, why, and when coherence engenders credibility, it is best to put this complication aside. Then we see that the story I have told could be told of a single epistemic agent as well. If the best explanation of the coherence of an agent's system of thought is that it is at least roughly true, and she has no overriding reason to think otherwise, she is justified. Anne is aware of what she thinks she saw, and what she thinks the other witnesses report. She is privy to the relevant background information about apparent sight lines and the like. Since her various relevant cognitive commitments mesh and the best explanation of their meshing is that they are at least roughly true, according to epistemological holists, she is justified in accepting them.

One might argue that even the best nineteenth-century novel does not pose as great a threat as we sometimes suppose. No matter how deeply immersed I am in the story, a single glance up from the page is enough to convince me that I am not in a drawing room in nineteenth-century England. The story, though internally coherent, manifestly fails to mesh with the rest of my experience. This is true, but the question is what to make of it. On the one hand, too restricted a cluster of mutually supportive claims seems inadequate to engender credibility. We can't make the story credible simply by ignoring everything else we believe. On the other hand, insisting that all our commitments need to cohere seems unduly demanding. If acceptability requires coherence with everything we accept (or with everything we accept for cognitive purposes; Lehrer, 1986), it is but a short step to skepticism. One wayward belief, however remote from current concerns, could discredit an entire constellation of beliefs. Theories that ground justification in coherence then face a problem of scope.



Worries about scope, however, seem not to do justice to the problem that confronts us here. Faced with a clash between the deliverances of the novel and those of my glance, it is obvious which I should accept. There is no temptation to resolve the tension by dismissing perceptual deliverances or taking *them* to be the fiction. They seem to possess an epistemic privilege that prevents considerations of coherence from overriding them. The capacity of perceptual deliverances to trump the claims of a tightly knit novel may seem conclusively to demonstrate that epistemological justification cannot consist in coherence.

The matter deserves further consideration though. Until the source of perception's epistemic privilege is clear, it is premature to rule coherence out. A variety of reasons have been offered. Foundationalists argue on a priori grounds that knowledge requires that there be some independently credible beliefs. They hold that perceptual deliverances are among the independently credible beliefs because perceptual deliverances derive at least some of their warrant from the circumstances in which they occur, not their relation to other deliverances. Exactly how credible they are is a matter of dispute (BonJour, 1985, pp. 26–30). But they must, foundationalists contend, have some measure of credibility that does not derive from their accord with other convictions. Reliabilists argue that a deliverance is epistemically acceptable if produced by a reliable mechanism. Some perceptual mechanisms are reliable, hence some perceptual deliverances are acceptable. Since the reliability of perceptual mechanisms is independent of the relations of their deliverances to other deliverances, perceptual deliverances are independently credible.

There are at least two separate insights here. The reliabilist argument targets the need for a link to the world. The reason for crediting the casual glance while dismissing the deliverances of the novel is that we take it that perception provides the link. The way the world is constrains our perceptual deliverances more immediately and directly than it does our other beliefs. Insofar as the contents of knowledge claims concern the way the world is, it makes sense that the constraints the world supplies should override other considerations. The foundationalist position underscores the idea that some deliverances – in particular, those of perception – seem at least *prima facie* credible independently of their connections to other beliefs.

What the objections show is that if perception is to provide the sort of check on theorizing that we think it should, egalitarianism vis-à-vis object-level deliverances will not do. An egalitarian theory would hold that each deliverance has an equal claim on our epistemic allegiance. On the principle of one man, one vote, there is no basis for privileging some deliverances over others. If a perceptual deliverance fails to cohere with an otherwise coherent theory, the perceptual deliverance ought to be rejected then, since the claims of the many outweigh the claims of the one. But no matter how comprehensive and integrated an empirical account is, no matter how many other beliefs the account manages to incorporate, observations should have the capacity to discredit it. They have that capacity only if the epistemic claims of perceptual deliverances at least sometimes outweigh those of theory. But it does not follow that perceptual deliverances must be utterly immune to revision or rejection on the basis of considerations of coherence. Nor does it follow that the epistemic privilege granted to perceptual deliverances is independent of coherence considerations.

If we think about our situation when we glance away from the novel, we recognize that we draw on more than the sentences comprising the novel and our current perceptual

deliverances. We tacitly rely on a fairly extensive and epistemologically informed understanding of novels and perception. We know enough about underlying mechanisms to have reason to credit some perceptual deliverances. We know enough about literature to realize that novels are typically literally false. That constitutes sufficient reason for even casual perceptual deliverances to override the claims of the novel.

Juxtaposing the novel with perception might seem to make the problem too easy, though. Regardless of what we think about perception, if we recognize that a novel is a work of fiction, we have reason to discount any direct claims it may seem to make on our epistemic allegiance. (I say direct claims because I believe that novels play a significant, albeit indirect, role in the advancement of understanding. But how they do so is not germane to this discussion (Elgin, 1996, pp. 183–200).) The serious challenge comes from a coherent factual account that conflicts with perceptual deliverances. If holism holds that such an account always overrides perceptual deliverances, it seems plainly unacceptable. However tightly woven an empirical account may be, we would be epistemically irresponsible to ignore recalcitrant evidence. Foundationalists take this latter point to be decisive: if observation can show a theory to be unjustified, then coherence cannot be the locus of justification.

This would be so, if observation worked in isolation. For then, owing to its epistemic privilege, one perceptual deliverance would have the capacity to discredit an entire system of thought. But this is a myth. Only observations we have reason to trust have the power to unseat theories. So it is not an observation in isolation, but an observation backed by reasons that actually discredits the theory.

The holist response to the challenge presented by observation is this: *a priori*, perceptual deliverances have no special weight. They are just deliverances jockeying for inclusion in coherent bodies of thought. But over time, as we attend to the fates of our various deliverances, we learn that the incorporation of some, but not others, yields accounts which are borne out by further experience, hence which retain their coherence over time. This gives us grounds for discrimination. We realize that the deliverances we take to be perceptual are more likely to be confirmed than spontaneous deliverances that just leap to mind. So we assign greater weight to perceptual deliverances than to passing thoughts. Moreover, we learn that not all perceptual deliverances are on a par. Those that are credible tend to come in mutually reinforcing streams, so isolated perceptual deliverances count for little. We begin to draw distinctions among perceptual deliverances. For example, we discover that peripheral vision is less trustworthy than central vision. So we have reason to discount what we see out of the corner of the eye. This is not to say that we dismiss the deliverances of peripheral vision out of hand, but that we demand more in the way of corroboration. Some of us discover that we are color blind or tone deaf or myopic. That is, we learn that our perceptions of colors, tones, or the dimensions of distant objects are not to be trusted. And so on. We come to assign different weights to perceptual deliverances depending on how well they accord with other things we take ourselves to have reason to credit – other appearances of the same object, the reports of other observers, the implications of our best theories about the visible properties of items of the kind in question, and so forth.

The issue is not simply how well a given content meshes with other things we believe, but how well a given content from a given source in given circumstances does. The weight we attach to perceptual deliverances derives from our understanding of the

world and our access to it. Initially, perhaps, this is just a matter of track records. Some perceptual deliverances seem to integrate better into acceptable systems of thought than spontaneous thoughts that just leap to mind. Later, as we develop physiological and psychological accounts of ourselves, which explain our perceptual mechanisms, we gain additional reasons to take some perceptual deliverances to be credible. The epistemic privilege that some perceptual deliverances enjoy then derives from an understanding of ourselves as perceiving organisms. That is, the reason for assigning those deliverances significant epistemic weight derives from the coherent account of perception that backs the assignment. Contrary to what foundationalists contend, the justification for privileging perception derives from the relation of perceptual judgments to the rest of our theory of ourselves as cognitive agents interacting with a mind-independent world.

The reliabilist account seems to fare slightly better. What justifies assigning my visual inputs significant epistemic weight seems to be that vision is a reliable perceptual mechanism. What justifies dismissing my forebodings is that premonition is not. This is not quite right though. It is not the brute reliability or unreliability of a source that supplies the justification, but an understanding of that reliability or unreliability. Even if my forebodings are accurate, so long as we have no reason to trust them, they bear little weight.

This argument explains both why some perceptual deliverances have the capacity to unsettle theory, and why those deliverances are not intrinsically privileged. They owe their epistemic status to their place in our evolving understanding of the world and our modes of access to it. This has two welcome consequences. The first is that the privilege they enjoy is revocable. When I learn that I am color blind, I need to revise my views about which of my visual deliverances are acceptable. The second is that non-perceptual deliverances can in principle be equally weighty. This is an advantage in accounting for the epistemic status of scientific evidence and of testimony.

A look at modern science shows that it is not just (or perhaps even mainly) bare perceptual deliverances that have the capacity to discredit theory. The outputs of measuring devices do too. In an effort to retain a tie to classical empiricism, some philosophers of science argue that measuring devices are simply extensions of our senses. Just as eyeglasses enable near-sighted people to see what otherwise they could not, telescopes and microscopes enable everyone to see what otherwise we could not. So if seeing something in suitable circumstances has sufficient weight to undermine a coherent cluster of claims, seeing something through a telescope or microscope should be able to do so too. This idea is not unreasonable so long as we restrict ourselves to devices like optical telescopes and microscopes. But it stretches the bounds of plausibility to contend that radio telescopes, electron microscopes, MRIs, and the like are also mere extenders of the sense of sight. It seems better to forgo the strained analogy and simply characterize such devices as detectors. Then an understanding of what they detect, how they detect, and why they should be trusted supplies reason to accord their outputs considerable weight. Even without the strained analogy, the argument for crediting the outputs of scientific instruments thus parallels the argument for crediting perceptual deliverances. For although they are not perceptual mechanisms, the devices are among our modes of access to the world.

Testimony poses a similar problem. We acquire many of our beliefs from the testimony of others, and consider those beliefs justified. Some philosophers say that

the justification for accepting testimony is *a priori*. *Ceteris paribus*, we are justified in accepting what people tell us. Others say it is inductive. We should believe only those who have shown themselves to be relevantly reliable in the past. The former seem to endorse gullibility, the latter to unduly limit acceptability. Something more sensitive is wanted. Evidently the question is not whether testimony *per se* is or is not *prima facie* acceptable. Some testimony is frankly incredible; some requires a good deal of corroboration; some is straightforwardly acceptable. The acceptability of a bit of testimony depends on how well its content coheres with other relevant deliverances, how well the belief that the testifier is competent with respect to her allegation coheres, and how well the belief that she is sincere coheres. Because of its meshing with our background beliefs, straightforwardly acceptable testimony scores high on all of these measures. Just as different perceptual deliverances are accorded different weights, so are different testimonial deliverances. Testimony with sufficiently strong backing can discredit a hitherto coherent cluster of beliefs.

Even though the deliverances of perception, testimony, and instrumental readings have no special standing *a priori*, in light of our developing theories of the world and our modes of access to it, some of them turn out to have considerable epistemic weight. This satisfies the demand that acceptable beliefs be appropriately constrained by the way the world is. It also reveals that holism has the resources to recognize that deliverances can differ in weight, some being more credible than others. The claims of the few can in suitable circumstances outweigh the claims of the many.

Achieving coherence is not just a matter of excluding untoward deliverances though. In the interests of systematicity, we may incorporate considerations we have no antecedent reason to believe. For example, although there is no direct evidence of positrons, symmetry considerations show that a physical theory that eschewed them would be significantly less coherent than one that acknowledged them. So physics' commitment to positrons is epistemically appropriate. Considerations we have no independent reason to believe can acquire tenability then because they strengthen the coherence of the systems they belong to.

The issue of scope remains. The totality of a person's beliefs and/or deliverances is not particularly coherent. Not only are there outliers and inconsistencies among beliefs, there are also clusters of beliefs that are relatively isolated from one another. Meg's cluster of beliefs about the pituitary gland, the evidence that bears on the acceptability of these beliefs, the trustworthiness of bits of testimony on the subject, and the proper methods for assessing such things has few and loose connections to her cluster of views about parliamentary procedure, the evidence that bears on these views, the trustworthiness of testimony about the subject, and the proper methods for assessing them. It seems that she could easily be badly wrong about the former without her error having any significant effect on the tenability of her views about the latter. Outliers and inconsistencies among beliefs are in principle relatively unproblematic. According to a holism, outliers lack justification. Because they lack suitable connections to other things we believe, we have no reason to credit them. Inconsistencies among beliefs conclusively demonstrate that some of the beliefs are false. But it is not obvious that mutual indifference of belief clusters is objectionable. It is not clear that we should consider Meg epistemically defective because of the lack of close ties between the two clusters. On the other hand, if the clusters of beliefs are too small and too numerous, complacency over their mutual indifference seems problematic. We do not want to

license ignoring inconvenient tensions among beliefs by consigning them to mutually irrelevant clusters.

The problem neither has nor needs an *a priori* resolution. Our evolving theories of the world and our access to it provide us with an appreciation of the relations in which our various clusters of beliefs should stand to one another and the requirements they should satisfy. Such a *laissez-faire* attitude might seem to allow for the acceptability of crazy constellations of views. If we leave it to our evolving theories to decide what range of considerations acceptable accounts must answer to, we may be forced to endorse isolated islands of claptrap. The worry is more apparent than real. We have theories about theories, which enable us to assess the reasons, methods, standards, and evidence that our various object-level theories appeal to. Some requirements, such as logical consistency, apply globally. Regardless of how far apart Meg's views about politics and endocrinology are, unless they can be conjoined without contradiction they are not all acceptable. Other requirements, like the need to respect judicial precedents or to accord with biochemical findings, are more limited in range. But even these do not enable us to isolate belief clusters entirely. Even if Meg's views about endocrinology and politics have few points of contact, her views about endocrinology and hematology have many.

Consistency requirements do more than rule out express contradictions. The requirement that like cases be treated alike demands that if a consideration has weight in one area but not in another, there be an acceptable reason for the difference. In order to be tenable, a system of mutually reinforcing claims must either answer to the logical and evidential standards to which other theories are subject or be backed by a tenable account of why those standards do not apply. Some theories have such backing. There are, for example, cogent reasons why mathematics is not subject to empirical testing. So infinitary mathematics is not threatened by the absence of empirical evidence for its findings. In epistemically objectionable cases, no such reasons are available. The claims of astrology, although mutually reinforcing, are epistemically unacceptable because they yield predictions that are either too vague to be tested or are not borne out when tested. Since astrology makes empirical claims, there are considerations to which it ought to be responsive which it fails to accommodate. To say that something cannot be ruled out *a priori* is not to say that it cannot be definitely and decisively ruled out.

Epistemological positions that construe knowledge as justified true belief generally treat being justified, being true, and being believed as three separate features of a propositional content. The standard objection to coherentism is that coherence among propositional contents is so easily achieved that it affords no reason to believe that the contents are true, hence no justification for them. This overlooks the fact that the contents in question are not just any propositional contents, they are belief contents or deliverance contents. That is, they are contents that present themselves as true. This makes a difference. For the fact that they present themselves as true gives us some slight reason to think that they are true. The word "slight" is crucial. I do not contend that we have sufficient reason to credit such contents. But at least two considerations speak in favor of granting them a slight measure of credibility. Beliefs form the basis for action, so the success of our actions affords evidence of the truth of the corresponding beliefs. Moreover, we learn from experience. Once we come to recognize that premonitions tend not to be borne out, we cease to credit them. We may continue to experience feelings of foreboding, for example, but they cease to qualify as deliverances.

Manifestly, these considerations are far too weak to demonstrate that beliefs or deliverances are epistemically justified. They do, however, give us reason to think that beliefs and deliverances have some claim on our epistemic allegiance. They have an epistemic edge. We have better reason to incorporate them into our systems of thought than to incorporate contents we are neutral about. Beliefs and deliverances are, I suggest, initially tenable. But initial tenability is a weak and precarious epistemic status. Considerations of overall coherence often require revision or rejection of initially tenable commitments. Initially tenable commitments can conflict. They may be mutually incompatible or non-cotenable. Or they may be sufficiently isolated that they are incapable of giving support to or gaining support from other things we believe. Then they cannot be incorporated into an epistemically acceptable system.

Epistemic acceptability, I contend, requires reflective equilibrium (Elgin, 1996; Rawls, 1971). A system of thought is in equilibrium if its elements are reasonable in light of one another. This is a matter of coherence. An equilibrium is reflective if the system is as reasonable as any available alternative in light of our initially tenable commitments. Such a system is not required to incorporate as many initially tenable commitments as possible. As we have seen, there are weighting factors that favor some incorporations over others. Moreover, rather than incorporating commitments, a system may show why we were misled into accepting them, or may include modifications of them.

The standards of reasonableness are second-order commitments, and are subject to the same sorts of considerations as our first-order deliverances. The fact that we accept them indicates that they are *prima facie* acceptable. But they can conflict, or fail to yield verdicts in cases where they should, or yield verdicts that we find unacceptable. Then they too are subject to revision or rejection in order to yield a comprehensive system of first- and second-order commitments that is on reflection something we can endorse.

Whether the sort of holism that results is a coherence theory is not clear. Using BonJour's (1985) categories, it might be classified as a very weak foundationalism or as a coherence theory. Deliverances derive their initial tenability from their status as deliverances. That suggests that something other than coherence is involved. But initially tenable commitments display at least two features that are not characteristic of standard foundational beliefs. First, there are no intrinsically privileged kinds of deliverances. The account does not insist that there is something epistemically special about perception or introspection or analyticity. It simply says that the fact that a consideration presents itself as true gives it a modest measure of tenability. Second, even that small measure of tenability is easily lost. Tenable theories are justified in part by reference to initially tenable deliverances, but they need not incorporate the deliverances by reference to which they are justified.

Whether we call such an epistemology a coherence theory does not in the end matter. The virtues of the theory are as follows. (i) It does not privilege any sorts of beliefs or representations *a priori*. What beliefs and representations are worthy of acceptance is something we learn by developing increasingly comprehensive, coherent accounts of the world and our access to it. (ii) It enables us to start from whatever deliverances we happen to have. But because it insists that we subject those deliverances to rigorous assessment, such a starting point is not question-begging. (iii) The standards of assessment are themselves the fruits of epistemic activity, and can change in response to feedback (Goodman, 1984, p. 69). (iv) Hence, everything is subject to revision. A system

of thought that we can on reflection accept today may be one that we cannot on reflection accept tomorrow. But so long as a system is in reflective equilibrium and the best of explanation of its being so is that it is at least roughly true, it and its components are justified. What results is neither certainty nor skepticism but a fallible, provisional, but reasonable epistemological stance.

## References

- Blanshard, B. (1939) *The Nature of Thought*. London: George Allen & Unwin.  
BonJour, L. (1985) *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.  
Elgin, C. (1996) *Considered Judgment*. Princeton, NJ: Princeton University Press.  
Goodman, N. (1984) *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.  
Lehrer, K. (1986) The coherence theory of knowledge. *Philosophical Topics* 14: 5–25.  
Lewis, C.I. (1946) *An Analysis of Knowledge and Valuation*. La Salle, IL: Open Court.  
Rawls, J. (1971) *A Theory of Justice*. Cambridge, MA: Harvard University Press.  
Rescher, N. (1973) *The Coherence Theory of Truth*. Oxford: Clarendon Press.

## Further Reading

- Adler, J. (1986) Knowing, betting, and cohering. *Philosophical Topics* 14: 243–257.  
Bender, J., ed. (1989) *The Current State of the Coherence Theory*. Dordrecht: Kluwer.  
Goodman, N. and Elgin, C. (1988) *Reconceptions*. Indianapolis: Hackett.  
Harman, G. (1973) *Thought*. Princeton, NJ: Princeton University Press.  
Lehrer, K. (1974) *Knowledge*. Oxford: Clarendon Press.  
Sellars, W. (1963) *Science, Perception and Reality*. London: Routledge & Kegan Paul.  
Sellars, W. (1968) *Science and Metaphysics*. London: Routledge & Kegan Paul.  
Sosa, E. (1980) The raft and the pyramid. *Midwest Studies in Philosophy* 5: 3–26.  
Sosa, E. (1985) The coherence of virtue and the virtue of coherence. *Synthese* 64: 3–28.  
Williams, M. (1980) Coherence, justification, and truth. *Review of Metaphysics* 37: 243–272.

# Why Coherence Is Not Enough: A Defense of Moderate Foundationalism

*James Van Cleve*

## I

Foundationalism has been characterized as the view that “the knowledge which a person has at any time is a structure or edifice, many parts and stages of which help to support each other, but which as a whole is supported by its own foundation” (Chisholm, 1964). To unpack the metaphor, we may say that the foundation consists of

*basic beliefs* – beliefs that a subject is justified in holding even in the absence of any justifying reason for them – and all other justified beliefs derive their justification at least in part from such basic beliefs.

The classical argument for foundationalism is an infinite regress argument going back to Aristotle: unless there were basic beliefs, every justified belief would rest on further justified beliefs, and so on without end. An infinite regress is not in fact the only alternative to basic beliefs, but the other alternatives are equally problematic, or so foundationalists maintain. To canvass all the options, let us set down four propositions that jointly imply the existence of an infinite regress of justified beliefs:

- 1 Some beliefs are justified.
- 2 No belief is justified unless some other belief serves as a reason for it.
- 3 One belief cannot serve as a reason justifying another unless the first is itself justified.
- 4 If A serves as a reason justifying B, then B cannot serve (directly or indirectly) as a reason justifying A.

These four propositions jointly entail the existence of an infinite regress of justified propositions. We are therefore faced with five alternatives: accept the infinite regress or reject one of the four assumptions.

*Skeptics* (of the universal ilk) deny 1, maintaining that no beliefs whatever are justified. *Foundationalists* deny 2, maintaining that some beliefs are justified in the absence of reasons. *Positists* (not to be confused with positivists) deny 3, maintaining that chains of justifying reasons can terminate in reasons that are not justified themselves, but are simply individual or societal posits.<sup>1</sup> *Coherentists* deny 4, maintaining that beliefs can be justified in virtue of relations of mutual support. *Infinatists* accept all four assumptions and the resulting infinite regress.

All five options have their takers. Skepticism and infinitism are both defended elsewhere in this volume (see the essays by Fumerton and Klein). Positism finds advocacy in Wittgenstein's *On Certainty* and in assorted postmodern thinkers (Wittgenstein, 1969). But I think it is fair to say that the leading contenders among the five options are foundationalism and coherentism.

Coherentism is sometimes characterized as a view that sanctions circular reasoning, but that is an oversimplified construal of it. Coherentists do not typically endorse simple loops in which A justifies B, B justifies C, and C justifies A; rather, they envision vast webs of belief in which everything is supported by some significant portion of the remaining beliefs: A by B and C, B by A, D, J, and K, and so on.

For their part, foundationalists do not typically deny the power of coherence to contribute to the overall epistemic status of a body of belief. They simply insist that coherence cannot do all the work on its own – there must be at least a modicum of intrinsic credibility or non-inferential warrant possessed by basic beliefs before coherence can have its amplifying effect.

Laurence Bonjour distinguishes three grades of foundationalism (Bonjour, 1985, pp. 26–30). According to *strong* foundationalism, basic beliefs are “not just adequately justified, but also *infallible, certain, indubitable, or incorrigible*” (Bonjour, 1985, pp. 26–27). According to *moderate* foundationalism, the non-inferential warrant possessed by basic beliefs need not amount to absolute certainty or any of the other



privileged statuses just mentioned, but it must be “sufficient by itself to satisfy the adequate-justification condition for knowledge” (BonJour, 1985, p. 26). Finally, according to *weak* foundationalism,

basic beliefs possess only a very low degree of epistemic justification on their own, a degree of justification insufficient by itself either to satisfy the adequate-justification condition for knowledge or to qualify them as acceptable justifying premises for further beliefs. Such beliefs are only “initially credible,” rather than fully justified. (BonJour, 1985, p. 28)

We must rely on coherence among such initially credible beliefs to amplify their level of warrant up to the point where it is adequate for knowledge.

As BonJour notes, weak foundationalism could be regarded as a hybrid view, mixing together foundational and coherentist elements. In fact, Susan Haack prefers to call it “foundherentism,” which she illustrates with the example of a crossword puzzle. Experience corresponds to the clues, which give an initial presumption in favor of certain beliefs or entries in the puzzle; the initial beliefs are then confirmed by the way in which they interlock with other entries (or sometimes discarded because they do not fit in). Thus does coherence amplify (or its absence erode) the initial warrant possessed by basic beliefs (Haack, 1993).<sup>2</sup>

But if coherence can elevate the epistemic status of a set of beliefs in this way, what prevents it from generating warrant entirely on its own, without any need for basic beliefs? This is a question that has been asked by several authors, including BonJour:

The basic idea is that an initially low degree of justification can somehow be magnified or amplified by coherence, to a degree adequate for knowledge. But how is this magnification or amplification supposed to work? How can coherence, not itself an independent source of justification on a foundationalist view, justify the rejection of some initially credible beliefs and enhance the justification of others? (BonJour, 1985, p. 29)

The implied suggestion is that if coherence can do what the weak foundationalist allows, it can also do what the thoroughgoing coherentist says it can do.

In the next section, I address this challenge. I argue that there is indeed a good rationale for the weak foundationalist’s insistence that before coherence can do its work, there must be initially credible inputs for it to work upon. But I also argue that the level of initial credibility cannot be as low as that envisioned by weak foundationalists, who should therefore upgrade their position to moderate foundationalism. Finally, in section III, I offer some critical reflections on the coherentist alternative defended by Catherine Elgin.

## II

An excellent example of a weak foundationalist theory is provided by C.I. Lewis’s theory of memory knowledge. It has two elements:

First; whatever is remembered, whether as explicit recollection or merely in the form of our sense of the past, is *prima facie* credible because so remembered. And second; when the whole range of empirical beliefs is taken into account, all of them more or less dependent

upon memorial knowledge, we find that those which are most credible can be assured by their mutual support, or as we shall put it, by their *congruence*. (Lewis, 1946, p. 334)<sup>3</sup>

Lewis defines a congruent set as one in which any member is more probable given the rest than it is on its own:

A set of statements, or a set of supposed facts asserted, will be said to be congruent if and only if they are so related that the antecedent probability of any one of them will be increased if the remainder of the set can be assumed as premises. (AKV, p. 338)

A point on which Lewis repeatedly insists is that congruence alone cannot generate probability or warrant. Rather, some of the statements must have initial credibility, which congruence can then amplify:

The feature of such corroboration through congruence that should impress us, is the requirement that the items exhibiting these congruent relationships must – some of them at least – be *independently given facts* or have a probability which is antecedent. (AKV, p. 352)

How *much* probability must the congruent items have? Only a “slight” amount, Lewis tells us, illustrating his point with the example of individually unreliable witnesses who tell the same story:

Our previous example [AKV, p. 239] of the relatively unreliable witnesses who independently tell the same circumstantial story, is another illustration of the logic of congruence; and one which is more closely typical of the importance of relations of congruence for determination of empirical truth in general. For any of these reports, taken singly, the extent to which it confirms what is reported may be slight. And antecedently, the probability of what is reported may also be small. But congruence of the reports establishes a high probability of what they agree upon, by principles of probability determination which are familiar: on any other hypothesis than that of truth telling, this agreement is highly unlikely. (AKV, p. 346)

But how much probability is a “slight” amount? In one place Lewis says, “Anything sensed as past is just a little more probable than that which is incompatible with what is remembered and that with respect to which memory is blank” (AKV, p. 358). He thereby implies that a remembered piece of information has a probability greater than 0.5, given that it is remembered. (This is because it is an axiom of the probability calculus that  $P(\sim h, e) = 1 - P(h, e)$ ; thus a proposition has greater probability than its negation if and only if it has probability greater than 0.5.) Thus according to Lewis’s version of weak foundationalism, the congruence of a set of remembered items can raise their level of justification arbitrarily high, but only if the items have initial credibility amounting to a probability (given that we seem to remember them) greater than 0.5.

To this latter aspect of Lewis’s theory BonJour has raised an objection. There is no need, he says, for Lewis’s requirement that memory reports or other cognitive deliverances have initial credibility:

What Lewis does not see, however, is that his own example shows quite convincingly that no antecedent degree of warrant or credibility is required. For as long as we are confident

that the reports of the various witnesses are genuinely independent of each other, a high enough degree of coherence among them will eventually dictate the hypothesis of truth telling as the only available explanation of their agreement – even, indeed, if those individually reports initially have a high degree of *negative* credibility, that is, are much more likely to be false than true (for example, in the case where all of the witnesses are known to be habitual liars). (BonJour, 1985, pp. 147–148)

We are now presented with a clear-cut issue to investigate: in order for the congruence of a set of items to raise their credibility to near 1, what level of antecedent credibility is required? Must it be greater than that of their negations and thus greater than 0.5, as Lewis maintains? Or may it be less than 0.5, as BonJour implies when he says that the reports of the witnesses (or of our memory) may be more likely false than true?

For light on this question, we may look to discussion of a traditional topic in probability theory: how to assess the probability that independent witnesses who agree in their testimony are telling the truth. This is a problem to which a number of classical authors have proposed answers. One of the standard answers is due to George Boole (1952, p. 364):<sup>4</sup>

Let  $p$  be the general probability that  $A$  speaks the truth,  $q$  the general probability that  $B$  speaks the truth; it is required to find the probability that, if they agree in a statement, they both speak the truth. Now agreement in the same statement implies that they either both speak truth, the probability of which beforehand is  $pq$ , or that they both speak falsehood, the probability of which beforehand is  $(1 - p)(1 - q)$ . Hence the probability beforehand that they will agree is  $pq(1 - p)(1 - q)$  and the probability that if they agree, they will agree in speaking the truth is accordingly expressed by the formula

$$w \left[ = P(A \text{ and } B \text{ speak truly, they agree}) \right] = \frac{pq}{pq + (1 - p)(1 - q)}$$

For an explanation of the rationale behind Boole's formula, I must refer the reader to what I have said elsewhere.<sup>5</sup> Here there is space only to note the bearing of his formula on the issue separating Lewis and BonJour. Suppose that  $A$  and  $B$  each tell the truth 60 percent of the time; that is, suppose that  $p$  and  $q$  are each equal to 0.6. The reader may verify that in this case,  $w = 0.69$ . That is, setting  $A$ 's credibility and  $B$ 's each equal to 0.6, the probability that  $X$  is true given that they each testify to it is 0.69. More generally, if we plug in any numbers greater than 0.5 as  $p$  and  $q$ ,  $w$  will be greater than the mean of  $p$  and  $q$ . So far we have an illustration of the point, common ground for Lewis and BonJour, that congruence can boost credibility.

But what if we plug in values of  $p$  and  $q$  equal to or less than 0.5? If 0.5 goes in, 0.5 comes out; and if  $p$  and  $q$  are each less than 0.5, the output value will be less than their mean. For example, if  $p$  and  $q$  are each 0.1,  $w$  is approximately 0.01. In other words, if the witnesses have what BonJour calls "negative credibility" (credibility less than 0.5), the probability of a statement given that they both testify to it is not enhanced but diminished! So if Boole's formula is correct, Lewis is vindicated and BonJour refuted: with initial credibilities less than 0.5, coherence makes things worse rather than better.

There is reason to be suspicious of Boole's formula, however. In deriving it, he tacitly assumes that there are only two possible answers to the questions put to the witnesses – true or false. That is why he can equate agreement with "both speak truly

or both speak falsely.” In a more realistic scenario, the witnesses would be asked multiple-choice questions, and if they agreed in giving the same answer out of ten possible choices (let us say), their agreement would be much more impressive.

A formula allowing for an arbitrary number of possible answers to questions has been devised by Michael Huemer (1997).<sup>6</sup> For simplicity’s sake, Huemer assumes that the witnesses have the same level of credibility, so  $p = q$ . If  $n$  is the number of possible answers and  $X$  is the answer on which independent witnesses  $A$  and  $B$  agree, then Huemer’s formula may be written as follows (omitting details of the derivation):

$$P(X, A \text{ says } X \text{ \& } B \text{ says } X) [= W] = \frac{np^2 - p^2}{np^2 - 2p + 1}$$

This formula agrees with Boole’s in the special case where  $n = 2$ , but gives dramatically different results in the cases not covered by Boole’s. Specifically, it enables coherence to have its amplifying affect even when the credibility level of the witnesses is below 0.5, just so long as it is greater than the chance or random guessing level of  $1/n$ . For example, if  $p$  is only 0.3, but there are ten possible answers (say, ten possible last digits in a glimpsed license plate number), then  $w = 0.62$ . If  $p = 0.3$  and  $n = 100$ , then  $w = 0.86$ . For any value of  $p$ , just so long as it is greater than zero, we can bring the final probability of  $X$  (i.e., its probability given that the witnesses agree on it) as close as we like to certainty by making  $n$  high enough. With a more complex version of the formula, we can also make  $w$  higher by increasing the number of witnesses.

If Huemer’s formula is correct, then, BonJour is vindicated and Lewis refuted. Initial credibilities need not be greater than 0.5; they need only be greater than the chance level of  $1/n$  (where  $n$  = the number of possible answers). If enough witnesses agree without collusion in giving the same answer from among a large enough number of choices, it becomes overwhelmingly likely that they are correct, even if their initial level of credibility was scarcely above zero.

Is BonJour right, then? Can coherence alone be a source of warrant without need of inputs with initial credibility? An answer of yes would be too hasty, for there is another requirement of initial credibility we have yet to consider. The requirement at which we have so far demurred is the requirement that what is reported must be more probable than not (and thus have probability greater than 0.5) given that a witness (or an ostensible memory) attests to it. If several individually unreliable reporters agree without collusion, then the fact to which they bear common witness may have high probability in the end. But in attaching a high final probability to the fact attested, we are of course taking for granted that the various witnesses *do* testify to it. If we had reason to think that the courtroom and all its proceedings were happening only in a dream or a novel, the fact that the ostensible reports hang together would count for little. And so it is with the reports of memory, the senses, and cognitive systems more generally: coherence among them lends high final credibility only on the assumption that the reports genuinely occur.

How, then, do we know *these* things: that witness  $A$  does say  $X$ , that I do ostensibly remember  $Y$ , that I do seem to see  $Z$ ? Many foundationalists would say that these are the grounds on which the rest of our knowledge rests, and that they must themselves be matters of basic knowledge.<sup>7</sup> Lewis himself famously maintained that nothing can be probable unless something is certain, and among the certainties he

placed the facts that I do have this or that presentation of sense or memory. His insistence on certainty is controversial,<sup>8</sup> but it seems to me that a good case can be made that there must at least be high intrinsic credibility – perhaps high enough to constitute knowledge – attaching to the facts that such-and-such cognitive states (be they experiences, ostensible memories, or beliefs at large) are actually taking place. If this is right, we must not only abjure pure coherentism: we must also adopt a moderate rather than a weak foundationalism.

I see only one plausible alternative to an assumption of high initial credibility or knowledge-sufficient warrant at the foundational level, and that is the view that the promptings of sense or memory function as *external* conditions of knowledge. An external condition of knowledge is a condition that makes knowledge possible regardless of whether it is itself known to obtain. For example, in Goldman's reliability theory, if a subject comes to believe *p* as the result of a reliable process, his belief is knowledge regardless of whether the subject knows anything about the reliability of the process (Goldman, 1979). Perhaps the facts that I have such-and-such ostensible perceptions or memories could function in this external way, contributing to my knowledge even if not themselves known. The idea would be that my ostensible perceivings and rememberings are not pieces of evidence on which I conditionalize when their epistemic status is high enough;<sup>9</sup> instead, they are facts whose mere obtaining confers credibility on their contents.

I turn now to another point at which I think a theory of knowledge that invokes coherence must make a concession to either foundationalism or externalism.

The example of the witnesses who agree is in one respect a drastically oversimplified case of coherence. The agreement of the witnesses is literal identity, or at least logical equivalence, of content: witness 1 says *X* and so does witness 2. But the coherence that figures in epistemology is typically a much looser sort of hanging together. The coherence of ostensible memories is not their all being memories that *p*, for the same *p* or something logically equivalent. Nor is the coherence of beliefs or cognitions generally like that. Rather, it is a type of coherence that is exemplified by the following items:

I seem to remember seeing a skunk last night;  
I seem to remember smelling a skunk last night;  
I seem to remember that the lid was on the garbage can when I went to bed;  
I now see that the can has been knocked over and trash strewn about;  
There was a skunk here last night;

and so on. In other words, it is not identity or even equivalence of content, but rather something like the relation Lewis calls congruence: a matter of each item being more probable given the rest than it is on its own.

What are these coherence-constituting relations of probability founded upon, and how do we know that they obtain?

One answer has been given by Russell: "It is only by assuming laws that one fact can make another probable or improbable" (Russell, 1948, p. 188). Perhaps Russell goes too far in requiring strict laws in order for one fact to make another probable, but it is plausible that we at least require rough empirical generalizations. Where do these generalizations come from? Presumably, they are inferred inductively from particular facts gathered by memory. And now the following difficulty emerges: ostensible

memories give rise to knowledge only with the help of coherence; coherence depends on laws or empirical generalizations; and such generalizations can be known only with the help of memory. In short, we cannot get coherence without the help of laws, and if memory does not suffice on its own to give knowledge of particular facts from which the laws are inferred, we cannot get laws without the help of coherence. It appears to follow that we cannot have any knowledge from memory unless the occurrence of ostensible memories is *prima facie* sufficient for knowledge. Such was Russell's own conclusion: 'memory is a premise of knowledge. ... When I say memory is a premise, I mean that among the facts upon which scientific laws are based, some are admitted solely because they are remembered' (Russell, 1948, pp. 188–189). Note the word "solely." Russell is saying that individual memories must be capable of giving rise to knowledge on their own, without benefit of coherence. This is compatible, of course, with allowing that the warrant provided by memory is defeasible, as Russell did allow. But the resulting view is nonetheless a foundationalism of memory knowledge stronger than that of Lewis, who required only an initial "slight" presumption in favor of the truth of any ostensible memory. Russell's view accords to memory greater epistemic powers than that: ostensibly remembering that *p* is a source of *prima facie* warrant that, if undefeated (and if *p* is true) is strong enough for knowing that *p*. In BonJour's terms, we have again advanced from weak to moderate foundationalism, this time as regards the contents of ostensible memories rather than the occurrences of them as mental events.

Russell's argument assumes that coherence and the laws that underlie it contribute to our knowledge only if they are themselves known. As in the case of the occurrences of our ostensible memories, one could challenge this assumption by going external, holding that coherence does its work regardless of whether the subject knows it obtains. This is the second point at which I believe coherentism can avoid a concession to foundationalism only by making a concession to externalism. The fact that *p*, *q*, and *r* do cohere with one another, as well as the facts that they are deliverances of our cognitive systems to begin with, are facts that must either function externally or be known foundationally.

### III

As the debate between foundationalists and coherentists has progressed, each side has moved in the direction of the other. Contemporary foundationalists are seldom foundationalists of the strong Cartesian variety: they do not insist that basic beliefs be absolutely certain. They also typically allow that the elements in a system of belief can acquire enhanced justification through their coherence. On the other side, many coherentists admit that coherence alone is not the sole source of justification – there must be some initially credible inputs before coherence can work its wonders. Is there anything more to disagree about, or do foundationalists and coherentists now meet in the middle?

There are indeed still points of difference. To highlight several of them, I shall discuss the broadly coherentist views of Catherine Elgin as developed in her book *Considered Judgment* (1996; cited as CJ hereafter) and in her contribution to this volume.

Elgin characterizes herself as a proponent of reflective equilibrium.<sup>10</sup> As she conceives of it, reflective equilibrium has two chief requirements: “The components of a system in reflective equilibrium must be reasonable in light of one another, and the system as a whole reasonable in light of our initially tenable commitments.” (CJ, p. 107; see also pp. ix, 13, and 127–128). It is by the second requirement that Elgin distinguishes her view from a pure coherentism: the components of a system in equilibrium must be answerable not just to one another, but also to our initially tenable commitments.<sup>11</sup>

The second requirement puts a “tether” on permissible systems (CJ, pp. 10, 107, 128), thereby enabling Elgin to avoid some of the objections to pure coherentism. For example, one of the standard objections to coherentism is that the contents of a consistent fairy tale would be a body of warranted propositions (see Schlick, 1973, p. 419). Not so for Elgin, since the propositions in the story may be reasonable in light of each other without being reasonable in light of our initially tenable commitments.

Of the various things we believe, which have initial tenability? In Elgin’s view, they *all* do, if we genuinely believe them: anything actually held has some initial tenability or presumption in its favor (CJ, pp. 101–102). The presumption may only be slight and it may be lost in the end, but it is there in the beginning.

What about the various principles of logic, evidence, and method whereby some things are reasonable in light of others? In Elgin’s view, these have the same status as everything else: they are initially tenable if held, and they may gain or lose in tenability depending on how they fit in with everything else (CJ, p. 104).

Why is Elgin’s view as so far set forth not simply a form of weak foundationalism, in which initially tenable claims function as basic beliefs? She cites two differences: unlike the justification that attaches to foundationalism’s basic claims, initial tenability can be lost; it can also be augmented through coherence (CJ, p. 110).

It is not clear to me, however, that either of these features should be regarded as a prerogative of coherentists alone. In a typology of possible foundationalisms, Roderick Firth has suggested that the minimal tenet of foundationalism is simply this: basic beliefs have some measure of initial warrant that is not derived from coherence. This level of warrant may be increased by coherence with other statements or diminished, even to the vanishing point, by lack of coherence with other statements.<sup>12</sup> Firth’s minimal view thus incorporates both of the features Elgin sees as antithetical to foundationalism.<sup>13</sup>

Nonetheless, I see two other questions on which foundationalists are apt to disagree with Elgin. First, are all commitments initially tenable, or only those in some specially marked out class? Second, are all commitments likewise revisable, or are some immune from subsequent rejection? On each question, Elgin takes the more egalitarian stand (CJ, pp. 101–102 and 121).

Using terminology from Michael Huemer, we may say that the first issue is the issue of phenomenal conservatism versus a more general doxastic conservatism (Huemer, 2001, pp. 99–115). Phenomenal conservatism is the view that if anything seems to be the case, one is *prima facie* justified in accepting it. The seemings can include perceptual seemings (it looks to me as if there is a red object over there), memorial seemings (I seem to remember being chased by a dog one day on my way to kindergarten), and intellectual seemings (it strikes me as self-evident that the relation of equality is transitive). Doxastic conservatism is the more sweepingly democratic view that anything the subject

believes has some presumption in its favor – the products of wishful thinking and superstition no less than the deliverances of perception and memory. Foundation-  
alists are typically phenomenal conservatives, while Elgin (in her book) is a doxastic conservative.

In her contribution to this volume, Elgin no longer espouses doxastic conservatism, or at any rate holds that coherentists need not be committed to it. She says that perceptual deliverances may be assigned special weight, provided they do not have it a priori. They have it only in virtue of coherence considerations: by accepting the deliverances of perception, we get systems that remain coherent over time (that is, they continue to jibe with new deliverances of perception). But might it not likewise be true that by accepting fantasies, we get systems that remain coherent with future contents of fantasy – especially if the fantasizer has a one-track mind? So it seems to me that there is a privileging of perception presupposed and not explained by this coherentist underpinning for it.

I turn now to the second point: are all commitments on a par as regards the possibility of revision? Elgin holds, with Quine, that in the quest for reflective equilibrium, anything may be revised. There is no commitment that may not be sacrificed in order to maximize the tenability of the entire system. Here I would like to protest that there are certain principles of logic, at least, that cannot be given up, because they are framework principles without which coherence could scarcely be defined. What would happen if we gave up the law of non-contradiction? It is not clear that there could any longer be such a thing as what Quine calls a recalcitrant experience, forcing changes elsewhere in the system. If a new deliverance stood in contradiction to things we already accepted, we could simply accept it with a “What – me worry?” shrug.

Here is another difficulty for a coherentism that holds everything revisable, at least if we understand this as “anything could be justifiably rejected,” symbolized as “(p) possibly  $J \sim p$ .” Suppose there is some proposition  $q$  (the law of non-contradiction, perhaps?) whose truth is necessary for anything to belong to a coherent system and therefore necessary for anything to be justified. Since  $\sim q$  entails that nothing is justified, we now have *possibly  $J$  (nothing is justified)*, which is absurd.<sup>14</sup>

A related question about the status of the rules and principles of logic and evidence is whether they have force only because a subject is committed to them. I gather Elgin would say yes, but I say no. Consider a system of beliefs containing the elements  $p$ ,  $q$ , and  $p \rightarrow \sim q$ . Suppose it is not a sheer fact of logic, independent of anything the subject believes, that the system is inconsistent and in need of revision. Could we make the system intolerably inconsistent by adding the principle: if any two of  $\{p, q, \text{ and } p \rightarrow \sim q\}$  are true, then the third must be false? No, for anyone who could live comfortably with the original system could live comfortably with the expanded system. Such is the lesson the tortoise taught Achilles (Carroll, 1895).

I wish to raise one more question about the role of logical and evidential relations in Elgin’s coherentism. Elgin defines both coherence and equilibrium in terms of the relation “ $p$  is reasonable in light of  $q$ ,  $r$ , and  $s$ .” What is the required epistemic status of this relation (or of the logical and other relations on which it supervenes)? Must such relations be known to hold among the propositions in a system before the propositions are warranted for the subject? And if so, how does such knowledge arise?

I see three possibilities to consider in answer to these questions. First, the holding of coherence-constituting relations might be regarded as an external condition of



knowledge, making knowledge possible regardless of whether the subject knows that such relations hold. Although this seems to me a good way for a coherentist to go, I gather that Elgin would not find it congenial. On more than one occasion, she expresses her dissatisfaction with externalist stratagems in epistemology (CJ, pp. 22, 46, 51; this volume, p. 251). Second, it might be held that logical relations and other relations of support are known to hold because they are necessary relations, apprehendable a priori. But this, too, is an option Elgin would reject. In the first place, it would be a concession to foundationalism; in the second place, she has Quinean qualms about there being any propositions at all that are true necessarily and known a priori (CJ, pp. 40–46 and 53–57). Third, it might be held that coherence-constituting relations are known to hold because the propositions saying that they hold are part of the best overall coherent system that is reasonable in light of our antecedent commitments.<sup>15</sup> But that way lies an infinite regress. A proposition *p* affirming a relation of coherence would be justified only because the subject is justified in believing that *p* belongs to a coherent system. That belief in turn would be justified only because the subject is justified in believing that *p belongs to a coherent system* belongs to a coherent system, and so on. Even if it is the same system every step of the way, we still get a regress in which ever more complicated propositions must be believed with justification to belong to the system. The untenability of such a regress suggests to me that we should go instead with one of the first two options, agreeing with the externalist that coherence propositions need not be known at all or with the foundationalist that they are known because they are either basic propositions or propositions inferable from basic propositions.

## IV

I have not lived up to the title of this essay, for I have not offered a complete defense of moderate foundationalism. I hope nonetheless to have shown that an internalist coherentism cannot be a satisfactory theory of justification. We must be either externalists or foundationalists, and if we are foundationalists, our foundationalism must be of the moderate rather than the weak variety.

## Notes

- 1 The distinction between positivism and foundationalism is lost on those who cannot hear the word “justified” as anything but a past participle, implying that some act or relation of justifying has occurred whereby a belief is justified by something else that serves as a reason for it. For foundationalists, “justified” simply connotes a favorable epistemic status, which a belief may have even though the subject has no reason for it. In this connection, another term, such as “evident” or “credible,” might be less misleading than “justified.”
- 2 Haack summarizes her theory in Haack (2000). She does not classify foundationalism as weak foundationalism, believing it essential to foundationalism that the foundations not receive support from other elements in the structure.
- 3 Hereafter I shall cite this work in the text as AKV.
- 4 This paper was first published in 1857.
- 5 James Van Cleve, “Can coherence generate warrant?” (2011).

- 6 I discuss Huemer's work at some length in "Can coherence generate warrant?"
- 7 For foundationalists who hold that propositions about the physical world are always derived, not basic, that A says X would not be basic after all. It would rest, however, on deliverances that are basic.
- 8 For further discussion of Lewis's view, see Van Cleve (1977). For a method of assigning probabilities in relation to evidence without assuming that the evidence is certain (in the sense of having probability 1), see Jeffrey (1983).
- 9 By "conditionalizing" I mean drawing the inference "h has probability n given evidence e, and e is certainly true; therefore, h has probability n."
- 10 The term was coined by John Rawls (1971, p. 20f.). It refers to a state of affairs in which specific judgments and general principles have been brought into agreement with each other through a process of mutual adjustment.
- 11 Elsewhere Elgin adds other requirements. One is that there must not be a competing system that is more tenable overall (Elgin, 1996, p. 107). Another is that the best explanation of the system's coherence must be that it is at least roughly true ("Non-foundationalist Epistemology: Holism, Coherence, and Tenability," this volume).
- 12 See Firth (1964, pp. 466–467) for his progressively weaker foundationalisms.
- 13 Even so staunch a foundationalist as Roderick Chisholm incorporates Elgin's two allegedly distinctive features to some extent. Although the warrant belonging to Chisholmian "self-presenting" propositions can be neither increased nor diminished by their relations to other propositions, the same is not true of the warrant possessed by "indirectly evident" propositions. Their epistemic status is defeasible, and it may be raised by concurrence (Chisholm's term for coherence). See Chisholm (1977, chapters 2 and 4).
- 14 In fairness to Elgin, I note that she may give special status to the law of non-contradiction. She says that tenable systems must at least be logically consistent, or else be all-entailing (CJ, p. 136). Against this rationale, however, I note that in revisionary relevance logic, contradictions do not entail everything.
- 15 Perhaps this is what coherentists should say in response to Russell's question about how the laws underlying coherence are known.

## References

- BonJour, L. (1985) *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.
- Boole, G. (1952) On the application of the theory of probabilities to the question of the combination of testimonies or judgments. In *Studies in Logic and Probability*. London: C.A. Watts.
- Carroll, L. (1895) What the tortoise said to Achilles. *Mind* 4: 275–280.
- Chisholm, R.M. (1964) The myth of the given. In *Philosophy*, ed. R.M. Chisholm *et al.* Englewood Cliffs, NJ: Prentice Hall.
- Chisholm, R.M. (1977) *Theory of Knowledge*, 2nd edn. Englewood Cliffs, NJ: Prentice Hall.
- Elgin, C. (1996) *Considered Judgment*. Princeton, NJ: Princeton University Press.
- Firth, R. (1964) Coherence, certainty, and epistemic priority. *Journal of Philosophy* 61: 545–557. Reprinted in *Empirical Knowledge*, ed. R. Chisholm and R.J. Swartz. Englewood Cliffs, NJ: Prentice Hall, 1973.
- Goldman, A. (1979) What is justified belief? In *Justification and Knowledge*, ed. G. Pappas. Dordrecht: Reidel.
- Haack, S. (1993) *Evidence and Inquiry*. Oxford: Blackwell.
- Haack, S. (2000) A foundherentist theory of empirical justification. In *Epistemology*, ed. E. Sosa and J. Kim. Oxford: Blackwell.

- Huemer, M. (1997) Probability and coherence justification. *Southern Journal of Philosophy* 35: 463–472.
- Huemer, M. (2001) *Skepticism and the Veil of Perception*. Lanham, MD: Rowman & Littlefield.
- Jeffrey, R. (1983) *The Logic of Decision*, 2nd edn. Chicago: University of Chicago Press.
- Lewis, C.I. (1946) *An Analysis of Knowledge and Valuation*. LaSalle, IL: Open Court.
- Rawls, J. (1971) *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Russell, B. (1948) *Human Knowledge: Its Scope and Limits*. New York: Simon & Schuster.
- Schlick, M. (1973) The foundation of knowledge. In *Empirical Knowledge*, ed. R. Chisholm and R.J. Swartz. Englewood Cliffs, NJ: Prentice Hall.
- Van Cleve, J. (1977) Probability and certainty: a re-examination of the Lewis–Reichenbach debate. *Philosophical Studies* 32: 323–334.
- Van Cleve, J. (2011) Can coherence generate warrant *ex nihilo*? Probability and the logic of concurring witnesses. *Philosophy and Phenomenological Research* 82: 337–380.
- Wittgenstein, L. (1969) *On Certainty*. Oxford: Blackwell.

## Reply to Van Cleve

*Catherine Z. Elgin*

Foundationalists consider some beliefs basic. A subject is justified in holding them “even in the absence of any justifying reason for them” (p. 256).<sup>1</sup> Other beliefs are justified by being suitably related to basic beliefs. Ostensible memories, beliefs about one’s current experiences, and the fruits of introspection standardly count as basic. Strong foundationalists consider basic beliefs certain. Moderate foundationalists hold that “the non-inferential warrant possessed by basic beliefs need not amount to absolute certainty ... but it must be ‘sufficient by itself to satisfy the adequate-justification condition for knowledge’ (BonJour, 1985, p. 26)’ (p. 257). Weak foundationalists set the level of credibility considerably lower. Despite its title, James Van Cleve’s paper is not so much a defense of moderate foundationalism as an attack on what he takes to be its chief rivals – coherentism and weak foundationalism. Playing offense rather than defense is a reasonable argumentative strategy; but his failure to defend the position may suggest that moderate foundationalism is epistemologically unproblematic. I will argue that this is not so, and defend coherentism and/or weak foundationalism against Van Cleve’s objections.

Each component in a coherent system supports and is supported by many others. Foundationalists acknowledge that coherence amplifies credibility, but insist that amplification can occur only where credibility is there to be amplified. If they are right, some measure of initial credibility is needed. Their only question is: How much? An unamplified sound retains its original volume. If coherence functions as a credibility amplifier, then a belief that fails to cohere with other beliefs should retain whatever level of credibility it began with. But failure to cohere deprives claims of credibility. Perhaps “amplification” is a bad metaphor. Still, one wants to know why failure to cohere discredits a claim.

Beliefs satisfying the adequate justification condition can have a probability of less than 1. If knowledge is closed under obvious implication, a subject who knows that *p*

and knows that  $q$  is evidentially in a position to know that  $p \& q$ . But if  $p$  and  $q$  are evidentially independent and each has a probability of less than 1, the probability of the conjunction is less than the probability of each conjunct. The more we conjoin, the lower the probability, eventually falling below the threshold for knowledge. Moderate foundationalism thus has reason to disvalue comprehensiveness. Granted, more complicated relations among propositions can enhance probability. But the closure principle for conjunction, which seems obvious and benign, is threatened.<sup>2</sup>

Moderate foundationalism might preserve closure by assigning probability 1 to the contents of all beliefs that satisfy the adequate justification standard. Then there is no doubt to aggregate. Adler (2002, pp. 250–255) argues that such a move is acceptable if we distinguish between credibility and confidence, and recognize that we can be less than fully confident in a belief to which we assign probability 1. Another alternative is to deny that the probability calculus provides the metric for credence. Whatever solution moderate foundationalists choose, they should acknowledge that their position has untoward consequences.

Weak foundationalism maintains that deliverances – items that present themselves as candidates for belief – have a slight measure of initial credibility, and that credibility is enhanced through their incorporation into a coherent system of beliefs. Coherentism denies that initial credibility is needed. I construed my position as a very weak form of foundationalism, since I was marking out the epistemological difference between deliverances and propositions a subject has no inclination to believe. On Van Cleve's characterization, my position is coherentist. Coherentists, he says, maintain “that beliefs can be justified in virtue of relations of mutual support” (p. 256). If we restrict the considerations we are evaluating to beliefs (or deliverances), we need not introduce initial credibility. Within the realm of beliefs and deliverances, coherence is enough. Inasmuch as Van Cleve argues against both coherentism and weak foundationalism, this is a point of clarification, not an objection, to what he says. But it is worth noting that the difference between coherentism and weak foundationalism turns on where we start.

Corroboration by independent witnesses can raise probability above the threshold for knowledge, even if the initial credibility of their testimony is slight, so long as there are enough witnesses and they have enough choices. This favors weak foundationalism. But, Van Cleve notes, Huemer's (1997) proof applies only to agreement on a single proposition. The sort of coherence we are interested in involves looser relations of support among distinct beliefs. So weak foundationalists cannot assume that Huemer has settled the debate in our favor. Perhaps an analogous theorem could be proved given a suitably rigorous description of the web of beliefs. Perhaps not. But if Huemer's result is relevant to the larger debate, it provides the weak foundationalism with grounds for optimism.

Van Cleve demurs. Even if corroboration vindicates testimony with low initial credibility, he says, we still have to know that the testimony actually takes place. We might after all be dreaming. “A good case can be made that there must be at least high intrinsic credibility – perhaps high enough to constitute knowledge – attaching to the facts that such-and-such cognitive states (be they experiences, ostensible memories, or beliefs at large) are actually taking place” (p. 261). I am not convinced. If the dream possibility is a skeptical challenge, it applies to Van Cleve's position as much as to mine. If not, then, as Descartes argued in Meditation V, we have reasons, albeit less than certain reasons, to believe that we are not dreaming.<sup>3</sup> The reasons need not be

more than initially credible; for coherence leverages weakly supported deliverances when the best explanation of their occurrence is that things are at least roughly as they seem. Even putatively basic beliefs may be far from sufficient for knowledge. I may be unsure whether I seem to hear a distant train whistle or only imagine that I hear it. If no one else on the platform reacts, no train arrives, and I recall my propensity for wishful thinking, I conclude that I imagined the sound; I did not even seem to hear it.

Van Cleve objects that coherence depends on empirical generalizations. Coherence cannot be what generates knowledge, since we need knowledge of particulars to justify those generalizations. But the tenability of a generalization does not rest on its etiology. A generalization is initially credible if it is a deliverance. If it is true and is suitably interwoven into a comprehensive web of beliefs, it is known. But it need not be already known to be a candidate for incorporation, and in the process of devising a system its credibility may both influence and be influenced by other things the subject is inclined to believe.

Although Quine contends that any commitment can be revised, he does not maintain that all are equally good candidates for revision. He advocates a principle of minimal mutilation: in revising, we should retain as many of our central beliefs as possible. But even central beliefs may require revision. “Inanimate objects are identical when their parts are identical” is a fundamental metaphysical principle. But if it holds universally, then “ $F=ma$ ” does not. Molecules in a viscous fluid move at different rates. In prototypical applications of “ $F=ma$ ,” forces act on objects like billiard balls that have a well-defined boundary. In viscous fluids, the “forces” on the “object” are effects on the momentum of molecules moving in and out of that “object.” Retaining “ $F=ma$ ” requires continually redefining what constitutes a single particle, letting different molecules comprise it at different times. Fluid mechanics preserves the law. Rather than insisting that all component molecules of a particle be the same from one instant to the next, they let the individual molecules come and go, but keep the average enclosed mass constant (Wilson, 2006, pp. 158–159). “ $F=ma$ ,” evidently, is so central a law of physics that scientists are willing to radically revise the criteria for the identity of a fluid particle over time in order to preserve it. Quine’s position explains why this is reasonable.

What about the law of non-contradiction? Could it ever be rational to revise it? Paraconsistent logicians say “yes” (Priest, 2002). They contend that the benefits of rejecting the law of non-contradiction outweigh the costs. I disagree, not because I think the law is unrevisable, but because I think that other revisions better accommodate the problem cases.

Even if the law of non-contradiction lies at the very center of the web of belief, so that it is always preferable to make revisions elsewhere in the system, the same does not hold of all the principles of classical logic, much less of the principles of evidence. The history of logic is rife with debates about which principles ought to be accepted. The law of excluded middle remains the subject of controversy. Evidential principles are even less secure. We endorse visual deliverances and reject premonitions not because the former are basic and the latter are not, but because visual deliverances have, and premonitions lack, impressive track record, and we have an acceptable theory of vision, but no clue how premonitions could connect to their subject matter. The support by an overarching, tenable system vindicates vision and discredits premonition.

Contra Van Cleve, fantasy poses no threat to coherentism. So long as the fantasizer realizes that he is fantasizing, the thoughts he entertains are not deliverances, hence

lack initial credibility. Confabulation is a more serious worry. A confabulator composes a coherent narrative by unconsciously ignoring, bracketing, or downplaying considerations that detract from the account he seeks to construct, and accepting unwarranted considerations that support it. Clearly a confabulated account has no claim to epistemic respectability. But self-deception undermines coherence (Adler, 2002, pp. 74–101). Confabulators achieve local coherence by sacrificing coherence across a broader range of beliefs. Suppose a father deceives himself into thinking that his daughter's dismal grades result from her teachers' failure to recognize her quirky brilliance. To sustain his belief, he overlooks factors he otherwise considers relevant to student performance – terrible study habits, unfinished homework, ignorance of the most basic facts about the subject matter. He violates his otherwise accepted methodological beliefs about how to judge such matters. He would not invoke such considerations to account for his paperboy's poor grades. He thus weakens the overall coherence of his belief system by carving out exceptions for a special case. Because self-deception weakens coherence, it does not undermine the contention that epistemic justification is grounded in coherence.

Delusions might seem to pose a stronger challenge, for they are more coherent than self-deceptive beliefs. A mental patient who believes he is Napoleon in exile interprets all his experiences in terms of his delusion. He takes nurses, doctors, and aides to be lackeys, courtiers, and guards; visitors to be loyal subjects; those who refuse to do his bidding to be traitors; those who tell him he is ill or mistaken in his beliefs to be part of the plot to prevent his retaking the throne. Unlike the self-deceptive father, he does not carve out special exceptions. He explains all his experiences in terms of the delusion. Whenever a tension occurs, he rejects the deliverance that conflicts with his identification with Napoleon. Although his beliefs are more coherent than the self-deceptive father's, his delusion would have to be extensive to achieve any reasonable level of coherence. He would, for example, have to believe that he and everyone around him was speaking French rather than English. He would have to believe that oil lamps or candles rather than electric lights were illuminating the area. He would have to believe that the vehicles he sees are horse-drawn carriages rather than cars and trucks. Still, there would be many everyday experiences for which he had no explanation – riding an elevator, watching television, or using a telephone. Such gaps deprive his worldview of the level of coherence that normal epistemic agents regularly achieve. Self-deception and delusion might seem to undermine coherentism. People, it seems, can manipulate themselves into believing nearly anything. But they cannot easily incorporate self-deceptive or delusive beliefs into a comprehensive, coherent belief system. Coherence, which might look like an unduly weak constraint, turns out to be a very strong one.<sup>4</sup>

I conclude that Van Cleve's objections to weak foundationalism and coherentism are not compelling. There is good reason to believe that through systematizing, we weave initially tenable deliverances into epistemically tenable webs of belief.

## **Acknowledgments**

I am indebted to Jonathan Adler and Samuel Elgin for useful comments on early drafts of this paper.

## Notes

- 1 Page numbers in parentheses refer to James Van Cleve's essay, "Why Coherence Is not Enough: A Defense of Moderate Foundationalism," this volume.
- 2 In discussing the lottery paradox, Kyburg (1997) argues that because of the aggregation of doubt, we should reject closure under conjunction. So perhaps the principle is not so benign as it looks.
- 3 I am grateful to Jonathan Adler for strengthening my argument here.
- 4 I am grateful to Jonathan Adler for helping me to articulate this point.

## References

- Adler, J. (2002) *Belief's Own Ethics*. Cambridge, MA: MIT Press.
- BonJour, L. (1985) *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.
- Huemer, M. (1997) Probability and coherence justification. *Southern Journal of Philosophy* 35: 463–472.
- Kyburg, H. (1997) The rule of adjunction and reasonable inference. *Journal of Philosophy* 94: 109–125.
- Priest, G. (2002) Paraconsistent logic. In *Handbook of Philosophical Logic*, 2nd edn, Vol. 6, ed. D. Gabbay and F. Guenther, pp. 287–393. Dordrecht: Kluwer.
- Wilson, M. (2006) *Wandering Significance*. Oxford: Clarendon Press.

# Reply to Elgin

*James Van Cleve*

I select five points for further debate.

*Should adequate justification be conjunctive?* Moderate foundationalists hold that basic beliefs need not be certain; they need only have a level of justification adequate for knowledge. Elgin notes as a potential problem for moderate foundationalism that it may have to renounce the conjunctivity of adequate justification. If we do epistemology within the framework of the probability calculus, to say that adequately justified beliefs need not be certain is to say that they need not have probability 1. It follows that a conjunction of adequately justified independent beliefs may fail to be adequately justified, since the probability of the conjunction will be given by multiplying several numbers less than 1, and the product may be less than whatever threshold is set for adequacy.

I agree with Elgin that adequacy should be conjunctive. She notes two ways for avoiding the result to the contrary: assign probability 1 to adequate justification after all, or deny that the probability calculus provides the metric for justification. I favor the latter alternative, for one and perhaps two reasons. First, one might construe adequate justification as a qualitative notion, not measurable by any number.<sup>1</sup> Second, even if one did assign a number to the level of justification adequate for knowledge,

one could very well deny that the laws of probability are the laws of rational belief. I cite two points of discrepancy. (i) In the probability calculus, any necessary truth (no matter how recondite and remote from our knowledge) gets assigned probability 1. By the same token, whatever is entailed by a proposition has a probability at least as high as what entails it. As often noted, these are utterly unrealistic as laws governing actual degrees of rational belief. Necessary truths may be far from certain in any epistemic sense, and the necessary consequences of the obvious may not be obvious. (ii) It is a theorem of the probability calculus that if  $e_1$  and  $e_2$  are pieces of evidence with probability 1, then for any proposition  $h$ ,  $P(h) = P(h, e_1) = P(h, e_2)$ . Let  $e_1$  = I am sensing redly,  $e_2$  = I am hearing F-ly, and  $h$  = there is a red object in front of me. Surely  $e_1$  lends a higher degree of rational support to  $h$  than  $e_2$  does, yet the conditional probability of  $h$  on either piece of evidence is the same.<sup>2</sup>

*Is everything in principle subject to revision?* Elgin holds with Quine that the answer is yes. I had suggested that the law of non-contradiction is an exception, being essential if recalcitrant experiences are to force revisions in the web of belief. Elgin cites the work of Priest, who has advocated giving up the law of non-contradiction. Yes, he is a *dialetheist*, someone who holds that some contradictions are true. But he is not a *trivialist*, someone who holds that everything, including all contradictions, is true. In paraconsistent logic, one may accept some contradictions without accepting them all (Priest and Tanaka, 2009). I suggest that it could never be rational to reject the statement *not all contradictions are true*.

*Must the facts of deliverance have high initial credibility?* Elgin (whom I should have classified as a weak foundationalist) maintains that deliverances need have only a small amount of initial credibility, much less than would be required for knowledge. I take it she is referring to the *contents* of deliverances – if I have an ostensible memory that  $p$  or seem to hear testimony that  $p$ ,  $p$  thereby acquires a slight amount of credibility for me. I had maintained that the *facts* of deliverance – that I *do* ostensibly remember that  $p$ , for instance – must have *high* credibility, perhaps even enough for knowledge. If there is not adequate justification for believing the deliverances are actually occurring, their coherence counts for little. Elgin replies that if I am raising the possibility of dreaming as a skeptical challenge, it threatens my own position. I say not so, for deliverances as I conceive of them are seemings to see, presentations of  $p$  as past, and so on, which (as Descartes notes in Meditation 2) can be known to be occurring dream or no dream.

*Are particulars epistemically prior to generals?* In developing a point of Russell's, I had presupposed a kind of epistemic priority: contingent general truths are known or adequately justified only if one first has adequate justification for particular propositions from which they may be inferred. Elgin does away with all such priority. Once particulars and generals are all there in the web of belief (however they got there), each has its modicum of initial credibility, and they all rise together to the level of knowledge when they are suitably coherent. Their coherence is explicated in part as their each being reasonable in light of the others. How is it known that a generalization is reasonable in light of its instances, or the instances reasonable in light of the generalization? Is this perhaps an a priori matter? Or is it an external condition of knowledge, making knowledge possible regardless of whether the subject has any justification for it? If neither of the foregoing, it is presumably a proposition that must already have been in the web alongside the particulars and the generals, getting its positive status



just as they do, some of it just by being delivered and more by cohering with the rest. But now we may inquire whence derives the reasonableness of it given its companions, and a regress looms.

*Are epistemic principles on a par with everything else in the web?* An epistemic principle is a general principle specifying the conditions under which various propositions have various levels of justification for a subject. Elgin says “yes” to our question, but I think the considerations of the previous paragraph show that some epistemic principles must have a different status, either being external conditions of knowledge or, if requiring to be known, being known a priori. Consider two of Elgin’s own epistemic principles – that any deliverance has a slight initial amount of credibility, and that a belief has a high amount of credibility if it is a member of a coherent set whose coherence is best explained by supposing its members true. If these are neither a priori nor external, they must have credibility for the subject that they derive by being members of a coherent set – a set whose members are reasonable in light of each other. If D is reasonable in light of A, B, and C, that is presumably because there is an epistemic principle to the effect that if S believes (with justification?) each of A, B, and C, then S is justified in believing D. If this principle is neither a priori nor external, the subject must be justified in believing it. Whence that justification? As before, a regress looms.

## Notes

- 1 This is the approach taken in Chisholm (1977). Of his five epistemic categories, only two are representable in the probability calculus, the top category (certainty) implying probability 1 and the bottom category (having some presumption in its favor) implying probability greater than 0.5. Being evident, the level of justification required for knowledge, is assigned no number.
- 2 So is the degree of confirmation of h on either piece of evidence, defining this either as the ratio  $P(h,e)/P(h)$  or the difference  $P(h,e) - P(h)$ .

## References

- Chisholm, R.M. (1977) *Theory of Knowledge*, 2nd edn. Englewood Cliffs, NJ: Prentice Hall.
- Priest, G. and Tanaka, K. (2009) Paraconsistent logic. In *The Stanford Encyclopedia of Philosophy* (Summer 2009 edn), ed. E.N. Zalta. <http://plato.stanford.edu/archives/sum2009/entries/logic-paraconsistent/> (last accessed March 14, 2013).

## Chapter Eleven

# Is Infinitism the Solution to the Regress Problem?

According to Peter Klein, the regress problem “concerns the ability of reasoning to increase the rational credibility of a questioned proposition.” A non-evident proposition is one about which there could be credible disagreement. Such propositions seem to be lacking in something important. How might we, through reasoning, supply some of what is lacking? Peter Klein argues that only infinitism provides a solution to this problem. A chain of reasons that can supply what’s lacking must be neither circular nor arbitrary. The result is that an appropriate chain of reasons must be non-repeating and infinite. This is infinitism. Carl Ginet argues that infinitism faces intuitive counterexamples involving simple perceptual and a priori beliefs, which seem acceptable in the absence of a non-repeating infinite chain of reasons. Such examples include the belief that there is a blue streak on the wall, and the belief that  $2+2=4$ . Ginet argues that infinitism entails skepticism about the justification, at least when it comes to humans’ beliefs. Ginet also argues that infinitism is conceptually problematic, on the grounds that it entails that inference alone can generate justification.

## Infinitism Is the Solution to the Regress Problem

*Peter Klein*

### **The Regress Problem**

The *locus classicus* of the regress problem is to be found in Sextus Empiricus’s *Outlines of Pyrrhonism*:

The later Skeptics hand down Five Modes leading to suspension, namely these: the first based on discrepancy, the second on the regress *ad infinitum*, the third on relativity, the

fourth on hypothesis, the fifth on circular reasoning. That based on discrepancy leads us to find that with regard to the object presented there has arisen both amongst ordinary people and amongst the philosophers an interminable conflict because of which we are unable either to choose a thing or reject it, and so fall back on suspension. The Mode based upon regress *ad infinitum* is that whereby we assert that the thing adduced as a proof of the matter proposed needs a further proof, and this again another, and so on *ad infinitum*, so that the consequence is suspension [of assent], as we possess no starting-point for our argument. The Mode based upon relativity ... is that whereby the object has such or such an appearance in relation to the subject judging and to the concomitant perceptions, but as to its real nature we suspend judgment. We have the Mode based upon hypothesis when the Dogmatists, being forced to recede *ad infinitum*, take as their starting-point something which they do not establish but claim to assume as granted simply and without demonstration. The Mode of circular reasoning is the form used when the proof itself which ought to establish the matter of inquiry requires confirmation derived from the matter; in this case, being unable to assume either in order to establish the other, we suspend judgment about both.<sup>1</sup>

Although the three alternative strategies for solving the regress will be the focus of this essay, a brief discussion of the two other modes will be useful in understanding what initiates the regress.

The Modes were recipes for avoiding *dogmatism*, that is, the disposition to assent to non-evident propositions when it is not settled whether they are true. One could locate such a non-evident proposition either by noting that there was credible disagreement about it or by merely recognizing that there could be credible disagreement. For in order to avoid epistemic hubris, the recognition that our epistemic peers could sincerely disagree with us about the truth of some proposition forces us to regard it as requiring reasons in order to rise to the desired level of credibility.

*The Regress Problem* can be put as follows: Which type of series of reasons and the account of warrant associated with it, if any, can increase the credibility of a non-evident proposition? Can a series with repeating propositions do so? Can one with a last member do so? Can one that is non-repeating and has no last member do so?

Foundationalists and coherentists typically address the trilemma in two steps. First, they cavalierly reject the infinitist option by alluding to our “finite mental capacity.” Second, they argue for one of the remaining options by disjunctive elimination.<sup>2</sup> We will consider the “finite mind” objection in due course. My point here is only that infinitism has been given a short shrift, if any shrift, by epistemologists.

The argument in this paper has three essential steps: *first*, I will argue that neither foundationalism nor coherentism can solve the regress problem; *second*, I will present an infinitist account of warrant and explain how reasoning in accord with it can solve the regress problem; *third*, I will argue that the best objections to infinitism fail.<sup>3</sup>

## **Step 1: Neither Foundationalism nor Coherentism Can Solve the Regress Problem**

The regress problem concerns the ability of reasoning to increase the rational credibility of a questioned proposition. It is not crucial what degree of credibility is at stake. The task is to produce an account of warrant, where “warrant” refers to the property possessed

by propositions or beliefs such that (i) true beliefs with that property are known<sup>4</sup> and (ii) reasoning in accordance with the dictates of that account increases our rational confidence in non-evident propositions.

My claim will be that neither foundationalism nor coherentism provides such an account of warrant. I will *not* be arguing that either account of warrant is incorrect. I will be arguing that neither account of warrant can provide a solution to the regress problem because neither account can be employed by a self-conscious practitioner to increase the rational credibility of a questioned proposition and, thus, a primary reason for adopting either foundationalism or coherentism has been eliminated.

Foundationalists and coherentists differ about the way in which warrant originates and is transferred. Varieties of foundationalism can be demarcated (i) by the features of basic propositions in which warrant arises, (ii) by the degree of warrant that arises initially, and (iii) by the rules of inference that transfer warrant. But all foundationalists think of warrant as arising autonomously in so-called basic propositions and being transferred to other propositions through permissible forms of inference.

Coherentists *could* think of warrant as transferring from one proposition to another. This form of coherentism, what I will call the “warrant-transfer” variety, holds that some proposition, *p*, transfers its warrant to another proposition which can, in turn, pass it to another proposition. Eventually the warrant is transferred back to *p*. This view endorses circular reasoning. Now, I don’t think any epistemologist explicitly advocated this view – although the critics of coherentism occasionally characterize it in this way, as did the Pyrrhonians and Aristotle.<sup>5</sup>

The second variety of coherentism – the form that *has* been advocated – is what I will call the “warrant-emergent” form. Warrant emerges from the structure of the mutually supporting propositions.<sup>6</sup> Warrant is *not* a property of a particular proposition except in the trivial sense that a proposition is warranted iff it is a member of such a set of mutually supporting propositions. As the set of propositions becomes increasingly comprehensive and the mutual support intensifies, the degree of warrant for each increases. This view eschews incestuous circular reasoning in which warrant is transferred from some ancestor proposition to its descendants and then back again to the ancestor.

Our question is whether the account of warrant underlying foundationalism or either of the two forms of coherentism provides a basis for a solution to the regress problem. Let us begin with foundationalism and imagine a dialogue between Fred, the Foundationalist, and Doris, the Doubter. (Fred/Doris could be subpersona if we are envisioning a Cartesian-style, *sotto voce* meditation). Fred asserts some proposition, say *p*. Doris says something – who knows what – that prompts Fred to believe that he had better have reason(s) for *p* in order to supply some missing credibility. So, Fred gives his reason, *r*<sub>1</sub>, for *p*. (*r*<sub>1</sub> could be a conjunction.) Now, Doris asks why *r*<sub>1</sub> is true. Fred gives another reason, *r*<sub>2</sub>. This goes on for a while until Fred (being a practicing foundationalist) arrives at what he takes to be a basic proposition, say *b*.

Doris will, of course, ask Fred for his reason for *b*. But Fred, being a self-conscious, circumspect foundationalist will tell Doris that *b* doesn’t need a reason in order to possess the autonomous bit of warrant. He will say that her question “Why do you believe that *x*?” though appropriate up to this point is no longer appropriate when “*b*” is substituted for “*x*” because *b* is basic. There is no reason that supplies the *autonomous* warrant that *b* has.<sup>7</sup>

Grant that foundationalism is true; b has some autonomous bit of warrant that arises because b has some foundational property, F, such that any proposition having F is autonomously warranted, and every non-basic proposition that depends upon b for its warrant would lose some of its warrant were b not autonomously warranted.

Doris should say to Fred, "I grant that b has autonomous warrant. But what I want to know is whether autonomously warranted propositions are, in virtue of that fact, somewhat likely to be true." Her worry becomes a "meta" worry. But she went meta, so to speak, because Fred went meta first.<sup>8</sup>

Given that with regard to any proposition, once we consider whether it is true, we must hold it, deny it, or withhold it (i.e., neither hold nor deny it), Fred is now faced with a trilemma:<sup>9</sup>

- 1 He can hold that autonomously warranted propositions are somewhat likely to be true in virtue of the fact that they are autonomously warranted.
- 2 He can deny that autonomously warranted propositions are somewhat likely to be true in virtue of the fact that they are autonomously warranted.
- 3 He can withhold whether autonomously warranted propositions are somewhat likely to be true in virtue of the fact that they are autonomously warranted.

If he takes alternative 2, then using b as a reason for the first non-basic proposition in the series is arbitrary. Holding b is not arbitrary. Doris has granted that b is autonomously warranted and she could grant that it is not arbitrary to hold a proposition that has autonomous warrant. But if Fred believed that such propositions were not even somewhat likely to be true in virtue of being autonomously warranted, how could he think that b could provide a good reason for thinking that the penultimate proposition was likely to be true? Fred thinks that the warrant for all of his beliefs rests on basic propositions. If he thought that b's possession of F was not the least bit truth conducive, then why is he using b and all the other basic propositions on which the warrant for his non-basic beliefs rests?

The same applies to alternative 3. Doris has asked whether the fact that b is autonomously warranted makes it at all likely that b is true. Fred responds that he doesn't have an opinion one way or the other. Fred thinks b is true, but he neither has a reason for thinking it is true, nor does he think that basic propositions are somewhat likely to be true because they are autonomously warranted. So, from Fred's point of view and Doris's, Fred ought not to use b as the basis for further beliefs. The mere fact that he thinks b is true is not sufficient for him to use b as a reason, unless he thinks that his thinking that b is true somehow makes it likely that b is true.

If he takes alternative 1, then using b as his reason for the penultimate proposition is not arbitrary, *but that is because the regress has continued*. Fred has a very good reason for believing b, namely b has F and propositions with F are likely to be true. Fred, now, could be asked to produce his reasons for thinking that b has F and that basic propositions are somewhat likely to be true in virtue of possessing feature F.

*Therefore:* foundationalism cannot solve the regress problem, even if it were true. A practicing foundationalist cannot increase the rational credibility of a questioned proposition through reasoning.

Let us turn to coherentism. The first form, the warrant-transfer form, is easily seen to be unable to solve the regress problem because Carl, the Coherentist, cannot increase

the credibility of some proposition, *p*, by citing *p* in its own evidential ancestry. If the reasoning is to increase the credibility of the questioned proposition for Carl, then that credibility will not already be cathected to the proposition. For if it were, then it is pointless to begin reasoning in the first place. Presumably that is what is wrong with circular reasoning. It cannot increase the credibility of a questioned proposition.

Indeed, the difficulty facing *all* warrant-transfer accounts (foundationalism and this type of coherentism) is more serious than that credibility will not be *added* to the questioned proposition by reasoning. There is a danger that credibility will actually diminish as the warrant is transferred.<sup>10</sup> If all of the inferences employed in the reasoning were deductive and if an appropriate form of closure holds, it would seem that credibility would not be lost. It could even increase. But if during the transfer of warrant some credibility were lost, as it would be, if the inference links were non-deductive, then the longer the series or the larger the circle, the more credibility would be lost.

The only escape from this difficulty for a warrant-transfer theorist is to (i) limit the number of transfers allowed, or (ii) require that there are sufficiently strong coherence relations to make up for the lost warrant, or (iii) require that enough of the transfers are by way of deduction. Those stipulations seem entirely *ad hoc*.

The second form of coherentism, the warrant-emergent form, seems more promising because it eschews circular reasoning, and warrant for propositions could increase as the number of threads in the web of propositions increases and/or the web becomes more tightly woven. But there is one problem with this form of coherentism. As others have pointed out, it is nothing but a type of foundationalism – one-step foundationalism.<sup>11</sup> In this case, the foundational property, *F*, which all warranted propositions have, is that each is a member of a set of coherent propositions. The Carl–Doris discussion would follow the same general pattern as the Fred–Doris discussion where the foundational property, *F*, is simply the proposition's membership in the set of coherent propositions.

Thus, Carl faces a trilemma similar to Fred's, discussed above. If he says either "no" or "I withhold," then vesting his credence in a coherent set of propositions is arbitrary. Why should he adopt a coherent set rather than an incoherent set? If he says that coherent propositions are likely to be true in virtue of the fact that they are coherent, then he faces the third horn.<sup>12</sup> For either the proposition "coherent sets are ipso facto likely to contain propositions that are true" is included in the initial coherent set or it isn't. If it is, then he has fallen back to the warrant-transfer – that is, question-begging – form of coherentism. If it isn't, then he has just added a new proposition to the coherent set and the regress has continued.

Now, coherentists might suggest that mere coherence is not sufficient to demarcate a set of warranted propositions. They could require, for example, that the set contain some propositions that have some further feature, namely that they are spontaneously endorsed or that their content has certain phenomenal properties (for example, see BonJour, 1985). But that is just another specification of the foundational *F*-property and the trilemma would reappear.

*Therefore:* coherentism cannot solve the regress problem, even if it were true. A practicing coherentist cannot increase the credibility of a questioned proposition through reasoning.

Perhaps there is no solution to the regress problem – that is, no way to add credibility to a proposition by reasoning. But before we come to that rather dismal conclusion, it would be appropriate to look at the third alternative.

## Step 2: Infinitism

What is infinitism? Infinitism is like the warrant-emergent form of coherentism because it holds that warrant for a questioned proposition emerges as the proposition becomes embedded in a set of propositions. Infinitism is like foundationalism because it holds that some propositions are epistemically prior to others. But some caution is needed if we are to be able to account for the coherentist intuition that (some) propositions are mutually supporting. For example, “all humans are mortal” is a reason for believing that “this human is mortal,” and the converse is true. Some have thought that the universal generalization is always epistemically prior to the particular, and others have thought that the particular is always epistemically prior to the generalization. Each view runs afoul of our reasoning practice. Sometimes we offer the generalization as a reason for the particular – when the particular is what is questioned. Sometimes we offer the particular as a reason for the generalization – when the generalization is questioned. But we cannot use the generalization as a reason for the particular and the particular as a reason for the generalization *in the course of one reasoning session*. That would be to fall into circular, question-begging reasoning.

What we seek is an account of warrant that is not a warrant-transfer view and is not warrant-emergent finite coherentism. There is only one option remaining. What we need is warrant-emergent infinitism. Such a view leads neither to the arbitrary employment of a so-called basic propositions nor to the endorsement of circular reasoning. It can solve the regress problem because it endorses a warrant-emergent form of reasoning in which warrant increases as the series of reasons lengthens.

Infinitism results from adopting the following two principles:<sup>13</sup>

*Principle of Avoiding Circularity (PAC):* for all propositions,  $x$ , if  $x$  is warranted for a person,  $S$ , at  $t$ , then for all  $y$ , if  $y$  is in the reason-ancestry of  $x$  for  $S$  at  $t$ , then  $x$  is not in the reason-ancestry of  $y$  for  $S$  at  $t$ .

*Principle of Avoiding Arbitrariness (PAA):* for all propositions,  $x$ , if  $x$  is warranted for a person,  $S$ , at  $t$ , then there is some reason,  $r_1$ , available to  $S$  for  $x$  at  $t$ ; and there is some reason,  $r_2$ , available to  $S$  for  $r_1$  at  $t$ , etc., and there is no last reason in the series.

PAC is readily understandable and requires no discussion. It simply recognizes that a warrant-transfer view cannot solve the regress problem by endorsing circular reasoning. PAA, on the other hand, introduces the notion of “available reasons” and some account of that is required.

There are two conditions that must be met in order for a proposition to be *available* to  $S$  as a *reason* for  $x$  at  $t$ . First, the proposition must be *available* to  $S$  at  $t$ ; that is, it must be appropriately “hooked up” to  $S$ ’s beliefs and other mental contents at  $t$ . In order for a proposition to be available in this sense it need not be occurrently believed or endorsed by  $S$  at  $t$ . For example, the proposition “ $352 + 226 = 578$ ” is available even though it might never be consciously entertained. Whether this is best understood as (a) a disposition to believe that  $352 + 226 = 578$  or (b) a second order disposition to form a disposition to believe that  $352 + 226 = 578$  is a matter of detail that can be put aside.

The second condition is that the proposition must be a *reason* for  $S$  at  $t$ . Now, what makes a proposition a reason need not be fleshed out here. That’s a good thing because the issue is a difficult one and there are many alternative accounts that could be

employed by the infinitist. It is here that infinitism can (but need not) make room for externalist accounts of justification and for a supervenience requirement in which the supervenience base is limited to non-normative facts.<sup>14</sup> For example, some proposition, say *p*, could be held to be a reason for *q* iff

- 1 *p* is true and it renders *q* probable; or
- 2 *p* would be accepted as a reason for *q* in the long run by the appropriate epistemic community; or
- 3 *p* would be offered as a reason for *q* by an epistemically virtuous individual; or
- 4 there is cognitive process available to *S* which reliably takes true beliefs that *p* into true beliefs that *q*.

There are other possible accounts. The point is that whatever the proper account of reasons is, coherentists, foundationalists, and infinitists will have to employ it because each view holds that there are reasons for at least some of our beliefs. So, this thorny issue can be set aside for the purposes of this essay.

Nevertheless, these two conditions make clear what infinitism is committed to, and, more importantly, what it is *not* committed to. For example, the mere existence of an infinite set of propositions, each of which entails the next in the series, is not sufficient for there to be the appropriate series of reasons available which could provide the missing credibility. As has been pointed out by others, there will be an infinite series of propositions each entailed by a previous one in the series for *every* proposition. The point is that not just *any* infinite series of propositions will do. The propositions must be available and they must be reasons.<sup>15</sup>

### Step 3: Replies to the Best Objections to Infinitism

It is now time to examine what I think are the two best objections to infinitism, beginning with the oldest. Recall what Sextus said:

The Mode based upon regress *ad infinitum* is that whereby we assert that the thing adduced as a proof of the matter proposed needs a further proof, and this again another, and so on *ad infinitum*, so that the consequence is suspension [of assent], as we possess no starting-point for our argument.

Now, if efficacious reasoning required that warrant originate in and be transferred from a basic proposition, this criticism would be just. But for the reasons given above, infinitism eschews such a view. The “starting point” of reasoning is, as Peirce says: doubt. A proposition becomes questionable and, consequently, it lacks the desired rational credibility. Reasoning scratches the itch. The infinitist holds that finding a reason for the questioned proposition, and then another for that reason, and so on, places it at the beginning of a series of propositions each of which gains warrant and rational credibility by being part of the series. Warrant increases not because we are getting closer to a basic proposition but rather because we are getting further from the questioned proposition. But the Pyrrhonist is correct that the infinitist’s conception of reasoning precludes assenting to a non-evident proposition. Dogmatism is incompatible



with practicing infinitism. Warrant, and with it rational credibility, increases as the series lengthens; but the matter is never completely settled.

In conclusion let me turn to the finite mind objection. Here is what John Williams says:

The [proposed] regress of justification of S's belief that p would certainly require that he holds an infinite number of beliefs. This is psychologically, if not logically, impossible. If a man can believe an infinite number of things, then there seems to be no reason why he cannot know an infinite number of things. Both possibilities contradict the common intuition that the human mind is finite. Only God could entertain an infinite number of beliefs. But surely God is not the only justified believer. (Williams, 1981, p. 85)

I hope that it is clear how to answer that objection. Infinitism does not require that we "hold" an infinite number of beliefs – if that means that there is some time at which an infinite number of beliefs are occurrent. Infinitism does require that there be an infinite non-repeating set of propositions each of which is an available reason for a preceding one.

But some philosophers have suggested that such a set cannot be available. Audi, for example, writes:

Let me suggest one reason to doubt that human beings are even capable of having infinite sets of beliefs. Consider the claim that we can have an infinite set of arithmetical beliefs say the 2 is twice 1, the 4 is twice 2, etc. Surely for a finite mind there will be some point or other at which the relevant proposition cannot be grasped ... and what we cannot grasp we cannot believe. I doubt that any other lines of argument show that we can have infinite sets of beliefs; nor, if we can, is it clear how infinite epistemic chains could account for any of our knowledge. (Audi, 1993, pp. 127–128)

Let us grant that such a set is not available to us. Of course, it does not follow that there could not be an infinite set of propositions available whose members do not increase in complexity. In fact, contra Audi, I think there is a simple argument to show that there is such a set.

Suppose we have a very limited set of concepts or vocabulary: {x is F, red, indexical "that"}. In other words, we can believe of an object: that is red. Now imagine that there are an infinite number of red objects. We could believe of each object that it is red. Those are different beliefs because the truth conditions of the propositions affirmed in the beliefs are distinct.

Are there an infinite number of red objects? I don't know. But that is not necessary for my argument. All I need to show is that a finite mind *can* have access to an infinite number of beliefs. And I have shown that.

Audi also claims that even if there were an infinite series of propositions each of which is available, it is not "clear how infinite epistemic chains could account for any of our knowledge." Now, if knowledge required actually completing the series, knowledge would not be possible. But why suppose that knowledge requires the highest possible degree of warrant or absolutely credible belief? As the series lengthens, warrant and credibility increase. Nothing prevents it increasing to the degree required for knowledge.

## Acknowledgments

I am indebted to the paper “Infinitism and degrees of justification” by Jeremy Fantl for making this point so clearly. I wish to thank Anne Ashbaugh, Troy Cross, Jeremy Fantl, Alvin Goldman, Brian McLaughlin, and Ernest Sosa for their help with this essay. Ancestor versions were read at the New Jersey Regional Philosophical Association, University of Colorado-Boulder, Montclair State University, and Bryn Mawr College. The discussions at the time and follow-up conversations and e-mails were very valuable, especially those with George Bealer, David Benfield, Michael Huemer, Aryeh Kosman, Michael Krausz, and Kenneth Richman.

## Notes

- 1 Sextus Empiricus, *Outlines of Pyrrhonism*, I, 166–9.
- 2 For a foundationalist employing this strategy, see Audi (1993, pp. 127–128); for a (former) coherentist employing it, see BonJour (1985, pp. 18–24).
- 3 Some of these objections, as well as others, are treated in more detail in Klein (1999, 2003, 2004a).
- 4 For this use of “warrant,” see Plantinga (1993, p. 3).
- 5 Aristotle, *Posterior Analytics*, 73a1–5.
- 6 For a defense of warrant-emergent coherentism, see BonJour (1985).
- 7 Reasons might supply additional warrant, but these are or ultimately depend upon other basic propositions for their warrant. Doris could ask what reason(s) he has for believing the conjunction of the basic propositions. Fred’s reply will be the same, “there is no reason for believing that conjunction other than each conjunct.”
- 8 For a similar argument, see BonJour (1985, pp. 9–14). I have discussed this elsewhere (see Klein, 1999).
- 9 Ernest Sosa discusses a similar issue in his 1994 (p. 107) and in “Two False Dichotomies” (2004), published in a volume of conference proceedings honoring Robert Fogelin, edited by Walter Sinnott-Armstrong. I have discussed Sosa’s views in Klein (2004b).
- 10 Thus, infinitism cannot endorse a warrant-transfer view. This potential problem for infinitism was originally suggested to me by Troy Cross.
- 11 I do not think the expressions “warrant-transfer coherentism” or “warrant-emergent coherentism” are original with me. But I do not recall where I first ran across those terms. Laurence BonJour (1985) distinguishes the two types of coherentism, as does Ernest Sosa (1980). Sosa also points out that warrant-emergent coherentism is a form of what he calls “formal foundationalism.” Thus, the claim that some forms of coherentism are actually forms of foundationalism is not original with me.
- 12 Ernest Sosa (1997) advocates seizing this horn of the trilemma. I have discussed this in Klein (forthcoming).
- 13 Note that PAC and PAA are necessary conditions for warrant. They are not intended to be jointly sufficient. At least a “non-overrider” clause and a “non-defeasibility” clause would need to be added in order to have a sufficient set of conditions. For discussions of those issues see Klein (1971, 1981, 2003).
- 14 This is important because it provides the basis for an answer to the objection that infinitism cannot account for the supervenience of the normative on the non-normative discussed in Goldman (1979), Sosa (1980), and Van Cleve (1992, especially pp. 350–351 and 356–367).
- 15 This objection to infinitism was developed in Post (1980, especially p. 34–35), and in Post (1987, pp. 84–92). For my reply, see Klein (1999, p. 312).

## References

- Audi, R. (1993) *The Structure of Justification*. New York: Cambridge University Press.
- BonJour, L. (1985) *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.
- Goldman, A. (1979) What is justified belief? In *Justification and Knowledge*, ed. G.S. Pappas. Dordrecht: Reidel.
- Klein, P. (1971) A proposed definition of propositional knowledge. *Journal of Philosophy* 67: 471–482.
- Klein, P. (1981) *Certainty: A Refutation of Scepticism*. Minneapolis: University of Minnesota Press.
- Klein, P. (1993) When infinite regresses are *not* vicious. *Philosophy and Phenomenological Research* 63: 718–729.
- Klein, P. (1999) Human knowledge and the infinite regress of reasons. *Philosophical Perspectives* 13: 297–332.
- Klein, P. (2003) Knowledge is true, non-defeated justified belief. In *Essential Knowledge*, ed. S. Luper. Harlow: Longman.
- Klein, P. (2004a) What is wrong with foundationalism is that it cannot solve the epistemic regress problem. *Philosophy and Phenomenological Research* 68: 166–171.
- Klein, P. (2004b) Skepticism: ascent and assent. In *Philosophers and Their Critics*, ed. J. Greco, pp. 112–125. Oxford: Blackwell.
- Plantinga, A. (1993) *Warrant: The Current Debate*. New York: Oxford University Press.
- Post, J. (1980) Infinite regresses of justification and of explanation. *Philosophical Studies* 38: 31–52.
- Post, J. (1987) *The Faces of Existence*. Ithaca, NY: Cornell University Press.
- Sosa, E. (1980) The raft and the pyramid. *Midwest Studies in Philosophy* 5: 3–25.
- Sosa, E. (1994) Philosophical skepticism and epistemic circularity. *Proceedings of the Aristotelian Society* 68: 263–290.
- Sosa, E. (1997) Reflective knowledge in the best circles. *Journal of Philosophy* 94: 410–430.
- Sosa, E. (2004) Two false dichotomies. In *Pyrrhonian Skepticism*, ed. W. Sinnott-Armstrong. Oxford: Oxford University Press.
- Van Cleve, J. (1992) Semantic supervenience and referential indeterminacy. *Journal of Philosophy* 89: 344–361.
- Williams, J. (1981) Justified belief and the infinite regress argument. *American Philosophical Quarterly* 18: 85–88.

# Infinitism Is Not the Solution to the Regress Problem

*Carl Ginet*

Many of our beliefs are justified beliefs: they are such that epistemic rationality would not forbid our holding them. And often what justifies a belief is the fact that the believer has (or has “available”: more on this later) other justified beliefs from which the belief in question can be properly inferred. That is to say, the justification of many

a justified belief is by inference from one or more other beliefs: it is *inferential justification*. For example, from my belief (based on observation) that our car is not in the driveway and my belief that my wife is the only one besides me who has a key to our car, I infer that my wife is not at home.

The premise beliefs must, of course, themselves be justified, if belief in what is inferred from them is to be thereby justified; and their justifications may be by inference from still other beliefs; and the justifications of those further beliefs may be inferential; and so on. Can this go on without end? If not, how can it end? Those questions are the regress problem.

I think that it cannot go on without end, that any ramifying structure of inferential justifications must end in justifications that are not inferential. Not so long ago I thought this truth so obvious as to need no argument. But Peter Klein has made me realize that it is not *that* obvious. (Besides his contribution to the present debate, see especially Klein, 1999. My page references will be to this latter paper.)

Klein holds that inferential justifications not only can ramify without end but *must* do so for any belief that is truly justified. He holds that every justification must be inferential: no other kind of justification is possible. (He also holds, and with this we must all agree, that no inferential justification can be circular, can be such that a belief is inferred ultimately from itself.) This is Klein's *infinetism*.

I, on the other hand, insist that inferential justification cannot ramify without end (or, rather, beginning), and that if justification is possible then non-inferential justification must be possible. Klein calls this view *foundationalism*. It might also be called *finitism*.

## Examples of Non-inferential Justification

One reason I think that non-inferential justification is possible is that I think we can give clear examples of it. Let me give two, one a priori and one a posteriori.

### A priori

Consider the following sentence:

A: Anything that lasts exactly one hour lasts exactly 60 minutes.

Someone who does not accept that what sentence A says is true – who doubts or denies that, or is uncertain whether, what it says is true – must be counted as one who does not understand what sentence A says (provided he has no specific reason to suspect that what it says, together with other equally evident-seeming propositions, entails a contradiction; this proviso is a complication I will hereafter usually omit to mention, as it is, for sentence A, always satisfied in actual cases).

Someone who does understand what sentence A says, and therefore believes it, is justified in believing it. The fact which constitutes his being justified in believing it is simply the fact that he understands what the sentence says. That he understands what it says entails that he believes what it says. So it cannot be that epistemic rationality requires that he ought not to believe it *even though* he understands it. Nor, surely, can it require that he ought not to understand what it says. If, as far as epistemic rationality

is concerned, he cannot be criticized for understanding it, then he cannot be criticized for what his understanding it requires, his believing it. Therefore, given that he understands what it says, he is justified in believing it.

The fact that he understands it *is* his justification for believing it, because that fact entails his believing it. This sort of justification does not appear to involve any sort of inferential justification: it does not entail any (available) belief in any premise such that what the sentence says is properly inferable from that premise. Therefore, one who understands, hence believes, what sentence A says is one who has a non-inferentially justified belief.

This belief and a similarly non-inferentially justified belief that

B: Anything that lasts 60 minutes lasts longer than anything that lasts just 55 minutes.

might provide the ultimate premises for an inferential justification for believing (what A and B obviously entail) that

C: Anything that lasts one hour lasts longer than anything that lasts just 55 minutes.

In such a case the subject's justification for believing C originates in her understanding, hence believing, A and B: the regress ends there.

## A posteriori

Suppose that my visual system is working properly and I see a blue smear on a white surface in good light a few feet in front of me. Suppose further that I believe that I see a blue smear on a white surface in good light a few feet in front of me and I am not aware of any reason to think that in this instance things may not be what they seem, to think that my visual system may not be working properly or that external circumstances may be conspiring to produce a visual illusion. Then I am justified in that belief.

What justifies my belief? The following two facts are sufficient: (1) my visual experience is as if (my visual experience represents that) I see a blue smear on a white surface in good light a few feet in front of me and (2) I am not aware of any reason to think that in this instance things may not be what they visually seem to me to be. Given those facts, there is no basis for faulting me for holding the belief. No good reason could be given for saying that, despite those facts, I ought not, I am unreasonable, to believe that I see a blue smear on a white surface.

If these facts do constitute a justification for that belief, that justification is obviously not inferential. It involves no further beliefs at all. Fact (1) is just the fact that I have a certain specific sort of visual experience, a fact which does not include or entail any particular belief on my part. My having this experience does not in itself include my believing that I have it (whether or not it is impossible for me to fail to believe that I have it if I do), nor does the visual experience include (or entail) my believing that what the experience represents as being there before me actually is there before me, that there is a blue smear on a white surface in good light a few feet in front of me. Fact (2) is completely negative. It is just the *absence* of any (available) belief that would be reason for me to suspect that, in this instance, things may not be what they visually seem to be; and this absence does not entail the *presence* of any particular (available) beliefs at all.

This non-inferentially justified belief, that I see a blue smear on a white surface, might be an ultimate premise of an (no doubt complex and extended, but finite) inferential justification for believing that my grandson was recently in the room with blue fingerpaint on his hands. Among the other ultimate premises in this justification would be memory beliefs non-inferentially justified in parallel fashion, by a combination of my seeming to remember that I came to know the propositions believed and my lacking any reason to think that in this instance my memory is not to be trusted.

## Objections to These Examples

One objection is suggested in Klein's comments on his Principle of Avoiding Arbitrariness – which is the principle that “For all  $x$ , if a person,  $S$ , has a justification for  $x$ , then there is some reason,  $r_1$ , available to  $S$  for  $x$ ; and there is some reason,  $r_2$ , available to  $S$  for  $r_1$ ; etc.” Klein (1999, p. 299) says:

Note that there are two features of this principle. The first is that it is reasons (as opposed to something else like appropriate causal conditions responsible for a belief) that are required whenever there is a justification for a belief. The second is that the chain of reasons cannot end with an arbitrary reason – one for which there is no further reason. [Both features] are needed to capture the well-founded intuition that *arbitrary beliefs*, beliefs for which no reason is available, should be avoided.

These remarks suggest that the facts cited in my alleged non-inferential justifications do not justify the beliefs in question because they do not entail that the believer has a *reason* for the belief. The argument seems to be this:

- 1 One's belief is unjustified if one lacks a good reason for it.
  - 2 Having a good reason for a belief *just means* having another belief from which the belief in question can be properly inferred – that is, having an inferential justification for it.
- ∴ To have a justification for a belief is to have an inferential justification for it.

It must be granted (it is a tautology) that one's belief is unjustified if one lacks a reason for it in the sense of *some sort of justification* for it. But it should not be granted, and it does not follow, that one's belief is unjustified if one lacks a reason for it in the sense of an *inferential justification* for it. If premises 1 and 2 seem intuitively acceptable to us when we read them, this can only be because we shift from one sense of “having a reason” to the other in going from one premise to the other.

I find it quite intelligible to say that my *reason* for believing that I saw a blue smear on a white surface was the fact that my visual experience was as if that were so, together with the fact that I was aware of no reason to think that in this instance things were not as they visually seemed to be. Hence I am inclined to allow that “having a reason” can be broadly used to cover non-inferential justifications and, sticking with that broad sense throughout the argument above, to accept premise 1 and reject premise 2. But the crucial point is that, if premise 2 is urged on the basis of the ordinary use of “having a reason,” then premise 1 cannot be taken as clear on the basis

of ordinary usage; it needs to be argued that being justified in a belief requires having a reason in the (supposedly ordinary) sense of having an inferential justification.

Specifically, it needs to be shown why any putative example of non-inferential justification (such as the ones I have given) cannot really be such. Klein considers the suggestion that a belief might be justified by some property P it has that does not entail the believer's having an inferential justification for it and responds as follows:

Pick your favorite account of the property, P. ... Why is having P truth-conducive? Now, either there is an answer available to that question or there isn't. ... If there is an answer, then the regress continues – at least one more step. ... If there isn't an answer, the [belief] ... is arbitrary. (Klein, 1999, p. 303)

... arguing that such beliefs are likely to be true because they possess a certain property, P, will not avoid the problem faced by foundationalism. Either [this] ... justification provides a reason for thinking the [believed] ... proposition is true (and hence the regress does not end) or it does not (hence, accepting the base proposition is arbitrary). (Klein, 1999, p. 304)

These remarks suggest the following counter to my claim that the facts cited in each of my examples do provide non-inferential justification for the belief in question: those facts are relevant to justifying the belief only if the believer has (available), as a reason for the belief, a further belief that when such facts obtain the belief is likely to be true; they cannot all by themselves justify the belief but can provide only a part of an inferential justification for it.

Thus, with respect to my putative example of a non-inferential a priori justification, Klein would seem to want to say the following: you propose that the belief in question (that anything that lasts exactly one hour lasts exactly sixty minutes) is justified by its having the property of being such that (a) the subject understands the proposition believed and (b) that proposition is such that believing it is a requisite of understanding it; but the subject is not justified in that belief unless it is also the case that he has (available), as inferential support for it, the further belief that the belief's having that property makes it likely to be true. And with respect to my putative example of a non-inferential a posteriori justification, he would seem to want to say: you propose that the belief in question (that I see a blue smear on a white surface in good light a few feet in front of me) is justified by its having the property of being accompanied by its visually seeming to me as if I see a blue smear on a white surface in good light a few feet in front of me and by my lacking awareness of any reason to think that in this instance things may not be what they visually seem to be; but you are not justified in that belief unless it is also true that you have (available), as inferential support for it, the further belief that the belief's having that property makes it likely to be true.

I will discuss just this second response, concerning my claim about the a posteriori justification of the perceptual belief (and leave it as an exercise for the reader to infer what I would say about the first response, to my claim concerning the a priori justification). We should note first that plenty of people who are justified in such a perceptual belief when they satisfy the sort of condition I specified do not in fact have any such further belief (that the perceptual belief's being accompanied by the perceptual experience, etc., makes it likely to be true) – perhaps because they have never entertained any such proposition, perhaps they even lack the concepts (e.g., of subjective visual experience or of probability) needed to entertain it.

In response to this Klein will point out that he is not requiring the subject actually to have the further belief but only that it be *available* to him, in the sense that, if the subject were to have the concepts needed to entertain the proposition in question and were to entertain it, he would accept it. (That Klein is prepared in this way to count as available to a subject a belief in a proposition that the subject does not yet possess the concepts to entertain is clear from his response (1999, pp. 308–309) to the challenge “to show that there can be an infinite number of reasons given a finite vocabulary each of which can be entertained by a human being.”)

Moreover, Klein might argue, this potential further belief is a *reason* for the perceptual belief in these circumstances because it would be unreasonable of the subject in these circumstances not to accept this further proposition (were he to entertain it) while continuing to hold the perceptual belief; and insofar as the believer acquires his perceptual belief because of the visual experience, he *acts as if* he had the further belief (that the perceptual belief’s being accompanied by the perceptual experience, etc., makes it likely to be true), he acts as one would who was motivated by such a further belief.

Perhaps so, I say, but only if he actually is motivated by this potential further belief can it be counted among his reasons for acquiring the perceptual belief. That a potential belief is *available* in Klein’s sense makes it perhaps a potential reason, but it is not enough to make it among the believer’s actual reasons. And only if it is among his actual reasons is it part of his actual justification. It is true that the subject must, if his perceptual belief is justified in these circumstances, have available (in Klein’s weak sense) this other belief. But this is only because his just having that perceptual belief in these circumstances entails that he has available (in Klein’s sense) that other belief; and from this it does not follow that its availability is any part of his actual justification. And the issue we are debating, I take it, is whether the *actual* justification for some beliefs can be non-inferential.

It is a trivial and uninteresting truth that, for any belief, the believer has a *potential* inferential justification. Consider a young child who has acquired enough understanding of elementary arithmetic to believe that  $2 + 3 = 5$ .

The proposition that

(\*) The smallest even prime added to the cube root of 27 equals the square root of 25.

is such that, were the child to understand and entertain it, it would be unreasonable of him not to accept it while continuing to believe that  $2 + 3 = 5$ ; so that in believing that  $2 + 3 = 5$  he acts as would one who believed the proposition (\*). Yet it is clear that the child’s potential belief in (\*) is not among his reasons for believing that  $2 + 3 = 5$  and not part of the story as to why his so believing is justified.

For any proposition P that any subject believes, there is another proposition such that, were he to entertain it, it would be unreasonable of him not to accept it while continuing to believe P (and thus he acts as would one who believed this further proposition), namely, a proposition of the form “ $P \vee (Q \wedge \neg Q)$ ” (where Q is any other proposition). It is clear that the subject’s potential belief in this proposition is not among his reasons for believing P and not part of the story as to why he is justified in this belief (if he is).

If I tell you that in his present circumstances S would accept the following proposition were he to entertain it:



- (L) It is likely to be true that one sees a blue smear on a white surface when one's visual experience is as if one were seeing a blue smear on a white surface.

I indirectly indicate to you what justifies S in believing that he sees a blue smear on a white surface – namely, the fact that his visual experience is as if he were seeing a blue smear on a white surface (and he is unaware of any reason to think that in this instance things are not what they visually seem): you can infer that he is disposed to believe (L) because his visual experience is of that sort. But of course reporting that fact about his potential belief in order to indirectly indicate his justification is not the same as reporting that that fact *is* (part of) his justification.

The notion of an available reason for a belief might be strengthened in such a way as to make it plausible that, if the belief that (L) is, in that stronger way, available to S as a reason for believing that he sees a blue smear, then it is a reason of his for so believing. An additional requirement that might do the trick is this: (L) is available to S as a reason for so believing only if S is disposed, upon entertaining and accepting (L), to believe that the fact that (L) was among his reasons for so believing. It would then, perhaps, be right to say that, if the belief that (L) is available to S as a reason in that stronger sense, then it is part of S's actual justification for believing that he sees a blue smear on a white surface that the belief that (L) is available to him. His *tacit* belief that (L), we might say, was a tacit reason of his for believing that he sees a blue smear on a white surface.

But of course it does not follow from this concession that the facts I cited earlier, of his having a certain sort of visual experience and lacking any reason to think that in this instance things are not what they seem, do not all by themselves provide S with a non-inferential justification. It does not follow, and I see no reason to concede, that if those facts obtained but the belief that (L) was *not* available to S in that stronger sense – surely a possible case – then S would not be justified in believing that he sees a blue smear on a white surface. We would still have no argument that a belief cannot be justified unless the believer has available (in the stronger sense) an inferential justification for it.

## Two Problems for Infitism

### First problem

Let us suppose that the notion of availability is strengthened in some such way as I have suggested, so as to make it plausible that a belief B2 that is available as a reason for a belief B1 is, not merely a potential reason, but among the believer's actual reasons for B1. Given that stronger (less easily satisfied) notion of availability, infinitism – the doctrine that a belief is justified only if the believer has available an inferential justification for it – would face the following question.

What reason can be given for thinking that any of our beliefs is such that the believer *has available in that stronger sense* an infinitely ramifying structure of inferential justifications, for thinking that infinitism has not laid on justification of belief a requirement that is never (or seldom) actually met? With regard to many people who acquire a basic perceptual belief when they have appropriate perceptual experience, it may be plausible

to suppose that, if they entertained the general proposition to the effect that when one has such experience the perceptual belief is likely to be true, they would not only believe it but also take this tacit general belief to be part of their reason for their perceptual belief. But is it equally plausible to suppose that they have available in that same strong sense still another belief which is their tacit reason for holding that tacit general belief? What would it be? And what would be their tacit reason for *that* tacit belief?

To argue that it is *in principle possible* for a human being to have available (in the stronger sense) an infinite number of beliefs would not be enough. To make it plausible that there actually occur justifications having the endlessly ramifying structure that infinitism says all justifications must have, the infinitist must provide representative examples of particular such structures possessed by cognitively normal human subjects – examples about which it would be credible that cases essentially like them actually occur. This would be none too easy to do.

It would, of course, be out of the question to specify individually all of the links in an endless chain. One can specify an infinite series only by providing a general way of (an algorithm for) finding a new member of the series no matter how (finitely) far the series has already been extended. Such an algorithm would have to tell us, with respect to a sort of belief we are sure is well justified – for example, my belief as I look out the window at my driveway that there is a car there – how to construct an endless inferentially linked chain of specific premise beliefs, which is such that it is plausible to think that, in the ordinary sort of circumstance in which (as we think) such a belief is justified, that infinite series of beliefs is available (in the stronger sense) to the holder of the belief. Specifying such an algorithm would be a formidable task. In fact, I am at a loss to see how it could be done.

## Second problem

A more important, deeper problem for infinitism is this: Inference cannot *originate* justification, it can only *transfer* it from premises to conclusion. And so it cannot be that, if there actually occurs justification, it is all inferential.

Inferential justification is analogous to instrumental value in this respect. Things have value as means to other things only if ultimately some things have value in themselves and not just as means to other things. The relation *x-is-a-means-to-y* can only transfer to *x* whatever value *y* has; it cannot *create* any value. But there can be no value to be transferred unless ultimately something else, something other than the means–end relation, does create value. Analogously, the relation *p-can-be-properly-inferred-from-q* can only transfer to *p* whatever justification *q* has; it cannot create any justification. But there can be no justification to be transferred unless ultimately something else, something other than the inferential relation, does create justification.

Jonathan Dancy (1985, p. 55) puts the point this way:

Justification by inference is conditional justification only; [when A is inferred from B and C] A's justification is conditional upon the justification of B and C. But if all justification is conditional in this sense, then nothing can be shown to be actually non-conditionally justified.

Klein (1999, p. 313) replies to Dancy's remarks as follows: "The answer is simply that although every proposition is only provisionally justified, that is good enough if one

does not insist that reasoning settle matters once and for all.” But this reply misses the point. From the denial that reason (or justification) can ever settle matters once and for all it follows only that every justification is provisional in the sense of *defeasible* – where a justification is defeasible if it is compatible with the facts F1 constituting the justification that there be other facts F2 that defeat the justification, that is, F2 and F1 combined would *not* justify what F1 alone does justify, so that facts F1 justify only provided that, only so long as, no defeating facts F2 turn up. That a justification is provisional in this sense does not entail that it is inferential. By conditional justification Dancy means not provisional, but inferential justification, where the fact that the premise beliefs are justified and have the inferential relation to the conclusion explains why the conclusion belief is justified. And the point of the objection is that an endless chain of inferential justifications can never ultimately explain why any link in the chain is justified.

Consider another analogy, between acceptable inference and justification on the one hand and deductively valid inference and truth on the other. Deductively valid inference preserves, but does not create, truth. If a set of premises are true then the property of truth will be “transferred” to any conclusion validly deducible from them. But, given a chain of propositions that are linked by the deductive relation, there is nothing in that relation itself that contributes to making any of those propositions true, no matter how extended the chain might be.

Analogously, acceptable inference preserves justification: if one has justification for believing the premises then one has justification for believing anything one recognizes as acceptably inferable from them. But, given a chain of beliefs linked by acceptable inference, there is nothing in the inferential relation itself that contributes to making any of those beliefs justified, nothing that explains why any of them is justified; and this is so no matter how extended the chain might be.

## References

- Dancy, J. (1985) *Introduction to Contemporary Epistemology*. Oxford: Blackwell.  
 Klein, P. (1999) Human knowledge and the infinite regress of reasons. In *Philosophical Perspectives*, 13 (*Epistemology*), ed. E. Tomberlin. Oxford: Blackwell.

# Reply to Ginet

*Peter Klein*

Let me begin by thanking Carl Ginet. He has helped me to see more clearly where infinitism and finitism can agree, and where and why they diverge.

First, I will list the areas where finitists and infinitists can agree. Second, I will list an area where I believe agreement can be reached. Third, I will discuss some remaining areas of disagreement and attempt to defend infinitism. Finally, I will conclude with a brief remark about the basic difference between the two views.

## Areas of Agreement between the Infinitist and the Finitist

- A1 A belief is justified for S when epistemic rationality “would not forbid ... holding” it (Ginet, p. 283).<sup>1</sup>
- A2 There are some beliefs that require having reasons in order to be *actually justified* for S. Those reasons must be available to S, and at least some of the those reasons must “motivate” the belief (Ginet, p. 288).
- A3 The process of giving reasons for a belief comes to an end in actual circumstances, and what’s more, the canons of epistemic rationality do not require that further reasons motivate the belief in order for it to be at least partially actually justified.
- A4 In a slightly extended use of “reason” such things as perceptual states, memories, or understanding the meaning of an expression are reasons that can make a belief at least partially justified. In the case discussed by Ginet, “no good reason could be given for saying that ... I am unreasonable to believe that I see a blue smear on a white surface” (Ginet, p. 285).

## An Area of Potential Agreement

Terms like “reasonable” or “justified” are comparative. It might be more reasonable for S\* to believe that p than it is for S to believe that p. Thus, in granting (in A4) that S is not unreasonable in believing p, it does not follow that S\* could not be more reasonable in believing that p. For example, if S has reasons for believing that p and S\* has those reasons plus some additional ones, then S\* is more reasonable in believing that p than S is in believing that p.

## Areas of Disagreement

Primary areas of disagreement concern the ability of finitism to solve the regress problem and the manner in which warrant originates and is transferred. The finitist holds that some belief, say the belief that there is a blue smear on a white surface, is autonomously warranted for me in virtue of (1) the visual experience of it being as if there is a blue smear on a white surface in good light a few feet in front of me and (2) the absence of any reason for believing that the visual experience is not veridical. The warrant that arises in such a fashion is transmitted to other beliefs by inference.

An infinitist should point out that even if warrant or justification could arise in that fashion and subsequently could be transferred by inference, such an account of warrant will not provide a basis for addressing the regress problem, i.e. to correctly describe how reasoning can increase the warrant for a proposition, say p. Why is p held to be true? A reason, say b, is provided. Simplifying matters, suppose b were autonomously warranted. The finitist claims that no further reason for b is required in order for S to be epistemically rational in holding p. The infinitist should respond in two ways:

- 1 The infinitist is *not* denying that b is actually justified to some degree or that p is justified by b to some degree. The infinitist is claiming that S would be better justified in believing p on the basis of believing b, if S also had a further reason for

holding b. Such a reason is that b-type propositions are likely to be true in virtue of, say, general truths about the causal history of beliefs with b-type contents.

- 2 If it is agreed that S is better off epistemically when S has a reason for believing that b-type propositions are likely to be true, then the infinitist will point out that the regress of warrant-producing reasons does not stop at b. Infinitism can explain how the warrant of b-type propositions, and hence p-type propositions, can be increased in ways that cannot be explained by the finitist. Further, were the finitist to concede that warrant can be augmented for “basic” propositions by further reasoning, then on what basis can the finitist deny that reasoning can produce warrant in the first place?

The finitist will probably demur here for two reasons. First, it will be claimed that for *any* proposition, p, there will always be another proposition such that if S failed to believe it and S\* did believe it, S\* would be better warranted than S in believing that p (Ginet, p. 288). Thus, it is too easy to increase the warrant of a belief. Second, there comes a point in our reasoning where we cannot imagine what the next reason in the chain of beliefs could be (Ginet, p. 290).

There is some tension between these two claims for if one can give a general recipe for constructing a further reason in every case, then it is easily imaginable how the chain could continue. Nevertheless, the infinitist should reject both claims.

The first claim rests on providing a general recipe for constructing a further reason for any proposition, say p. I have discussed similar objections elsewhere and pointed out that it is not sufficient that there be such a proposition that can serve as a reason for p; it must also be “available” to S (see Klein, 1999, pp. 311–312). But Ginet’s objection is not so easily handled. He asks us to consider any proposition, p, and correctly points out that there will be another proposition available (in the sense that it is appropriately hooked up with S’s beliefs) that is such that if S were not to believe it, S would not be as well justified as S\* would be were S\* to believe it. That proposition is  $[p \vee (q \ \& \ \sim q)]$ . But it should be recalled that according to infinitism there are at least two necessary conditions of justification: (i) the Principle of Avoiding Arbitrariness (PAA), which generates an infinite series of propositions; and (ii) the Principle of Avoiding Circularity (PAC), which blocks circular reasoning. PAC was not the primary concern of Ginet’s response, because he grants that coherentism is not the correct view of warrant. Nevertheless, it is that second principle which is of use here. PAC is: for all propositions, x, if x is warranted for a person, S, at t, then for all y, if y is in the reason-ancestry of x for S at t, then x is not in the reason-ancestry of y for S at t. Now, since  $[p \vee (q \ \& \ \sim q)]$  is equivalent to p, a chain which included the former as an ancestor of the latter would violate PAC. Hence, S is not better justified in believing p were S also to believe that  $[p \vee (q \ \& \ \sim q)]$ ; but were S not to believe  $[p \vee (q \ \& \ \sim q)]$ , then S would not be as well justified in believing that p. In other words, not believing an equivalent proposition can lower the degree of justification, but believing it cannot increase the justification.

Now to the second claim. Ginet suggests that there is a point in the reason giving process such that no further reasons could be given.

With regard to many people who acquire a basic perceptual belief when they have appropriate perceptual experience, it may be plausible to suppose that, if they entertained the general proposition to the effect that when one has such experience the perceptual

belief is likely to be true, they would not only believe it but also take this tacit general belief to be part of their reason for their perceptual belief. But is it equally plausible to suppose that they have available in that same strong sense still another belief which is their tacit reason for holding that tacit general belief? What would it be? And what would be their tacit reason for *that* tacit reason? (Ginet, p. 290)

Recall the *types* of tacit reasons that have been adduced for holding the tacit general belief. Descartes was faced with just this problem in the *Meditations*, namely: Do we have any reasons for thinking that our perceptual equipment typically yields the truth? We know his type of answer: there are a priori reasons available that show that the equipment is reliable. The currently more fashionable type of answer is based upon a posteriori reasoning involving mechanisms posited by evolutionary biology. Thus, I suggest it is easy to imagine how the reasoning could continue because we have good examples of such reasoning. Will reasoning in support of *that* tacit reason ultimately beg the question? I don't believe it need do so and have argued for that elsewhere (see Klein, 2004).

### **Basic Difference: Reasoning Can Originate Warrant**

For Ginet, and many epistemologists, the primary reason for rejecting infinitism is that it is committed to the view that all warrant originates by inference (see Klein, 1999, pp. 310–311).

A more important, deeper problem for infinitism is this: inference cannot originate justification, it can only transfer it from premises to conclusion. And so it cannot be that, if there actually occurs justification, it is all inferential. (Ginet, p. 290)

This objection would be conclusive if inferential justification were correctly construed (as Ginet suggests) as analogous to instrumental value. Just as there must be non-instrumentally valued objects, there must be non-inferentially justified beliefs.

But infinitism rejects (or ought to reject) this analogy.<sup>2</sup> Infinitism, like the plausible versions of coherentism, depicts justification as emerging when the set of propositions that are appropriately adduced as reasons expands. Of course, were the foundationalist to insist on thinking of warrant as originating in some propositions and then being transferred by inference to other propositions, he or she would be begging the question at hand. For it is that very concept of warrant that infinitism is challenging.

### **Acknowledgments**

I wish to thank Anne Ashbaugh, Alvin Goldman, and Ernest Sosa for their discussions about the issues in this essay.

### **Notes**

- 1 All references to Ginet refer to his "Infinitism Is Not the Solution to the Regress Problem," this volume.

- 2 Ginet rightly criticizes my response to Dancy's objection on this point. I did not see clearly at that stage in the development of infinitism that it must reject the analogy. I have tried to correct that in Klein (2004).

## References

- Klein, P. (1999) Human knowledge and the infinite regress of reasons. *Philosophical Perspectives* 13: 297–332.
- Klein, P. (2004) Skepticism: ascent and assent. In *Philosophers and Their Critics*, ed. J. Greco, pp. 112–125. Oxford: Blackwell.

# Reply to Klein

*Carl Ginet*

## 1 Fred and Doris

In Klein's dialogue between Fred and Doris, foundationalist Fred doesn't do as well as I think he should. Let me revisit the scene and try to give him some help.

Let the proposition that Fred has asserted be

Little Benny has been in the living room with blue paint on his hands.

Doris asks Fred why he thinks this is so and Fred answers:

There's a blue smear on the wall.

And Doris asks Fred why he thinks that is so and Fred replies by asserting, among other things (such as that he has just come into the living room), the following:

(P) I see a blue smear on a white surface right in front of me.

If Doris has the temerity to ask Fred why he thinks *that* is so, Fred could with some justice reply, "What do you mean? Didn't you hear me? I said I see it right in front of me." But Fred is a parent, as well as an epistemologist, and has learned to be patient with insatiable askers of "Why?" So he instead says, "I have no reason for thinking that I see a blue smear on a white surface, in the sense of a premise from which I infer it, but it is eminently reasonable of me, as it would be of anyone, to think that I see such a thing when, as is the case, I am prompted to do so by the fact that

(C) My visual experience is as if I see such a thing and I am aware of no reason to think that my visual experience might in this case be misleading me.

Doris now asks Fred whether, if he is indeed being reasonable in believing P in circumstance C, isn't he also obliged to believe both of the following two propositions (should he entertain them)?

(R1) When C obtains, it is likely to be true that I see a blue smear on a white surface.

(R2) C does obtain.

And isn't it only because these beliefs are available to him as (good) reasons for believing P that he is justified in believing P?

I would advise Fred to reply as follows. I do believe R1 and R2 (now that I consider them) and I see that their truth provides reason to believe P. And it would be unreasonable of me not to do this, while still believing P and thinking my doing so to be justified. An equally good reason for believing P would be constituted by R2 together with

(R1\*) When C obtains, it is reasonable for me to believe that I see a blue smear on a white surface.

I am not persuaded, however, that the availability of either of these reasons for believing P constitutes my actual reason for believing P. But let that pass. Let us suppose, for the sake of this discussion, that the availability of justified belief in, say, R1\* and R2 was my reason for believing P and hence constituted (at least part of) my justification for believing it.

Doris now asks Fred whether, if he is justified in believing R1\* and R2, he must not have available further beliefs as reasons for those beliefs. If he were to listen to me, Fred would respond as follows.

No, I need no premise from which to infer R1\*. This is because R1\* is a basic a priori principle constitutive of the concept of justification for belief in a perceptual proposition like P: it is a principle such that understanding it requires accepting it.

Nor do I need a premise from which to infer R2. This is because the following is a basic a priori principle constitutive of the concept of justification for a conscious-state proposition like R2:

(R3) When C obtains and one is prompted by that fact to believe that C obtains, then that belief is justified.

It follows from this principle that, given that my belief that C obtains is prompted by the fact that C obtains, I am justified in that belief, whether or not I have available further beliefs that support it. It is clear that R3 could not serve as a premise in an inference justifying belief in R2 (that C obtains), because the other premise would have to be R2! R3 is not part of an inferential justification I have for believing that C obtains, but rather the basic principle of justification that entails that I am non-inferentially justified in believing that C obtains by the fact that my belief is prompted by C's obtaining.

So the lesson I draw from my version of the dialogue is that inferential justification of perceptual beliefs need not regress further than non-inferentially justified belief (or available belief) in conscious-state propositions like R2 and self-evident principles of justification like R1\*.



## **Inferentialism Drives Out Infinitism?**

In the last part of Klein's essay things take a startling turn. On p. 280 Klein says:

The infinitist holds that finding a reason for the questioned proposition, and then another for that reason, etc. places it at the beginning of a series of propositions each of which gains warrant and rational credibility by being part of the series. Warrant increases not because we are getting closer to a basic proposition but rather because we are getting further from the questioned proposition. ... Warrant, and with it rational credibility, increases as the series lengthens.

And in his last paragraph, in response to Audi's remark that even if there were an infinite series of propositions each of which is available, it is not "clear how infinite epistemic chains could account for any of our knowledge," Klein says:

Now, if knowledge required actually completing the series, knowledge would not be possible. But why suppose that knowledge requires the highest possible degree of warrant or absolutely credible belief? As the series lengthens, warrant and credibility increase. Nothing prevents it increasing to the degree required for knowledge.

In my essay I said that the (to my mind) most severe difficulty with Klein's infinitism is that it is committed to the thesis that inference alone can create justification. Here Klein seems to embrace this commitment wholeheartedly, holding that the longer the chain of inferential justification for a given belief the greater the justification created, and that, if the chain is long enough (but still finite), the justification can "increase to the degree required for knowledge." This seems to give us the result that knowledge does not require infinitely long chains of inferential justification after all: infinitism gives way to inferentialism. Worse yet, given Klein's thesis that inferential justification is the only sort of justification there can be, we seem to get the result that one could start with a belief (or set of beliefs) that is totally unjustified, because it lacks any inferential justification, and by spinning out a long enough chain of inference from it reach a belief that has the degree of justification required for knowledge.

These results are so counterintuitive that I hesitate to attribute them to Klein. But how else are we to interpret the quoted remarks?

## Chapter Twelve

# Can Evidence Be Permissive?

Roger White defends a negative answer to this question. He begins with what is known as the Uniqueness principle: the total evidence relevant to proposition P permits one and only one attitude towards P. Philosophers who deny Uniqueness are permissivists. They hold that there are possible cases in which the total evidence regarding P allows for more than just one attitude towards P; it might, for example, allow believing P and suspending judgment about P. According to what White calls strong permissivism, one and the same body of evidence might even permit believing P and permit believing  $\sim$ P. Using a variety of thought experiments involving belief-inducing pills, White argues that strong permissivism should be rejected. In his response to White, Thomas Kelly distinguishes between *intrapersonal* and *interpersonal* slack. If my total evidence permits two different attitudes towards P, that's intrapersonal slack; if you and I have the same total evidence and our shared evidence permits one attitude for you and another for me, that's interpersonal slack. Suppose you value collecting truths more than avoiding error; I value error avoidance more than truth collection. In that case, our shared evidence might make believing P permissible for you and suspending judgment about P permissible for me. Further avenues towards forms of interpersonal slack are open to subjective Bayesians. Hence, Kelly argues, there is a significant gap between principles against intrapersonal slack and principles against interpersonal slack, a gap that he thinks is not easy to bridge.

## Evidence Can Be Permissive

*Thomas Kelly*

Roger White's official statement of the thesis that he defends reads as follows:

**Uniqueness:** If an agent whose total evidence is E is fully rational in taking doxastic attitude D to P, then necessarily, any subject with total evidence E who takes a different attitude to P is less than fully rational.

Following Roger, I'll call someone who denies Uniqueness a *permissivist*. In what follows, I'll argue against Uniqueness and defend Permissivism.

## 1 The Strength of Uniqueness

At an intuitive level, one immediate attraction of Permissivism is this: Uniqueness is an extremely strong thesis. We can think of Uniqueness as one possible answer to the following question: How much *slack* exists between the evidence and what it's reasonable to believe given the evidence? In these terms, the friend of Uniqueness thinks that there is never any slack, ever. On the other hand, the permissivist thinks that in at least some possible cases, there is at least a little bit of slack. As this suggests, a permissivist might very well think that there are many cases in which there is no slack at all, where there is one and only one response to the evidence that's the fully rational response.

I mention this possibility – that Uniqueness is false, even though there are many non-permissive cases – in part because of my conviction that this is where the truth lies. Suppose that I pull a coin out of my pocket at random in order to flip it. I invite you to consider the proposition that *the coin will land heads rather than tails*. How much credence should you invest in this proposition? Here it's quite natural to think that, given plausible assumptions about your evidence, you should divide your credence evenly between this proposition and its negation, and that if you did anything other than that, you would be responding less than perfectly to your evidence. This natural verdict is one that a permissivist can embrace. (Although of course, not every permissivist will embrace it.) Moreover, a permissivist might clear-headedly hold that the great majority of cases are non-permissive, in the way that this one at least initially appears to be.

One respect in which permissivism is a very modest thesis, then, is that it's compatible with there being relatively few permissive cases. Another respect in which it's a very modest thesis is that the permissivist might think that what permissive cases there are, aren't all that permissive. At this point, it will be helpful to describe a realistic example that (unlike the coin case) seems to be a good candidate for a permissive case, at least as far as pre-theoretical intuition is concerned.

Suppose that six months before the US presidential election, it is quite unclear whether the Democratic or the Republican nominee will win. (Although it is clear that one or the other will.) I possess a large body of information that I take to bear on this question. Some of this information makes it more likely that the Democrat will win, while some of it makes that outcome less likely. On balance, I regard it as somewhat more likely that the Democrat will win than not, so I invest somewhat more credence in that proposition than in its negation. If I met someone who had exactly my evidence but was *extremely confident* that the Democrat will win, then I would regard this person as less reasonable than I am. (Perhaps he's in the grips of wishful thinking, or alternatively, pessimistic despair, and that accounts for why he's so confident.) Similarly, if I met someone who had exactly my evidence but thought that the Republican was

going to win, it would be natural for me to think that this person had made some kind of mistake in responding to our shared evidence. Suppose, however, that you and I agree on the basis of our common evidence that the Democrat is more likely than not to be elected. We similarly agree that although this outcome is more likely than the alternative, it's far from a sure thing. The only difference between us is this: you're a bit more cautious about the Democrat's prospects, and so give a bit less credence to the proposition that the Democrat will win than I do. Here there seems little pressure for me to conclude that you are less reasonable than I am. Moreover, the natural verdict about the case is that it's consistent with everything that's been stipulated so far that you and I might both be fully reasonable in our opinions about the election, despite the fact that those opinions are not identical. But if adding that further detail to the story does not render the story incoherent, then Uniqueness is false.

Again, someone might deny Uniqueness while thinking that what permissive cases there are resemble this one in relevant respects. So Uniqueness seems very strong. How strong is it exactly? Perhaps it matters here how we think about the psychological states to which it is taken to apply. To my mind, uniqueness seems most plausible when we think about belief in a maximally coarse-grained way, so that there are only three options with respect to a given proposition that one has considered: belief, disbelief, or suspension of judgment. On the other hand, as we begin to think about belief in an increasingly fine-grained way, the more counterintuitive Uniqueness becomes. Consider a thought experiment. Suppose that when we meet the Alpha Centaurians, they differ from us in only one important respect: they routinely take up doxastic attitudes towards propositions that are extremely fine-grained compared to our own. So, for example, the Alpha Centaurians really do have psychological states such as *believing to degree .5436497 that the Democrat will win*, or *believing to degree .5122894 that it will rain tomorrow*. I assume that this is a perfectly coherent possibility. (We might even have empirical evidence that they have such attitudes; it shows up in their betting behavior, and so on.) The friend of Uniqueness might insist that, for any possible evidential situation, the evidence in that situation singles out some one, exact degree of belief that it is uniquely reasonable for the Alpha Centaurians to have, any slight deviation from which already counts as a deviation from perfect rationality. Moreover, this will be so no matter how fine-grained we make the propositional attitudes of the Alpha Centaurians. But as one cuts up the psychology more and more finely, Uniqueness looks increasingly counterintuitive. Even if we are inclined to think that the epistemic facts (i.e., facts about what it's reasonable to believe, given the evidence) are *sharp* and not fuzzy, could there really be no limit to their sharpness? At some point, one wants to say, there must be a range of (presumably adjacent) mutually exclusive attitudes, any one of which would be reasonable to hold, and none of which is any more reasonable than any other within the range.

What should the friend of Uniqueness say about this? I think that the best move for her at this point is to appeal to so-called "mushy credence." It's not really that there is some range of permissible options. Rather, the uniquely reasonable thing for the Alpha Centaurians to do is to go vague over the ostensibly permissible range. On this way of thinking about it, one way of falling short of perfect reasonableness is to have overly precise degrees of belief: that amounts to treating your evidence as though it carries information that it doesn't carry. (And if the Alpha Centaurians are constitutionally

incapable of having these coarser attitudes, then they are constitutionally incapable of full rationality.)

Although natural, the appeal to mushy credence in order to defuse the challenge carries risks, inasmuch as whether the mushy credence picture is ultimately viable is currently the subject of intense debate.<sup>1</sup> I don't propose to enter into that debate here. Instead, I'll simply note that it seems that the friend of Uniqueness has strong incentive to hope that this vigorously contested issue is resolved in one way rather than another.

## 2 A Jamesian argument for Permissivism

What has been said so far concerns only the intuitive (im)plausibility of Uniqueness. But even if it would be surprising if Uniqueness turned out to be true, perhaps that's where the arguments lead. In this section, I'll sketch one argumentative route by which someone might arrive at the conclusion that Uniqueness is false. For reasons that I'll explain, I think that someone who arrives at the conclusion that Uniqueness is false in this way should not feel especially threatened by the kinds of arguments offered by Roger.

How then might the permissivist be thinking about things? Consider first a point emphasized by William James in his classic essay "The Will to Believe" (1897). James noted that philosophers often talk about the importance of attaining truth and avoiding error, but that such talk tends to mask certain complexities. On the one hand, there is the goal of *not believing what is false*, a goal that can be successfully achieved with respect to a given issue by suspending judgment on that issue. On the other hand, there is the goal of *believing what is true*, for which suspending judgment is obviously insufficient.

Moreover, as James also emphasized, these two cognitive desiderata can pull in opposite directions. In general, the more value one gives to not believing what's false about some issue, the more it behooves one to be relatively cautious or conservative in forming beliefs about that issue. That is, the more weight one gives to not believing something false, the more it makes sense to hold out until there is a great deal of evidence that *p* is true before taking up the belief that *p*. On the other hand, the more one values not missing out on believing the truth, the more it makes sense to take a somewhat more liberal attitude about how much evidence one expects before taking up the relevant belief. That is, to the extent that one is concerned to avoid *not believing *p* when *p* is in fact true*, one shouldn't wait until there is overwhelming evidence in favor of *p* before taking up the corresponding belief.

My suggestion is that James's observation is potentially highly relevant to our assessment of Uniqueness. Suppose that the evidence that you and I have that bears on some hypothesis *H* is *E*. Although it's clear enough that *E* supports *H* over not-*H*, it's not as though *E* is overwhelming evidence that *H* is true. Indeed, let's suppose that this is a marginal case, in that *E* is just *barely sufficient* to justify believing *H*: if *E* were any less supportive than it is, believing *H* on its basis would be positively *unreasonable*. Recognizing that *E* suffices to justify belief in *H*, I take up the belief in response. I notice, however, that you don't take up the same belief, despite having the same evidence. Let's further stipulate that it's not as though you are dogmatically averse to believing *H*, or anything like that: in fact, if the evidence for *H* grows any stronger, than you too will become an *H*-believer in response.

In these circumstances, is there any chance that your refraining from believing H is reasonable, given that my believing H is reasonable? As someone who believes H, am I committed to thinking that you're guilty of making some kind of mistake, that you've misjudged the probative force of our shared evidence? Before attempting to answer these questions, let's add one further detail to the story. With respect to the question at hand, you're a bit more concerned than I am to avoid believing what's false, while I'm a bit more concerned than you are to not miss out on believing what's true in virtue of suspending judgment. That is, there is a subtle difference in our cognitive goals, or rather, in the relative weights that we give to the two cognitive goals with respect to the question at hand.

Once this further stipulation is added, your not believing H on the basis of evidence that is only marginally sufficient to justify such belief seems eminently reasonable. As an H-believer, if I learned that we differed in our cognitive goals in this way, I would be disinclined to conclude that the manner in which you are responding to our shared evidence is unreasonable, even though it differs from my own. In fact, I might even think that if you were responding to the evidence in any other way than you are, then *that* would be unreasonable, given your cognitive goals. Moreover, notice that making such a judgment has no tendency to make me insecure in my conviction that I am also responding to the evidence in a reasonable way, given my cognitive goals. The upshot: subtly different ways of responding to the same body of evidence seem equally reasonable, given corresponding differences in the weights that we give to our shared cognitive goals.

Notice that this route to rejecting Uniqueness does not depend on thinking that "anything goes" with respect to the relative weights that can be permissibly assigned to the two cognitive goals, or even that there is much in the way of permissible variation here at all. So long as there are at least some possible cases in which it is reasonable for different individuals to give at least somewhat different weights to the goals, then this can affect how much evidence they should hold out for before they take up the relevant belief. There will then be possible bodies of evidence that fall within the relevant margin, bodies of evidence relative to which belief is a perfectly reasonable response on the part of the person who is somewhat more concerned to believe the truth, and relative to which suspension of judgment is a perfectly reasonable response on the part of the person who is somewhat more concerned to avoid believing what is false.

It might be objected that this route to rejecting Uniqueness depends on thinking about belief as an all-or-nothing matter, as opposed to a matter of degree. According to this line of thought, the "James point" only comes into play when one combines a fine-grained notion of evidence with a coarse-grained picture of belief. For once that combination is in place, then it seems like the following kind of threshold question is appropriate: How much evidence does one need that p is true before it becomes appropriate to believe p? (Presumably, just a little bit of evidence that p is true isn't enough.) And once questions about *where* the evidential threshold is located are put in play, it becomes natural to ask why the threshold is where it is, as opposed to someplace higher or lower. It is at this point that James's observation seems to become relevant, inasmuch as it is natural to think that one of the factors that can make a difference to where the threshold is located is the relative weight given to the two cognitive goals. Intuitively, as more relative weight is given to not believing what's false, that tends to exert some upward pressure on the threshold. (More evidence will be required, before

it makes sense to take up the belief.) On the other hand, as more weight is given to not missing out on the truth by suspending judgment, that tends to exert some downward pressure on the threshold.

The suggestion of the objector is that (i) James's observation about the potentially competing cognitive goals only gets traction against the background of this threshold picture, but that (ii) we can and should dispense with the threshold picture by doing epistemology in terms of credences or degrees of belief as opposed to all-or-nothing beliefs. Once we think in terms of more fine-grained doxastic states, there is no longer any question about where the threshold is, or which factors play a role in determining where it lies, because there is no need for a threshold at all. The only rule is: proportion your credence to the strength of your evidence. When one's evidence for *p* is very weak, one should invest very little credence in *p*; as one's evidence for *p* grows stronger, one's credence should rise accordingly. Thus, there is never any question about how much evidence one needs before belief (as opposed to suspension of judgment) is appropriate.

This is a tempting line of thought. In fact, for most of the time that I have been thinking about these issues, I believed that it was correct. I now think that it is mistaken. Joseph Rachiele (unpublished) argues compellingly that "the James point" holds even in theoretical frameworks that employ credences rather than all-or-nothing beliefs. For even if we do our theorizing in terms of credences, there will still be different dimensions relative to which we can evaluate the accuracy of those credences. Thus, one natural goal is that of minimizing the gradational inaccuracy of one's credences.<sup>2</sup> Relative to this goal, one set of credences is more accurate than another just in case it has a lower mean gradational inaccuracy. Another desideratum is that of lowering the *variance* in the gradational inaccuracy of one's credences. Even if one set of credences is superior to a second set in having lower mean gradational inaccuracy, the second set might be superior with respect to the variance property. Significantly, neither of these cognitive desiderata seems to be lexically prior to the other (Rachiele, unpublished, pp. 11–12). Although these two accuracy-related desiderata are complementary, the fact that they are distinct means that trade-offs will sometimes be necessary. (Compare: although the goals of believing truths and not believing falsehoods are complementary – doing well with respect to one is generally helpful with respect to the other – the fact that they are different goals creates the need for trade-offs; the optimal strategy for the achievement of one is not the optimal strategy for the achievement of the other.) On the plausible assumption that different individuals might reasonably differ, at least marginally, in how they resolve these trade-offs, different patterns of belief revision might be appropriate relative to the different resolutions. The upshot is that, to the extent that it works at all, the Jamesian route to vindicating a permissive epistemology sketched in this section works just as well in a framework that employs credences instead of all-or-nothing beliefs.

### **3 Interpersonal versus Intrapersonal Slack**

The permissivist should not rest her case on this Jamesian line of thought.<sup>3</sup> But even if it ultimately fails to undermine Uniqueness, I believe that there is an important lesson to be learned from it. The lesson concerns the need to distinguish sharply between

statements of Uniqueness that have what I will call *interpersonal import* from those that do not.

As noted above, someone who is impressed with James's point might think that the following kind of case is possible: if you are somewhat more concerned than I am to avoid believing what's false about whether *p*, and I am somewhat more concerned than you are to not miss out on believing the truth about *p* by suspending judgment, then there are possible bodies of evidence *E* such that:

- (1) The uniquely reasonable response for you is to suspend judgment about whether *p*, and
- (2) The uniquely reasonable response for me is to believe *p*.

Generalizing this, one might arrive at a view that is *permissive across individuals* but that is *impermissive with respect to the range of options open to any particular individual*. Someone who holds a view of this kind is prepared to countenance *interpersonal* slack (different individuals possessing the same evidence might believe differently, and each be reasonable in believing as they do), but deny the existence of *intrapersonal* slack (for any given individual, there is a uniquely reasonable thing for *her* to believe given her evidence). Roger's official statement of Uniqueness is clearly inconsistent with this kind of view; in this sense, it has interpersonal import. But other principles in the near neighborhood might lack such import. As a possible example, consider Roger's statement of Uniqueness in his seminal 2005 paper on the topic, which I will call Uniqueness\*:

**Uniqueness\*:** Given one's total evidence, there is a unique rational doxastic attitude that one can take to any proposition. (2005, p. 445)

On what I take to be its most natural reading – at least, its most natural reading when it is read in isolation – this principle says the following: *there is no slack for a single subject*. (Once you specify what her evidence is, that locks in what it is reasonable for her to believe.) But the principle is silent on whether some other individual with the same total evidence might take up a different attitude towards the same proposition that's fully reasonable. It thus lacks interpersonal import. When read in this way, Uniqueness\* is significantly weaker than Uniqueness, which explicitly rules out the possibility of interpersonal slack.<sup>4</sup>

Although principles that lack interpersonal import raise philosophically interesting questions in their own right, I believe that there are good reasons to think that the issue that philosophers have been concerned with in the literature on this topic concerns the truth of principles that *do* have interpersonal import, like Uniqueness. First, many philosophers (including Roger in his contribution to this volume) have suggested that there are important connections between this debate and the debate over the epistemic significance of disagreement.<sup>5</sup> And it is hard to see why a principle that did not have any interpersonal import would be thought relevant to the latter debate.

More importantly, certain views in epistemology that everyone would be inclined to treat as paradigms of “permissive” views seem to be consistent with uniqueness principles that lack interpersonal import. Consider, for example, a subjective Bayesian who thinks that the only rational constraints on one's doxastic corpus are the following:



(i) one's initial probability distribution must be coherent (beyond that, "anything goes"), and (ii) one must update one's credences by conditionalization upon gaining new information. The subjective Bayesian should presumably count not only as a Permissivist, but as an "Extreme Permissivist" in Roger's sense. For she thinks that even if you and I have exactly the same evidence, I might be extremely confident that the Democrat is going to win the election, and you might be extremely confident that the Republican is going to win (while both being perfectly reasonable). Nevertheless, the subjective Bayesian might very well accept Uniqueness\*, given a reading of that principle on which it lacks interpersonal import. Given the total evidence that I have, there really is one place that I should be, and if I were anywhere else, I would be less than fully reasonable. What the subjective Bayesian will deny is that it follows from this that *you* are less than fully reasonable, if you are somewhere else.

What I have argued for thus far in this section is the following. First, there is a significant gap between statements of uniqueness that have interpersonal import and those that lack such import: the former are significantly stronger than the latter, as witnessed by the fact that there are positions in contemporary epistemology with actual, flesh-and-blood proponents that are inconsistent with the former and consistent with latter. Second, the debate in the literature on this topic is really about whether the stronger principles are true. Notably, however, many of the kinds of considerations that friends of Uniqueness offer in its favor actually seem best suited to establishing the *weaker* principles, principles that lack interpersonal import. For example, both Roger's "arbitrariness argument" and his "arbitrary switching" cases invite us to consider how things look from the perspective of a *single subject*, and whether we can make good sense of the possibility that such a subject might be faced with a choice between incompatible but perfectly rational options with respect to his or her beliefs. As I understand them, these arguments have the form of *reductio ad absurdum* arguments. We are invited to suppose (for purposes of *reductio*) that a particular subject is in a permissive case and knows that she is. Roger then proceeds to ingeniously draw out the many apparent absurdities that seem to follow from these suppositions. For example, the subject might decide to switch her opinions randomly back and forth between the ostensibly permissible options, by popping a pill, or some other mechanism that has nothing to do with the truth, and then rationally maintain her latest opinion in the full knowledge that this is how she had arrived at it. We are then invited to conclude that this shows that there is something absurd about the original supposition, namely that there could be such cases.<sup>6</sup>

However, I don't think that arguments of this general form could possibly establish anything as strong as Uniqueness, a principle that has interpersonal as well as intra-personal import. This is because a theorist might very well *agree* with the conclusion that there is something incoherent or absurd about the supposition that a person could be in a situation in which she had rationally permissible doxastic options, while holding that some other person (say, someone with a different prior probability distribution) might reasonably believe something else on the basis of the same evidence. The kind of subjective Bayesian described above is an example of such a theorist. Notice that this possible combination of view is no mere occupier of logical space, something cooked up in order to avoid having to accept Uniqueness; rather, it follows immediately from independently motivated positions in epistemology that have prominent defenders.

Of course, that isn't the end of the story. If the kind of arbitrariness arguments put forward by Roger do suffice to establish that there is no intrapersonal slack, then one might attempt to argue from that intermediate conclusion or lemma to the stronger conclusion that there is no interpersonal slack, either. For example, suppose that the following bridge principle could be established:

BRIDGE: If it is currently reasonable for some subject S1 to hold doxastic attitude D1 towards P on the basis of evidence E, and it either is or would be reasonable for some other possible subject S2 to hold a different doxastic attitude D2 towards P on the basis of evidence E, then it is also currently reasonable for S1 to hold doxastic attitude D2 instead of D1 towards P on the basis of evidence E.

If the principle BRIDGE could be established, and if Roger's arguments suffice to show that there is no intrapersonal slack, then we could conclude that there is no interpersonal slack either, by reasoning in the following way:

If there were a case that was interpersonally permissive, then there would be a case that was intrapersonally permissive (by BRIDGE). But Roger's arguments show that there are no intrapersonally permissive cases. Therefore, there are no interpersonally permissive cases, either.

However, the principle BRIDGE is far from obvious. Indeed, many would flatly deny that it is true. In any case, it's the kind of thing for which we should insist on arguments. In the absence of actually looking at what arguments might be offered in its favor, it's difficult to say anything very definitive about the prospects for establishing it (or some sufficiently close principle). So here let me simply record my conviction that the gap between "no intrapersonal slack" and "no interpersonal slack" will not be an easy one to bridge, and that there will be plenty of promising points along the way for the permissivist to dig in her heels.

Notice, for example, that any reason that might be offered for thinking that conditionalization is the rule that governs belief change over time will cast doubt on intrapersonal slack (given one's initial prior probability distribution, and the evidence that one has accumulated since then, there is some particular probability distribution that one would have now if one were ideally rational), but *won't* be a reason for thinking that there is no interpersonal slack.

More generally, the fact that there are substantive coherence requirements that constrain permissible combinations of beliefs at the intrapersonal level (what I rationally believe constrains what else I can rationally believe), requirements that do not in general carry over to the interpersonal level (what I rationally believe does not constrain what you can rationally believe, in anything like the same way), generates obstacles for the project of arguing from the putative absence of intrapersonal slack to the non-existence of interpersonal slack. For example, in Roger's "belief toggling" cases, we are asked to place ourselves in the situation of an agent who can, by means of a pill, swap his current belief that p for a belief that not-p. (The case is designed to bring out the odd consequences of extreme permissiveness, or at least, of taking oneself to be in an extremely permissive case, in which believing either p or not-p on the basis of one's evidence would be perfectly reasonable.) But it seems that *everyone* – including extreme

permissivists – will have good reason to deny that one could end up with a fully reasonable belief that not-*p* in this way. After all, the proposition *p* stands in logical and evidential relations to countless other propositions that are potential objects of belief (or disbelief) for me. So if I am currently a *p*-believer who is fully rational, the fact that I am fully rational depends in part on the fact that my belief that *p* perfectly coheres with a large number of other doxastic attitudes that I take towards other propositions. When I contemplate swapping my current belief that *p* for a belief that not-*p*, I should recognize this as a change that is bound to make me less coherent – and therefore, less rational – than I am now. This seems like a good reason to decline to take the pill. But the extreme permissivist can say this, along with everyone else. For it is enough for the truth of extreme permissivism if the following is possible: some other person with my evidence is fully rational in believing not-*p* rather than *p*. If there is such a person, then she will presumably differ from me a great deal in her doxastic states, inasmuch as her belief that not-*p* will cohere perfectly well with all of her other doxastic attitudes towards propositions that stand in logical and evidential relations to not-*p*. Of course, the extreme permissivist should also say that the fully rational not-*p*-believer has a good reason to decline to take a belief-toggling pill that will reverse her belief about whether *p*, inasmuch as such a change is bound to make her less coherent, and therefore, less rational, than she is now.<sup>7</sup>

My advice to the permissivist, then, is that she should resist the slide from

Given that my evidence is *E*, there is some doxastic attitude *D* that is the only fully rational doxastic attitude for me to take towards proposition *p*

to

Given evidence *E*, there is some doxastic attitude *D* that is the only fully rational doxastic attitude for anyone to take towards proposition *p* (including all of those with different prior probability distributions, or those who assign different weights to the cognitive goals, etc.)

But Uniqueness requires the truth of the latter claim.

## 4 Evidential Support

In addition to considerations having to do with arbitrariness, Roger also offers an argument that appeals to the nature of evidential support:

1. Necessarily, it is rational for *S* to believe *P* iff *S*'s total evidence supports *P*.
2. If *E* supports *P* then necessarily *E* supports *P*.
3. It cannot be that *E* supports *P* and *E* supports not-*P*.
4. Therefore, if an agent whose total evidence is *E* is rational in believing *P*, then it is impossible for an agent with total evidence *E* to rationally believe not-*P* (p. 314).

Notice that this argument is directed at extreme permissiveness, so one could accept it while consistently denying Uniqueness. Nevertheless, it is worth exploring how the argument might be resisted.

One point of potential resistance that will appeal to many is this: the argument relies on the assumption that the relation of evidential support should be understood as a two-place relation (“E supports P”) as opposed to a three-place relation (“E supports P relative to background Z”). It is uncontroversial that whether a particular *piece* of evidence supports a given hypothesis often depends on considerations of background knowledge or theory. In the context of Roger’s argument, however, what matters is whether the relation of evidential support should be understood as a two- or three-place relation when what is at issue is the bearing of one’s *total* evidence on particular hypotheses. Suppose that we take one’s total evidence E to include everything that one has learned. Notably, even on this inclusive understanding of what is included in E, orthodox confirmation theorists will insist that the relation of support should be understood as a three-place relation, inasmuch as whether evidence E supports P (or the extent to which it supports P) will depend on the agent’s initial probability distribution. A philosopher who thinks that the support relation is a three-place relation will thus insist on rewriting the premises of Roger’s argument to reflect that fact:

- 1 \* Necessarily, it is rational for S to believe P iff S’s total evidence supports p relative to S’s prior probability distribution.
- 2 \* If E supports P relative to a prior probability distribution then necessarily E supports P relative to that prior probability distribution.
- 3 \* It cannot be that E supports P relative to a probability distribution and E supports not-P relative to that prior probability distribution.

Once the premises are rewritten in this way, however, even an extreme permissivist can happily accept them, for she can then point out that the argument from 1\*–3\* to 4 is invalid. Rather, what follows from premises 1\*–3\* is something like the following:

- 4 \* If an agent whose total evidence E is rational in believing P given her prior probability distribution, then it is impossible for an agent with total evidence E and the same prior probability distribution to rationally believe not-P.

But this conclusion falls well short of the original conclusion 4 and is consistent with extreme permissivism. For even if specifying an agent’s total evidence and her prior probability distribution suffices to pin down some doxastic attitude as the uniquely reasonable one, it does not follow that merely specifying her total evidence suffices to do the same. More specifically, an extreme permissivist might hold that while an agent with total evidence E might be reasonable in believing (or investing high credence in) P given her prior probability distribution, another agent with the same total evidence might be reasonable in believing (or investing high credence in) not-P given his different prior probability distribution. (Here again the gap between “no intrapersonal slack” and “no interpersonal slack” is significant.)<sup>8</sup>

Of course, even if evidential support is in fact better understood as a three-place relation than a two-place relation, it doesn’t follow that extreme permissivism is true. For it might be that there are substantive rationality constraints on prior probability distributions (that is, constraints beyond that of coherence), constraints that guarantee that it is impossible for an extremely permissive case to arise. Even if *that* is true, however, it doesn’t follow that Uniqueness is true, for the rationality constraints might be

such as to allow for at least some moderately permissive cases. What would vindicate Uniqueness is if it turned out that there is some uniquely reasonable prior probability distribution, which at least in this context would be tantamount to thinking of the relation of evidential support as a two-place rather than a three-place relation. So what the permissivist should claim is this: (i) the relation of evidential support is best understood as a three-place relation, and (ii) there is no uniquely rational starting point for all agents. But of course, many philosophers are already committed to thinking exactly this.

I don't imagine that any of this is news to Roger. In fact, he is quite modest in his claims for the argument. As he puts it: "My point is just that avoiding this conclusion [i.e., 4. above] appears to require a departure from very natural ways of thinking about evidence and rationality (p. 315)." I think that that's completely fair. In particular, I think that (for example) understanding the evidential support relation as a three-place rather than a two-place relation *does* involve a certain "departure from very natural ways of thinking about evidence," inasmuch as much of our ordinary thought and talk about evidence suggests the latter understanding when taken at face value. However, the fact that this way of avoiding the conclusion of the argument involves a departure from very natural ways of thinking about evidence and rationality should not be confused with the claim that it is an ad hoc response to the argument, or even that it should be regarded as a costly one. After all, the fact that many contemporary philosophers think that (i) and (ii) are true is not attributable to their desire to avoid the conclusion of Roger's argument, or any similar argument. Rather, what popularity (i) and (ii) enjoy is largely due to a common perception that these are among the lessons to have emerged from the systematic investigation of the nature of confirmation that has been pursued by philosophers and others in the decades since World War II.

Here the general trajectory of confirmation theory in the twentieth century is perhaps significant. Carnap's original vision for an "inductive logic" was that of a system that would assign a unique "degree of confirmation" that would attach to any hypothesis given a particular body of evidence. (The fact that this was a desideratum is perhaps a testament to the naturalness of thinking about the relation of evidential support in the way that Roger's argument requires.) But Carnap ultimately abandoned this ambitious vision as unworkable, and he and many of those who followed him in the development of quantitative confirmation theory came to advocate more liberal accounts of confirmation. Thus, for many contemporary philosophers the assumptions about evidential support that are needed to resist Roger's argument are independently motivated: in replying to Roger's argument along the lines suggested here, such philosophers need not say anything that they did not already believe about evidence or rationality. From such a perspective, even if resisting the argument does involve a departure from a very natural way of thinking about evidence and rationality, whatever theoretical costs are involved in such a departure have already been judged worth paying.

Of course, perhaps those who embraced more liberal views of confirmation did so for bad reasons. Notably, Roger has recently attempted to rehabilitate a version of the Principle of Indifference, a project that many had written off as hopeless.<sup>9</sup> Success in that venture would undoubtedly lend the argument considered in this section a dialectical effectiveness that it currently lacks. For this reason as well as for others, the debate over epistemic permissiveness is surely a long way from over.

## Acknowledgments

This paper grew out of two APA sessions that took place seven years apart: in 2005, I served as a commentator on Roger's seminal paper "Epistemic Permissiveness," and in 2012 I took part in an invited symposium (along with Roger and E.J. Coffman) on "The Uniqueness Thesis." I'd like to thank the audiences present on those occasions for their feedback and Michael Titelbaum for organizing the symposium and inviting me to participate in it. I also discussed these issues in two seminars at Princeton University, one of which was co-taught with Bas van Fraassen; thanks to Bas and the students for much helpful discussion. Finally, special thanks to Roger for the intellectual stimulation provided by his work on this topic and our exchanges about it over the years. Although I've been thinking about his challenging arguments for a long time now, I certainly don't pretend to have gotten close to the bottom of them here.

## Notes

- 1 Recent critiques include Elga (2010) and (somewhat ironically, if my sense of the dialectic with respect to the permissiveness question is on the right track) White (2009). A recent defense of mushy credence is Joyce (2010).
- 2 If we measure credences with real numbers, we can measure the gradational inaccuracy of a credence by taking the absolute value of the difference between that credence and the actual truth value of the target proposition (where "the actual truth value of the target proposition" = 1 just in case the proposition is true, and 0 just in case the proposition is false). For a useful discussion of gradational accuracy, see Joyce (1998).
- 3 Having just noted why I am unconvinced by one natural objection, let me mention what I take to be a better (even if more idiosyncratic) reason for skepticism. James's point seems to depend upon thinking about epistemic rationality in a particular way. Specifically, it seems to depend on thinking that epistemic rationality is really a special case of instrumental or means-end rationality, namely instrumental rationality in the service of one's cognitive goals, goals such as believing what's true and not believing what's false. This is an extremely natural way of thinking about epistemic rationality, and I believe that it is widely accepted within contemporary epistemology (even if many of those who accept it do so only implicitly). Nevertheless, I think that there are good reasons to be skeptical of the general picture. On this, see my "Epistemic Rationality as Instrumental Rationality: A Critique" (2003).
- 4 Although I have yet to see the point appear in print, the importance of distinguishing between principles that have interpersonal import and principles that lack such import in discussions of uniqueness is one on which a number of us have apparently independently converged, including Lee (manuscript), Meacham (manuscript), and Rachiele (manuscript).
- 5 See, for example, Feldman (2006), Christensen (2007), Kelly (2010), Ballantyne and Coffman (2011, 2012), Matheson (2011), and Cohen (2013). On the epistemology of disagreement, see especially Christensen (2009), Kelly (2005), and the essays collected in Feldman and Warfield (2010).
- 6 It is sometimes objected to this style of argument that it will inevitably fall short of showing that there are no permissive cases; rather, at best it shows that even if there are permissive cases, one could never know that one was in one. In effect, the objection is that *I'm in a permissive case* might be a "blind spot proposition," in the sense of Sorensen (1988). Like Roger, I doubt that this objection ultimately has much force, inasmuch as the assumption needed

to close the gap in the argument, namely *if one were in a permissive case, then at least in principle one could know that one was* seems extremely plausible.

- 7 So perhaps we should think of the pills as altering not simply one's doxastic attitude towards the target proposition *p*, but as altering a large cluster of one's opinions, namely, all of those opinions about propositions that stand in logical or evidential relations to the proposition *p*. However, it's not obvious that such a change in the case is innocent, or that once the case is changed in this way it elicits the same intuitive responses that the original version was designed to elicit (at least in my case, it doesn't).
- 8 Meacham (manuscript) emphasizes the difference made by thinking about the support relation as three place rather than two place in the course of criticizing a similar argument in White (2005).
- 9 See White (2009). For a critique, see Meacham (manuscript).

## References

- Ballantyne, N. and Coffman, E.J. (2011) Uniqueness, evidence, and rationality. *Philosophers' Imprint* 11 (18).
- Ballantyne, N. and Coffman, E.J. (2012) Conciliationism and uniqueness. *Australasian Journal of Philosophy* 90 (4): 657–670.
- Christensen, D. (2007) Epistemology of disagreement: the good news. *Philosophical Review* 116 (2): 187–217.
- Christensen, D. (2009) Disagreement as evidence: the epistemology of controversy. *Philosophy Compass* 4 (5): 756–767.
- Cohen, S. (2013) A tentative defense of the equal weight view. In *The Epistemology of Disagreement: New Essays*, ed. D. Christensen and J. Lackey. Oxford: Oxford University Press.
- Elga, A. (2010) Subjective probabilities should be sharp. *Philosophers' Imprint* 10 (5).
- Feldman, R. (2006) Epistemological puzzles about disagreement. In *Epistemology Futures*, ed. S. Hetherington, pp. 216–236. Oxford: Oxford University Press.
- Feldman, R. and Warfield, T. eds. (2010) *Disagreement*. Oxford: Oxford University Press.
- James, W. (1897) The will to believe. In *The Will to Believe and Other Essays in Popular Philosophy*. New York: Longmans, Green, and Co.
- Joyce, J.M. (1998) A nonpragmatic vindication of probabilism. *Philosophy of Science* 65 (4): 575–603.
- Joyce, J.M. (2010) A defense of imprecise credences in inference and decision making. *Philosophical Perspectives* 24 (1): 281–323.
- Kelly, T. (2003) Epistemic rationality as instrumental rationality: a critique. *Philosophy and Phenomenological Research* 66 (3): 612–640.
- Kelly, T. (2005) The epistemic significance of disagreement. In *Oxford Studies in Epistemology*, Vol. 1, ed. J. Hawthorne and T. Gendler, pp. 167–196. Oxford: Oxford University Press.
- Kelly, T. (2010) Peer disagreement and higher order evidence. In *Disagreement*, ed. R. Feldman and T. Warfield, pp. 111–174. Oxford: Oxford University Press.
- Lee, M. (manuscript) Sharpening up the Uniqueness thesis.
- Matheson, J. (2011) The case for rational Uniqueness. *Logic and Episteme* 2 (3): 359–373.
- Meacham, C.J.G. (manuscript) Impermissive Bayesianism.
- Rachiele, J. (unpublished).
- Sorensen, R.A. (1988) *Blindspots*. Oxford: Oxford University Press.
- White, R. (2005) Epistemic permissiveness. *Philosophical Perspectives* 19: 445–459.
- White, R. (2009) Evidential symmetry and mushy credence. In *Oxford Studies in Epistemology*, pp. 445–459. Oxford: Oxford University Press.

## Further Reading

- Conee, E. (2010) Rational disagreement defended. In *Disagreement*, ed. R. Feldman and T. Warfield, pp. 69–90. Oxford: Oxford University Press.
- Goldman, A. (2010) Epistemic relativism and reasonable disagreement. In *Disagreement*, ed. R. Feldman and T. Warfield, pp. 187–215. Oxford: Oxford University Press.

# Evidence Cannot Be Permissive

*Roger White*

## 1 Permissivism and Disagreement

In figuring out what to believe we look to our evidence. What sort of guidance does the evidence give? How stringent are the constraints it puts on rational belief? To sharpen our question, consider Alice and Bob who happen each to have the same body of evidence E pertaining to some matter P. They are independent inquirers unaware of each other's opinions. Suppose further that Alice and Bob are *ideally rational agents*, and so their beliefs are rationally impeccable. Might their opinions on whether P differ? If so, by how much? Is it possible, given the description above, that Alice is convinced that P while Bob thinks that not-P? If not, might it still be that Alice is a little more confident that P than Bob is?

Proponents of the Uniqueness thesis will say that Alice and Bob must be of exactly the same opinion.

**Uniqueness:** There is just one rationally permissible doxastic attitude one can take, given a particular body of evidence.

Or more precisely,

**Uniqueness:** If an agent whose total evidence is E is fully rational in taking doxastic attitude D to P, then necessarily, any subject with total evidence E who takes a different attitude to P is less than fully rational.<sup>1</sup>

Those who deny Uniqueness are *permissivists* about epistemic rationality.<sup>2</sup> According to permissivism, the evidence leaves us with some leeway as to what to believe. (Permissivists may disagree about how much leeway is allowed in general and in particular cases.)

It will be useful here to have labels for a stronger and weaker permissive thesis.

**Strong permissivism:** There are cases in which it is rationally permissible to believe P, but it is also rationally permissible to believe not-P instead, given the very same evidence.

**Moderate permissivism:** There are cases in which there is more than one rationally permissible *degree of confidence* one can have in P, given the same evidence.



Call a case in which an agent's evidence E does not determine a uniquely rational attitude to take to a proposition P a *permissive case*. A permissivist believes that there are some permissive cases, but needn't insist that all cases are permissive, or that they are all strongly permissive.

The issue of permissivism is importantly related to the epistemological problem of *peer disagreement*. But the two issues need to be distinguished. Consider now Carol and Dave who likewise have the same evidence (and they know this), but are not ideally rational (or at least don't take themselves to be). While recognizing their fallibility they share a mutual respect for one another. Carol, for instance, judges Dave to be no more or less likely than herself in general to arrive at rational, and indeed correct, conclusions in response to evidence. As it happens, Carol and Dave reach opposite conclusions: Carol believes that P while Dave thinks that not-P. When Carol learns that Dave thinks differently, how should this affect her opinion about P? Once both are aware of their difference of opinion can they rationally maintain their respective views, or must they somehow converge? This is the central issue in the epistemology of disagreement.<sup>3</sup>

While the issues are distinct, how we answer the question of permissivism would appear to have consequences for the problem of disagreement. The exact relation is open to debate, but on the face of it permissivism allows you to "stick to your guns" in the face of disagreement. If either opinion is rationally permissible, why should it bother Carol that Dave thinks differently? If Uniqueness holds then the matter becomes more pressing. Given their disagreement and shared evidence it follows from Uniqueness that at least one of Carol and Dave is failing to respond rationally to the evidence. Unless Carol is entitled to pin the blame entirely on Dave, she has reason to suspect that her opinion is not fully rational. And that would appear to give her reason to doubt her opinion.

## 2 The Case against Permissivism

I will mostly focus on strong permissivism. Many of the points I make can be adapted to make a case against moderate permissivism also. In the interests of space I will not pursue these developments in detail here.

### 2.1 Evidential support

Here are three premises that make up an argument against strong permissivism.<sup>4</sup> First, we commonly appeal to relations of *evidential support* among propositions. We can ask, for instance, whether certain meteorological data support the hypothesis that human activity is a major cause of global warming. It is by virtue of these evidential support relations that an agent is rational in her beliefs. In short, it is rational for an agent to believe P just in case her total evidence E *supports* P.

Second, as I'll argue in what follows, these evidential support relations hold *necessarily*. If E supports P then necessarily E supports P. Someone might question the necessity claim as follows. That the gas gauge reads Full supports the conclusion that the tank is full. But it need not. Suppose we know that the gauge is stuck on Full, or even that the wiring is switched so that it tends to read Full only when the tank is empty. In these cases the gauge's reading Full seems to support no conclusion or the

opposite conclusion. However, I think the right way to think about the matter is that these are just cases in which our *total evidence* is different. Our evidence concerning the gas tank involves a lot more than just how the gauge reads. We have a lot of information about the purpose of gauges and their reliability. It is not clear that the reading of the gauge all by itself does much if anything to support the tank's being full. By varying the additional evidence, of course, we can vary which conclusion is supported. But when we consider the totality of what we have to go on in assessing the state of the gas tank – that it reads Full, that gas gauges are typically reliable, and so on – it is hard to make sense of the idea that all of *that* information might have supported a different conclusion.

The point can be supported by two further considerations. If evidential support were contingent it would be unclear how we can assess what our evidence supports. I am trying to assess whether it will rain today. I survey the evidence and consider which conclusion it supports. Suppose it could be that while the weather reports, color of the sky, and so on *actually* support the conclusion that it will rain, that very evidence *might* have supported a different conclusion, say that it will be sunny all day. How am I to tell whether I'm in a possible world in which this evidence does support rain rather than one in which it doesn't? Will some further evidence indicate this? But then is it a further contingent matter what this extra evidence supports? Surely the fact of the matter is as follows: In principle I can examine the total evidence and recognize that it supports the conclusion of rain. I can recognize this without further investigation because the evidence *necessarily* supports this conclusion. There may be worlds in which the very same evidence is *misleading* – worlds in which we have all the same evidence and yet it doesn't rain (the actual world might even be one of these). But there are no worlds in which it supports a different conclusion than it actually supports.

Suppose again that this were not so. Our total evidence E does in fact support the standard thesis C of climate change. But now suppose that that very evidence need not have supported this conclusion. If this were so then it would seem that industry lobbyists could in principle manipulate the world such that E supports not-C instead. Hence they could make it the case that we can rationally *believe* that carbon emissions do no harm (and rationally act on this!) without having done anything to prevent disastrous climate change. This seems absurd.

Third, evidential support is *unidirectional*. It cannot be that E supports P but also that it supports not-P. Whatever is evidence for P is evidence against not-P. If it could be that the evidence supports both P and not-P then apparently one could rationally hold both contradictory opinions at once. But that can't be right. Putting these points together we have a simple argument against strong permissivism.

1. Necessarily, it is rational for S to believe P iff S's total evidence supports P.
2. If E supports P then necessarily E supports P.
3. It cannot be that E supports P *and* E supports not-P.
4. Therefore, if an agent whose total evidence is E is rational in believing P, then it is impossible for an agent with total evidence E to rationally believe not-P.

The argument is valid. There are, as usual, sophisticated ways in which one might deny one or more of these premises. I will not try to explore all of these here.<sup>5</sup> My point is

just that avoiding this conclusion appears to require a departure from very natural ways of thinking about evidence and rationality.

## 2.2 Arbitrariness

As a jury member it is my responsibility to do my best to arrive at the truth about the defendant's guilt or innocence. The responsible way to go about this is to examine the available evidence and seek to proportion my belief to it. There are lazier ways of going about it. I have two magical belief-inducing pills, one is marked GUILTY, the other NOT GUILTY. Ingesting a pill will give me the labeled belief. Clearly it is irresponsible to take such a pill. Why? Arbitrarily taking a pill while I have no clue as to whether the defendant is guilty gives me only a 50 percent chance of arriving at the correct verdict. If instead I base my belief on the evidence I have a much better chance of getting at the truth. There is no guarantee of course – evidence can be misleading – but I have a much better chance. Indeed, if there is evidence available strongly supporting one verdict, then it is highly probable that it supports the correct verdict.

Now suppose that permissivism is true and that the present case is a permissive one and that I know this about the case. (Perhaps the Epistemology Oracle has revealed to me that this is one of those cases.) What reason do I have to form my belief by an examination of the evidence rather than just popping a pill? If either conclusion can be rationally held given the evidence, why not just randomly pick one? In a non-permissive case where the evidence directs us to a particular conclusion, following the evidence is a reliable means of pursuing the truth. It is hard to see how this could be so in a permissive case (if there could be such). In such a case forming a belief that is rationally permitted by the evidence leaves it underdetermined what my conclusion will be. Whatever ends up causing me to accept one conclusion over the other will involve some non-evidentiary factor. And how could some factor other than *evidence* increase my chances of arriving at a true conclusion? Think of it this way. Suppose a hundred fully rational agents are given evidence E. Now of course in a non-permissive case we can predict that they will all arrive at the same conclusion and they will likely be right. But in a permissive case surely we can't expect this. If either conclusion can be rationally held it would be natural to expect around a 50–50 split of opinions. In this case only about half of the inquirers will be correct in their conclusions. If I am one of these inquirers, how can I sensibly expect to be one of the lucky ones? So it appears that in a permissive case I have no good reason to form my beliefs by examining the evidence rather than just popping a pill. But this is absurd. Surely it is always wiser to let the evidence guide us in inquiry.

Here is a different way of thinking about it. Suppose I have no idea whether the defendant is guilty, but I take one of the pills, the GUILTY one say, and find myself with the conviction that he is guilty. Could I reasonably maintain my belief while recognizing that I formed it just by popping a pill? Surely not. Prior to ingesting the pill I must have only 50 percent confidence that the result will be a true belief. Once I believe that the defendant is guilty and recognize that this is what I believe, I can't coherently doubt that my belief on the matter is true. But wouldn't it be strange to conclude, "What a stroke of luck! I now believe that the defendant is guilty, and he is guilty. So I must have gotten lucky in selecting the right pill." If this doesn't seem bizarre enough, imagine taking *many* such pills by random selection from a bag. Could

you sensibly think that you are lucky enough each time to pick the right pill? It seems that it can't be rational to maintain my belief in these circumstances. But now according to permissivism, it may be that whichever pill I take my resulting belief is a rationally permissible response to the evidence. So it seems that a permissivist is committed to the surprising conclusion that I can rationally maintain my conviction that I have remarkable luck at randomly forming true beliefs.

It might be noted that as a result of taking the pill, my belief, while rationally permissible, will not be well-founded. While the resulting opinion will conform to the evidence, it will not be appropriately *based* on this evidence. Such a belief may be *propositionally* but not *doxastically* justified, and to this extent it will be defective. Might this explain why a permissivist should prefer reasoned inquiry over pill popping? I can't see that it does. In inquiry my first concern is to arrive at a *true* conclusion regarding the defendant's guilt. And it is not clear why I should be so concerned with having my beliefs appropriately based unless this is conducive to the goal of getting things right. Consider again a *non*-permissive case. Suppose that the Epistemology Oracle reveals to me that the belief I have just formed was not in fact based on my evidence, but rather formed in some random way (whether or not it is supported by the evidence). In this case it would be a matter of luck if my belief were supported by the evidence. And unsupported beliefs are less likely to be true. So if I realize that my belief was not based on the evidence I should doubt that my belief has evidential support and hence doubt that it is true. But here the worry stems not from the lack of appropriate basing as such, but rather from the doubt raised as to whether my belief is supported by the evidence. The Oracle may continue, "Luckily for you in this case your conclusion that P is the uniquely rational conclusion to draw from your evidence E, even though your possession of this evidence isn't what caused you to first believe P." Having heard this I can sensibly maintain my belief in P since if it is the only rational conclusion to draw from my evidence then it is likely to be true. The matter is different if permissivism is correct. If this is a permissive case then there isn't the risk that a randomly formed belief will be rationally impermissible given the evidence. Whichever doxastic attitude I end up with – believing P, disbelieving P, or suspending judgment about P – will be fine as far as the evidence goes. Nevertheless, the knowledge that my resulting belief is a rationally permissible response to the evidence can do nothing to suggest that my belief is true, since in this case rationality does not distinguish between the true and the false: both are deemed to be rational conclusions from the evidence.

In any event, we can sidestep worries about well-foundedness by building it into the belief pills. There are more sophisticated pills that not only induce a belief that P, but cause me to base my belief on the evidence in the appropriate way. In an alleged permissive case there are possible worlds in which I examine the evidence and thereby come to believe P basing it on this evidence. There are other worlds where I examine this same evidence and come to the conclusion that not-P, basing it on this same evidence. A P-belief pill just ensures that one of the former set of worlds is actualized. If I take one of these pills then I will have a well-founded rational belief and there should be no reason to abandon it.

But consider just how odd this situation is. I have no idea whether P is true and have no idea whether the belief I'm about to form by taking a pill will be true. I now choose a P-pill and hence know that soon I will believe P, but presumably I still have no idea whether my belief will be true (I know of course that I will soon *think* that my belief

is true). As the pill takes effect I can think to myself, “What do you know? P, and I believe P, so by a stroke of luck I picked the pill that gave me a true belief. I’m glad I didn’t take the other pill because if I had I would be mistaken, even though I would still be rational.” But now I can take a not-P-pill. This will revert my mind back into a state of considering E and lead me to rationally base my conclusion that not-P on E. At this point I will rationally think, “So, it turns out I was wrong in thinking that P. But that’s okay. I’m glad I took this not-P-pill and corrected myself.” And of course I can switch back again. Each time I toggle my beliefs in this manner I am relieved to find that my resulting opinion is true. The absurdity of this should make us wonder whether permissive cases are possible.

## 2.3 Reflection argument

The following seems to be a plausible general principle:

**Reflection:** If I know that tomorrow I will come to rationally believe P on the basis of new evidence without having lost any of my old evidence, then it is rational for me to believe P now.<sup>6</sup>

Consider now a different case. Instead of a P-pill and a not-P-pill we have a Truth pill and a Falsity pill. The Truth pill gives me a true belief regarding P and the Falsity pill gives me a false belief. (In either case the pill I take will cause me to base my belief on the available evidence.) You will flip a fair coin and feed me a pill depending on how the coin lands: Truth if the coin lands *heads*; False if it lands *tails*. (You will not let me know how the coin landed or which pill you have dropped in my mouth.) Suppose I come to believe P. Assuming that this is a permissive case my belief will be rational, and since I will know that P is the conclusion I have reached, I will also rationally believe that I have a true belief concerning P. But the same goes if I take the other pill and believe  $\neg P$ : from my rational belief in  $\neg P$  and my knowledge that I believe it I will rationally conclude that I have a true belief concerning P. So I can know in advance that regardless of which pill I take and whether P is true, upon examining the evidence and taking a pill I will rationally believe that my belief regarding P is true. It follows by reflection that I can rationally believe *now* that my future belief will be true. But I know it will be true only if the coin lands *heads*. It follows from strong permissivism and reflection that I can rationally predict in advance that this fair coin will land heads! But that is absurd.

The strong permissivist of course will just have to deny reflection.<sup>7</sup> Perhaps that can be done, but let’s note how plausible this principle seems. Suppose I know that tomorrow I will believe that P. There are circumstances in which this would be regrettable. Perhaps I know that I soon will have forgotten some important information that would otherwise prevent me from concluding that P. Or perhaps I suspect that I will go mad and believe (what I can now recognize to be) nutty things like P. In such a case I have reason to distrust my future opinion and may want to try to prevent myself from forming such a belief tomorrow. But suppose that none of this is the case and that I’m quite sure that everything will go swimmingly. I will just consider all the evidence and form a rational opinion on the matter. It is hard to see what reason I could have to prevent my future self from forming such an opinion. What better way is there to go

about inquiry than to believe rationally on the evidence? And if I willingly anticipate forming this belief, what is to stop me forming it now?

The situation seems even stranger when we consider the connection with decision making. Before the coin is tossed I will surely have credence  $1/2$  that the coin will land heads and hence that I will form a true belief concerning P. So I will naturally be unwilling to take a bet where I win \$1 if my resulting belief is true, but lose \$2 if it is false. But I know that once I take the pill I will rationally form some opinion and hence rationally conclude that my belief on the matter is true. At this point, according to strong permissivism, I will rationally take the bet that my belief concerning P is true. Surely I have no reason to be reluctant to take the pill since it will simply result in me forming a perfectly rational belief on the basis of more evidence than I have to go on now. So it seems that I must *now* approve of the bet that I will take once I've ingested the pill. But I can't now approve of this bet since I'm only 50 percent confident that the coin will land heads. It appears my attitudes are incoherent.

## 2.4 Arbitrary switching

Fleeing from a tiger escaped from the zoo, you come to a fork in the road. One path leads safely home, the other off a cliff. Given your evidence you rationally believe that path A will lead you home. But you take your case to be a strongly permissive one where a rational agent with your evidence might conclude that B is the safe path to take instead. You don't think that B is the safe path, but you think that you could rationally think so given your evidence. Why not then switch your belief? A quick pill could change your state to one of a rational and well-founded conviction that B is the path that leads home. It is hard to see what could be wrong with taking such a pill. As a permissivist you don't think the worse of someone who has the contrary opinion. So you can't think there would be anything wrong with you epistemically speaking if you had believed that B is the safe path to begin with. Why then be wedded to your actual belief? Of course once you take the pill and rationally believe that B is the safe path you will rationally take that path. But if you can see nothing wrong with taking an action with that result, you might as well just go ahead and take path B. The scenery on that route is rather more enjoyable after all. But that is just bonkers. Insofar as you have considered the evidence and concluded that A is the only path that leads home, you can't sensibly think that it would be okay to take the other path, or even to take a pill which will result in your going the other way. The challenge is to say why this is so in a way that is consistent with strong permissivism.

Here is what I take to be the most promising line of response to this challenge. It can be developed in more or less sophisticated ways. I will stick to a simple formulation to get across the key idea.<sup>8</sup> I have asked the strong permissivist, "Assuming that the contrary opinion about the paths can be rationally held with the same evidence, why wouldn't it be rational to take a pill that switches your opinion?" The permissivist has an answer that might seem like a no-brainer: "Because if I do that I'll fall off a cliff! I don't judge someone who thinks otherwise as necessarily any less *rational*. But they are sadly *mistaken*. And of course I don't want to be mistaken on this matter." There are subtle issues here. The permissivist has described her deliberation in a way that seems reasonable. But to answer the question "Why would it be irrational for you to take the pill?" we need to identify what it is about her situation that makes this

choice irrational. This cannot be the *fact* that path B runs off a cliff. It might not even be a fact. Even if B is in fact the safe path but the agent rationally thinks otherwise it would be crazy for her to change her belief on the basis of no further evidence. Instead she might say, "Well I *believe* that taking the pill would result in my falling off the cliff. And I don't want that!" But now it seems we don't have an adequate answer to my question. Sure, you happen to believe that changing beliefs would result in your death since you believe that B leads off a cliff. But the question is, why be wedded to that belief? Upon taking the pill you will no longer think that taking the pill has the result that you run off the cliff. Indeed you will be relieved to conclude that the pill has saved you from that fate. And according to strong permissivism you will be rational in concluding this. In general, it doesn't sound like an adequate answer to the question, "Why not act so as to make it the case that P, rather than Q?" to say "Because it's not the case that P: Q is the case." The answer to the question "Why shouldn't you change your belief from A to B?" can't be "Because I don't believe B; I believe A."

Still, perhaps there does seem to be something right about the agent's original response. In practical deliberation we are focused outward on the world. It does seem appropriate to think, "Doing A will result in my death. So I won't do it." The fact (if it is a fact) that doing A will result in my death can't be the fact in virtue of which it is irrational for me to do A. Nevertheless, it is appropriate in practical deliberation to ask myself whether doing A will lead to my death and to act on the conclusion that I draw. (The relevant question in deliberation is not whether I *believe* that doing A will lead to my death, but whether it *will* lead to my death. For it is my death that I'm worried about, not my psychological states.) So the permissivist may insist that by displaying her apparently sensible line of reasoning she reveals why she is rational in refraining from doing A.

The response still seems unsatisfying. Consider a case involving only moderate permissiveness. You don't know whether either path leads home. Perhaps they both do, perhaps neither does. But on examination of the evidence (carefully attended to as you flee from the tiger!) you are somewhat more confident that path A leads to safety than you are that B does. Naturally you choose A. But as a moderate permissivist you take it to be rationally permissible for you to be somewhat less confident that A is safe and maybe a little more that B is safe such that the balance tips over in favor of B. The question arises again: why not take a pill to effect these small changes in your degrees of confidence in these hypotheses? But here the answer cannot be, "If I did so I would take path B and die!" You don't even believe this. Perhaps you even believe that B *will* lead you home; you are just not as sure of this as you are that A will. Instead you may say, "I'm more *likely* to get home safely if I take path A." But how are we to understand "more likely"? You can't be thinking that the *objective chance* of making it home is greater. The relevant chances of getting home may be around 0 or 1 depending on where the path leads (if you're on the right path you can be sure to get to your home before the tiger catches up). So the only way you could believe that the chance of getting home via A is greater is if you thought that A leads home but B does not. But you don't think that in this case.

A natural way to understand it is in terms of evidential or epistemic probability, which amounts to something like *the evidence more strongly supports the conclusion that A leads home than that B does*. If you believe this to be the case then naturally you will be motivated to take path A. But this does not fit with permissivism. It can't

be fully rational to put more confidence in a hypothesis that enjoys less evidential support.

For a permissivist it looks as though the likelihood here can only be understood as subjective probability, or credence. This is a little odd, however, as it doesn't seem that when I have the thought that *A is more likely to lead me home* I am just reflecting on my own psychology. And anyway, the thought *my credence that A leads home is greater than my credence that B does* can't play the same role in practical deliberation as the thought *only A will lead me home* does. In decision making I am concerned with the consequences of my actions, not with my psychological states. There does not seem to be any fact that you can appeal to in your deliberations in this case to make sense of your reluctance to switch that is consistent with permissivism.<sup>9</sup> So we have an odd situation. If strong permissivism is plausible then a fortiori moderate permissivism is also. Yet our best response to the arbitrary switching objection works, if it works at all, only in defense of strong permissivism.

A second point to note here is that it is clearly not always appropriate to appeal to the fact that P as a reason to avoid changing my belief about P. Even if I rationally believe that it is path A that leads home, it is typically a good idea to be open to, or even seek out, further evidence on the matter. In doing so I open myself to the possibility of changing my opinion. For it may turn out that new evidence clearly suggests that it is B that takes me home, and being rational I will change my view accordingly. It would be absurd for me to reason, "Path B runs off a cliff. So anyone who believes otherwise, even rationally, on the basis of further evidence, is unfortunately mistaken and will come to a bad end. I must take whatever steps I can to avoid this myself by shielding myself from any new evidence on the matter."<sup>10</sup>

Or consider a variation on this case. I might simply come to think that I was confused in judging that A is the safe path. Without new evidence I might reassess the case and conclude that actually the evidence suggests that it is B that is safe. Perhaps I am so fixated on the plan to take path A home that I initially continue to think that it leads to safety even after I've come to think that the evidence suggests otherwise. It would be perverse to think, "Well, now I see clearly that the only *rational* opinion to hold is that path B will lead me home, given my evidence. However, the fact remains that it is only path A that leads home, so I should avoid changing my opinion even if the alternative opinion is rational." In each case it seems entirely appropriate to let my judgments about *what it would be rational to believe* override my convictions about *what is actually the case*. But if that is so, it is hard to see why, in cases I take to be permissive, I can't give myself permission to change my opinion to one that I judge to be rational even if false by my current lights.

Note further that the anti-permissivist has a good explanation as to why he is always open to obtaining new evidence and conforming his belief and action to it. While of course I take my current opinion that P to be true, I also judge that *conditional on my obtaining further evidence such that my total evidence supports the opposite conclusion*, my current opinion is likely to be false. So I am not concerned that obtaining new evidence will lead me away from the truth (although that is always possible, just as it is possible that I am now mistaken). Compare this with the permissivist's predicament. Suppose I believe P but have the opportunity to obtain substantial new evidence on the matter. I don't yet know what this new evidence will involve, but the



Epistemology Oracle informs me that it will remain rationally permissible to take either of the two contrary opinions on P in the light of the new evidence. Even though I know that I'm perfectly rational, there is no predicting which belief I will end up with if I examine the new evidence since the Oracle informs me that either belief can be rationally held on the new evidence. Upon examining this new evidence it might strike me that I was mistaken and I may (rationally) come to believe  $\neg P$ . If we are to take the response sketched above to the arbitrary switching problem ("I won't switch my belief from A to B because then my belief will be false"), then I must take it that I have reason to refuse to view the new evidence. For there is a good chance that viewing the new evidence will result in my believing what I now take to be false. I could perhaps try to steel myself against the possibility of changing my opinion. But even if I can do this, it only follows that there is no point in considering the new evidence as it will have no effect on my opinion. But refusing to consider new evidence is absurd. Surely I should always take the opportunity to form a rational opinion on the basis of more evidence than I currently have.

I will leave you with you with one last odd consequence of strong permissivism. You have a hundred beliefs B1, B2, ..., B100, which you take to be permissive cases. I offer you a belief-toggle pill for B1. (The pill replaces B1 with a belief in the negation of its content.) "No thanks," you reply. "While the resulting belief would be rational it would also be false. And I don't want that." I slip it into your coffee anyway. Perhaps you don't thank me for violating your wishes, but you are more than happy with the result. As you see it now, your new opinion is not only rational but true. You couldn't have been more wrong in thinking that my pill would lead you into error. Quite the opposite! Or so you must now rationally think. I offer you a belief-toggle pill for B2. "No thanks," you reply. "While the resulting belief would be rational it would also be false. And I don't want that." I slip it into your coffee anyway ...

At some point shouldn't you start to trust me more? It's not that you have any reason to suppose that I'm an authority on the subjects of B1–B100 that you should defer to. I haven't any clue about these matters. I just enjoy experiments with belief pills. Suppose I feed you belief-toggle pills for the first 60 of these beliefs. You must find the results rather surprising. Of course you are not surprised that you *think* that your new beliefs are true. You knew that would happen. But you are surely surprised to find that they *are* true, which is what you now must take them to be. You are surprised to find your new beliefs true for several reasons. First of all, you had every reason to suppose that your old beliefs B1–B60 were true. You were after all quite rational in holding these beliefs given your evidence. And you haven't lost any of your evidence, so you must still have every reason you ever had to suppose that B1–B60 were true. But it turns out that none of them were! (Or so you must rationally think, on permissivist assumptions.) Still, it has all turned out okay by a remarkably lucky coincidence. Fortunately for you, those 60 of your of 100 beliefs that were mistaken have been corrected by a fist-full of randomly chosen belief-toggle pills! What are the odds of that? The rest of your beliefs you take to be true. So had you taken more pills or less pills, or a different selection (e.g., by counting backwards from B100, or taking the even-numbered pills), then your resulting beliefs would be partially or mostly false. But you must take it that you were lucky enough that the bunch chosen happen to line up perfectly with your prior mistakes. Perhaps you have an Epistemic Guardian Angel looking out for you!

Common wisdom has it that examining the evidence and forming rational beliefs on the basis of this evidence is a good means, indeed the best means, to forming true beliefs and avoiding error. The result so far must shake your confidence in this wisdom. You formed a hundred rational beliefs upon careful examination of the evidence yet 60 percent of them were mistaken! In stark contrast, the method of arbitrarily toggling your beliefs seems to have worked wonderfully. It has a 100 percent track record of leading you to the truth so far. Could this be an accident? Perhaps then you should decide to trust that the pills will have an enlightening effect and take the remaining 40. But there is no way to coherently think so. If you think that the belief-toggle pill will lead you to the truth regarding P, then you must take your current belief that P to be false. You can change that belief to not-P, but then you must think that the pill will change your belief back to P, which you must think is false given that you believe not-P. There is no coherent option except to think that all future belief-toggle pills will have the unfortunate effect of changing your opinion for the worse, even though you think they've done just the opposite so far.

I offer you a belief-toggle pill for B61. "No thanks," you reply. "While the resulting belief would be rational it would also be false. And I don't want that." I slip it into your coffee anyway ...

## Acknowledgments

Too many people have helped me think about this topic. But the main ones to thank are Tom Kelly and Miriam Schoenfield. I presented some of this material at the Central APA in Chicago and at MIT. Matthias Steup provided helpful corrections and feedback on a draft.

## Notes

- 1 I take the label "Uniqueness" from Feldman (2007) who defends a position along these lines. Other defenses of Uniqueness can be found in Christensen (2007), Sosa (forthcoming), and White (2005).
- 2 Permissivism is clearly the majority view among epistemologists. And many views that are widely held seem to entail it. See Ballantyne and Coffman (2011), Douven (2009), and White (2005) for details. Some recent defenders of permissivism include Ballantyne and Coffman (2011), Bruckner and Bundy (2012), Douven (2009), Meacham (manuscript), Schoenfield (forthcoming), and Titelbaum (2010). Some of these contain helpful critiques of the arguments I make here (derived from my earlier work), which I haven't had the space to address.
- 3 Christensen (2009) provides a useful overview of the disagreement debate.
- 4 Here I am clarifying an argument I've presented before (White, 2005) which was recently challenged by Ballantyne and Coffman (2011).
- 5 The interested reader might look at Schoenfield (forthcoming), and Meacham (manuscript).
- 6 This is a modification of the principle defended by van Fraassen (1984). One source of potential counterexamples involves beliefs expressed with the essential use of indexicals like "now." Rather than get distracted by this issue we can assume that the proposition P that we discuss here involves nothing of the sort.
- 7 See Schoenfield (forthcoming) for a discussion of how to do this.

- 8 See Schoenfield (forthcoming) for a more subtle discussion.
- 9 For an account of probabilistic assertions that might appeal to the permissivist here see Moss (2013).
- 10 This is one version of the puzzle introduced by Kripke (2011).

## References

- Ballantyne, N. and Coffman, E.J. (2011) Uniqueness, evidence, and rationality. *Philosophers' Imprint* 11 (18).
- Bruckner, A. and Bundy, A. (2012) On "epistemic permissiveness." *Synthese* 188: 165–177.
- Christensen, D. (2007) Epistemology of disagreement: the good news. *Philosophical Review* 116: 187–217.
- Christensen, D. (2009) Disagreement as evidence: the epistemology of controversy. *Philosophy Compass* 4: 756–767.
- Douven, I. (2009) Uniqueness revisited. *American Philosophical Quarterly* 46: 347–361.
- Feldman, R. (2007) Reasonable religious disagreement. In *Philosophers without God*, ed. L. Antony, pp. 197–214. Oxford: Oxford University Press.
- Kripke, S. (2011) Two paradoxes of knowledge. In *Philosophical Troubles: Collected Papers*, Vol. 1, pp. 27–51. Oxford: Oxford University Press.
- Meacham, C. (manuscript) Impermissive Bayesianism.
- Moss, S. (2013) Epistemology formalized. *Philosophical Review* 122 (1): 1–43.
- Schoenfield, M. (forthcoming) Permission to believe: why permissivism is true and what it tells us about irrelevant influences on belief. *Noûs*.
- Sosa, E. (forthcoming) The epistemology of disagreement. In *The Epistemology of Disagreement: New Essays*, ed. J. Lackey and D. Christensen. Oxford: Oxford University Press.
- Titelbaum, M. (2010) Not enough there: evidence, reasons, and language independence. *Philosophical Perspectives* 24: 477–528.
- van Fraassen, B. (1984) Belief and the will. *Journal of Philosophy* 81: 235–256.
- White, R. (2005) Epistemic permissiveness. *Philosophical Perspectives* 19: 445–459.

## Chapter Thirteen

# Is Justification Internal?

Internalists claim that only what is internal to our minds can be the sort of thing that justifies our beliefs. Externalists reject this requirement. **In what sense could a justifier be claimed to be internal to the mind? The question could be answered by saying that what justifies our beliefs are good reasons, and that only mental items – such as sense experiences, intuitions, introspective states, or memories – qualify as good reasons.** The internalist constraint might be justified by claiming that (i) justified belief is responsibly held belief and (ii) only internal factors determine whether a belief is held responsibly. To defend externalism, John Greco rejects (ii). He argues that whether a belief is responsibly held is at least in part a function of its etiology (its causal history), and thus partially an external matter. Another defense of internalism is based on the claim that externalism makes the rejection of skepticism too easy. Greco's reply is that, vis-à-vis the task of rejecting skepticism, externalism is better situated since **internalism engenders regress problems** and thus makes it impossible to advance an effective anti-skeptical argument. In the last section of his essay, Greco argues that all interesting concepts of epistemic evaluation are externalist because they all make etiology relevant.

Richard Feldman defends internalism. Identifying justification with having good reasons, where whether one has good reasons is solely a function of what mental state one is in, Feldman argues that there are at least two interesting epistemic concepts: consistency among one's beliefs, and a belief's being justified by virtue of being supported by good reasons. **Feldman accepts that there is an important epistemically evaluative status – well-foundedness – that makes etiology relevant. But, Feldman argues, if having good reasons is part of well-foundedness, and well-foundedness is necessary for knowledge, then having good reasons is necessary for knowledge and therefore qualifies as interesting.**

# Justification Is Not Internal

John Greco

## 1 The Internalism–Externalism Debate in Epistemology

When we say that someone knows something we are making a value judgment – we are saying that there is something intellectually good or right about the person’s belief, or about the way she believes it, or perhaps about her. We are saying, for example, that her belief is intellectually better than someone else’s mere opinion. Notice that we might make this sort of value judgment even if the two persons agree. Suppose that two people agree that the earth is the third planet from the sun. Nevertheless, we might think that one person knows this while the other person merely believes it. If so, we are making a value judgment – we are saying that there is something intellectually better going on in the case of knowledge. Another way to put the point is to say that **knowledge is a normative notion**. There is something normatively better about the case of knowledge, as opposed to the case of mere opinion, or even the case of true opinion. Finally, saying that someone knows something is not the only sort of value judgment we can make about her belief. **Even if we agree that some belief falls short of knowledge, we might nevertheless judge that it is justified, or rational, or reasonable, or responsible**. In each such case, we are saying that there is something normatively better about the case in question, as opposed to the case of mere opinion, or even the case of true opinion.

Internalism in epistemology is a thesis about the nature of this sort of normativity. More precisely, it is a thesis about what sorts of factors determine the epistemically normative (or evaluative) status of belief. **Internalists claim that the epistemic status of a belief is entirely determined by factors that are relevantly “internal” to the believer’s perspective on things**. That is, when a person *S* has some belief *b*, whether *b* is justified (or rational, or reasonable, or responsible) for *S* is entirely a function of factors that are relevantly internal to *S*’s perspective. By contrast, the “externalist” in epistemology denies this. **The externalist says that the epistemic status of a belief is *not* entirely determined by factors that are internal to the believer’s perspective**. When internalism and externalism are characterized this way, a number of things become apparent.

First, internalism is a rather strong thesis, in the sense that it says that epistemic status is *entirely* a function of internal factors. By contrast, the denial of internalism is a relatively weak thesis. Externalism in epistemology holds that *some* factors that are relevant to epistemic status are not internal to the believer’s perspective. Second, it is apparent that there are a number of kinds of epistemically normative status, corresponding to a number of kinds of epistemic evaluation. As already noted, we can say that a belief is justified, or rational, or reasonable, or intellectually responsible, and **these need not mean the same thing**. It is possible, then, to be an internalist about some kinds of epistemic status and an externalist about others. Hence there are a variety of internalisms and a corresponding variety of externalisms.

Third, we get different understandings of internalism (and externalism) depending on different ways that we understand “internal to *S*’s perspective.” The most common

way to understand the phrase is as follows. Some factor F is internal to S's perspective just in case S has some sort of privileged access to whether F obtains. For example, a factor F is relevantly internal to S's perspective if S can know by reflection alone whether F obtains.<sup>1</sup> A related, though not equivalent, understanding of "internal to S's perspective" is as follows. Some factor F is internal to S's perspective just in case F constitutes part of S's mental life.<sup>2</sup> For example, a person's perceptual experience counts as internal on this understanding, since how things appear perceptually to S is part of S's mental life in the relevant sense. Also, any belief or any representation that S has about how things are would be internal on this understanding, since a person's beliefs and other representations are also part of her mental life. These two understandings are related because **it is plausible to think that one has privileged access to what goes on in one's mental life, and perhaps *only* to what goes on in one's mental life.** In that case, the two understandings would amount to the same thing for practical purposes. Internalism would then be the thesis that epistemic status (of some specified sort) is entirely a function of factors that are part of one's mental life, and to which one therefore has privileged access.

Finally, it is apparent that some varieties of internalism are initially more plausible than others. That is, some sorts of epistemic evaluation are obviously externalist on the above understandings. Most importantly, and perhaps most obviously, **whether a belief counts as *knowledge* is an external matter,** if only because a belief counts as knowledge only if it is true, and whether a belief is true is typically an external matter.

There is another reason why knowledge and many other sorts of epistemic evaluation must be understood as externalist, however. Consider that we can evaluate both persons and their beliefs in two very different ways. Broadly speaking, we can evaluate persons and their beliefs either from an "objective" point of view or from a "subjective" point of view. From the objective point of view, we can ask whether there is a "good fit" between the person's cognitive powers and the world. For example, we can ask whether the person has a sound understanding of the world around her, or whether she has a good memory, or accurate vision. Also from this point of view, we can ask whether a person's methods of investigation are "reliable," in the sense that they are likely to produce accurate results. Notice that when a person gets positive evaluations along these dimensions, her relevant beliefs will as well: her beliefs will be largely true, or objectively probable, or objectively well formed, or reliably formed. On the other hand, there is a second broad category of epistemic evaluation. This sort concerns not **whether a belief is objectively well formed, but whether it is subjectively well formed. It asks not about objective fitness, but about subjective appropriateness.**

Common sense tells us that these two sorts of evaluation can come apart. For example, suppose that someone learns the history of his country from unreliable testimony. Although the person has every reason to believe the books that he reads and the people that teach him, his understanding of history is in fact the result of systematic lies and other sorts of deception. How should we evaluate this person's beliefs epistemically? By hypothesis, they are not well formed, objectively speaking: they are based on lies and deceptions. Nevertheless, there are clear senses in which the person's beliefs might be subjectively well formed. If the person has been deceived through no fault of his own, we might fairly say that his beliefs are intellectually responsible, or perhaps epistemically rational. We are inclined to say similar things about the victim of a convincing hallucination. Suppose that Descartes believes there

is a fire before him, and that he believes this on the basis of vivid sensory experience embedded in a broad and coherent set of background beliefs. But suppose also that Descartes is the victim of a massive and systematic illusion. The illusion, we can imagine, is undetectable and occurs through no fault of his own. Again, any epistemic evaluation of Descartes's belief will fall into one of two broad categories, or some combination of these. The belief can be evaluated in terms of its objective fit (in which case it fares poorly), or it can be evaluated in terms of its subjective appropriateness (in which case it fares well).

And now to the point: internalism is pretty much a nonstarter with respect to evaluations of the first category. Evaluations from an objective point of view involve factors such as accuracy, reliability, and appropriate causal relations to one's environment. And these are paradigmatically external factors. That is, they are factors that cannot be understood as internal to our cognitive perspective, whether we understand "internal" in terms of privileged access or in terms of what goes on in one's mental life. This is why there are no internalist theories of knowledge. Knowledge, it would seem, requires both objective and subjective factors. Put another way, a belief counts as knowledge only if it is *both* objectively well formed and subjectively appropriate. But since the former sort of status involves external factors, knowledge itself is external. Internalism, therefore, is best understood as a thesis about the second broad category of epistemic evaluation: it is a thesis about what factors determine subjective appropriateness.

We have now arrived at the following understanding of internalism:

- (I) Whether a belief *b* is subjectively appropriate for a person *S* is entirely a matter of factors that are internal to *S*'s perspective.<sup>3</sup>

Suppose we use the term "epistemic justification" to name the sort of subjective appropriateness that is required for knowledge. A standard form of internalism says that (I) holds with regard to epistemic justification (for example, see Ginet, 1975; Chisholm, 1977; Conee and Feldman, 2001). Alternatively, one might think that (I) holds for other kinds of subjective appropriateness, independently of their connection to knowledge. For example, one might think that rationality is an important and independent normative property, and that whether a belief is rational is entirely a matter of factors that are internal to *S*'s perspective.

I will argue that internalism is false in all of its varieties. More exactly, I will argue that internalism is false in all of its *interesting* varieties – it is false as a thesis about any interesting or important sort of normative epistemic status. Most importantly, internalism is false as a thesis about epistemic justification, or the kind of subjective appropriateness that is required for knowledge. That is bad enough, but in fact the situation for internalism is much worse. Internalism is false as a thesis about any interesting or important sort of epistemic evaluation, and any corresponding sort of epistemic normativity. In section 2, I will look at three considerations that are commonly put forward in favor of internalism as a thesis about epistemic justification, and I will argue that none of these adequately motivates the position. In fact, all three considerations motivate externalism about epistemic justification. In section 3, I will give a general argument against internalism in all its (interesting) varieties.

## 2 Three Motivations for Internalism

Three considerations are commonly put forward in favor of internalism, where internalism is understood as a thesis about epistemic justification, or the sort of justification that is required for knowledge. I will consider these in turn.

### 2.1 Epistemic justification as epistemic responsibility

The first consideration begins with an assumption about the nature of epistemic justification. Namely, a belief *b* is epistemically justified for a person *S* just in case *S*'s belief *b* is epistemically responsible. However, the argument continues, epistemic responsibility is entirely a matter of factors that are internal to *S*'s perspective. Therefore, epistemic justification is entirely a matter of factors that are internal to *S*'s perspective.<sup>4</sup>

The essentials of the argument can be stated this way:

1. A belief *b* is epistemically justified for a person *S* just in case *S*'s believing *b* is epistemically responsible.
2. Epistemic responsibility is entirely a matter of factors that are internal to *S*'s perspective.

Therefore,

3. Epistemic justification is entirely a matter of factors that are internal to *S*'s perspective. (1, 2)

Let us grant for the sake of argument that epistemic justification is a matter of epistemic responsibility. In other words, let us grant premise 1 of the argument. Nevertheless, premise 2 of the argument is false. Specifically, it is not true that epistemic responsibility is entirely a matter of factors that are internal to *S*'s perspective. Two sorts of considerations establish this point.

First, the notion of responsibility is closely tied to the notions of blame and praise. For example, judgments concerning whether a person is *morally* responsible with respect to some action or event are often equivalent to judgments about whether the person is morally blameworthy with respect to the action or event. Similarly, judgments concerning whether a person is *epistemically* responsible with respect to some belief *b* are often equivalent to judgments about whether the person is epistemically blameworthy with respect to *b*. And now the point is this: whether a person is epistemically blameworthy for holding some belief is partly a function of the person's prior behavior: if *S*'s reasons for believing *b* are the result of prior negligence, then *S* is not now blameless in believing *b*. An example will illustrate the point.

Example 1. Maria believes that Dean Martin is Italian. She believes this because she seems to remember clearly that it is so, and she presently has no reason for doubting her belief. But suppose also that Maria first came to this belief carelessly and irresponsibly (although she has now forgotten this). Many years ago, she formed her belief on the basis of testimony from her mother, who believes that all good singers are Italian. At the time Maria knew that her mother was an unreliable source in these matters, and she realized that it was not rational to accept her mother's testimony.



Clearly, Maria is not now blameless in believing that Dean Martin is Italian. Again, prior negligence is a factor determining present responsibility, and even if that negligence is not internal to S's perspective. Therefore, premise 2 of the argument above is false.

A second consideration also establishes that premise 2 is false. First, we can make a distinction between (a) *having* good reasons for what one believes, and (b) believing *for* good reasons. Anyone who knows the axioms of arithmetic *has* good reasons for believing a theorem in the system. But unless one puts two and two together, so to speak, one does not believe the theorem in question *for* the right reasons.<sup>5</sup> And now the point is this: a belief is epistemically praiseworthy only if it is believed for the right reasons. Two examples illustrate the point.

Example 2. A math student knows all the relevant axioms but doesn't see how the axioms support a theorem that must be proven on the exam. Eventually he reasons fallaciously to the theorem, and believes it on the basis of his fallacious reasoning.

Example 3. Charlie is a wishful thinker and believes that he is about to arrive at his destination on time. He has good reasons for believing this, including his memory of train schedules, maps, the correct time at departure and at various stops, and so forth. However, none of these things is behind his belief – he does not believe what he does *because* he has these reasons. Rather, it is his wishful thinking that causes his belief. Accordingly, he would believe that he is about to arrive on time even if he were not.

Clearly, the math student's belief about the theorem is not praiseworthy. Likewise, Charlie is not praiseworthy in believing that he will arrive on time.

The moral to draw from both examples is that "etiology matters" for epistemic responsibility. In other words, whether a belief counts as epistemically responsible depends, in part, on how the belief was formed. Since these beliefs were formed on the basis of bad reasons rather than good reasons, they are not epistemically praiseworthy. In fact, the same moral can be applied to example 1. Prior negligence also figures into the etiology of a belief, and is a factor in determining whether a belief is epistemically responsible. And of course, the etiology of a belief concerns factors that are external to the believer's perspective. Putting all this together, we may conclude that epistemic responsibility is not entirely a matter of factors that are internal to S's perspective. Accordingly, understanding epistemic justification in terms of epistemic responsibility does not motivate internalism about epistemic justification. In fact, it motivates externalism about epistemic justification (see Greco, 1990).

## 2.2 Like believers have like justification

A second consideration that is sometimes put forward in favor of internalism invokes a strong intuition about epistemic justification. Namely, in many cases it seems that believers who are alike in terms of internal perspective must also be alike in terms of epistemic justification. The point is often illustrated by considering Descartes's victim of an evil deceiver. Suppose that the victim is exactly like you in terms of internal perspective. Even if the victim lacks knowledge, the argument goes, surely his beliefs are as well justified as yours are. If you are justified in believing that there is a table before you, and if the victim's perspective is exactly as yours, then he must be justified in believing that there is a table before him.<sup>6</sup>

The considerations about epistemic responsibility above suffice to counter this line of reasoning, however. The problem is that two believers might be alike internally, and yet different regarding the causal genesis of their beliefs. And once again, etiology matters. Suppose that two persons arrive at the same internal perspective, but that one does so in a way that is epistemically responsible, whereas the other does so on the basis of carelessness, thick-headedness, and stupidity. The two persons will not be alike in epistemic justification, although they share the same internal perspective.

## 2.3 Replying to skepticism

A third consideration invoked in favor of internalism is that externalism makes an answer to skepticism too easy. Philosophical problems are supposed to be difficult. If the externalist has an easy answer to the problem of skepticism, this argument goes, then that is good reason to think that externalism is false. At the very least, it is good reason to think that the externalist has changed the subject, that he is no longer talking about our traditional notions of justification and knowledge.<sup>7</sup>

How does externalism make an answer to skepticism too easy? The idea is roughly as follows. According to the skeptic, one can know via sense perception only if one knows that sense perception is reliable. Similarly, one can know by inductive reasoning only if one knows that inductive reasoning is reliable. This creates problems for the internalist, because it is hard to see how one can mount a non-circular argument to the desired conclusions about the reliability of one's cognitive powers. There is no such problem for the externalist, however, since the externalist can deny the initial assumption of the skeptical argument. For example, an externalist can insist that sense perception gives rise to knowledge so long as sense perception *is* reliable. There need be no requirement, on an externalist account, that one *know* that one's perception is reliable. What is more, on an externalist account one seemingly *can* know that one's cognitive powers are reliable, and easily so. For example, one can use reliable perception to check up on perception, and then reason from there that perception is reliable. Similarly, one can use reliable induction to check up on induction, and then reason from there that induction is reliable.

In this context Richard Fumerton writes,

All of this will, of course, drive the skeptic crazy. You cannot *use* perception to justify the reliability of perception! ... You cannot *use* induction to justify the reliability of induction! Such attempts to respond to the skeptic's concerns involve blatant, indeed pathetic, circularity. (Fumerton, 1995, p. 177)

The fundamental objection to externalism can be easily summarized. If we understand epistemic concepts as the externalist suggests we do, then there would be no objection in principle to using perception to justify reliance on perception ... and induction to justify reliance on induction. But there is no philosophically interesting concept of justification or knowledge that would allow us to use a kind of reasoning to justify the legitimacy of using that reasoning. Therefore, the externalist has failed to analyze a philosophically interesting concept of justification or knowledge. (Fumerton, 1995, p. 180)

The problem with this argument against externalism is that it is self-defeating. In effect, the argument claims that only internalism can give a satisfying reply to traditional

skeptical concerns. On the contrary, I want to argue, if one concedes internalism then it is *impossible* to give a satisfying reply to traditional skeptical concerns. Specifically, if one concedes that epistemic justification is internalist, then the skeptic has all he needs to construct skeptical arguments that are otherwise sound. Put simply, **internalism about epistemic justification guarantees skepticism about epistemic justification**.

We may see the point if we consider non-basic knowledge, or knowledge based on evidential grounds.<sup>8</sup> Presumably, S knows p on the basis of evidence E only if E is a reliable indication that p is true. For example, consider the case where S believes that there is a bird in the tree on the basis of her sensory evidence. Presumably, S *knows* that there is a bird in the tree only if the sensory evidence that S has is indeed a *reliable* indication that there is a bird in the tree. That seems to be something that anyone should concede. But now something important follows from this. Namely, the reliability of S's evidence is one factor that determines whether S's belief has epistemic justification. On the assumption of internalism, then, S knows p on the basis of E only if E's reliability is something that is within S's perspective.

But now what does *that* mean? In what sense could the reliability of one's evidence be within one's perspective? Presumably, the internalist will have to accept something like this: in cases of knowledge, S's belief that "E is a reliable indication that p" is itself epistemically justified. But now how could that be? How could S have epistemic justification for this belief about her evidence? On the assumption of internalism, it is hard to see how she could.

Consider propositions of the form "E is a reliable indication that p." For example, consider the belief that such and such sensory evidence is a reliable indication that there is a bird in the tree. Clearly, this is itself a belief about the world. That is, it is a belief about the character of one's sensory appearances, and about the relationship between those sorts of appearances and real birds and trees. Now this sort of belief will not be knowable a priori. Rather, it is the sort of belief that is known, if at all, on the basis of empirical evidence. And therefore we are threatened with a regress or a circle. **That is, if S can know that there is a bird in the tree only if she knows that her evidence for this is reliable, and if she can know that her evidence is reliable only if both (a) she has evidence for *this*, and (b) she has evidence that this new evidence is reliable, then there would seem to be no end to this sort of problem. For presumably S's belief that her new evidence is reliable will require further evidence, and it will now be necessary that S know that *this* evidence is reliable, and so on.**

This problem was illustrated by the example of knowing that there is a bird in the tree on the basis of sensory appearances. But it is really a very general problem, which arises in any case where a belief about the world is based on empirical evidence. For in any such case, the belief that one's evidence is reliable will itself be a belief about the world based on empirical evidence, and so we will be off and running. For example, consider the belief that all crows are black, which is based on inductive evidence involving past observations of crows. If internalism is true, then one is justified in believing that all crows are black only if one is justified in believing that one's past observations are a reliable indication of one's present belief about crows. But this belief about one's evidence is itself a belief about the world, and will itself require empirical evidence.

One strategy for avoiding this sort of problem is to attempt an a priori argument to the effect that one's evidence is reliable. That is, one might try to show that one's evidence is reliable, but without using further empirical evidence to do so. That would

stop the regress in its tracks. But this strategy is a dead end. In principle, it would require showing that our sensory evidence *must* be a reliable indication of our perceptual beliefs, and that our inductive evidence *must* be a reliable indication of our inductive beliefs. But neither of these things is true. Rather, it is at most a contingent fact about us and our world, not a necessary fact about our evidence, that sensory appearances are a reliable indication of perceptual beliefs. Likewise, it is at most a contingent fact about us and our world, not a necessary fact about our evidence, that past observations are a reliable indication of unobserved cases.

The line of reasoning set out above is closely analogous to Hume's skeptical reasoning. Just like our internalist, Hume believed that one's empirical evidence gives rise to knowledge only if one knows that one's evidence is reliable. For example, Hume thought one must know that, in general, sensory appearances are a reliable guide to reality. Likewise, he thought one must know that, in general, observed cases are a reliable indication of unobserved cases. But there is no way to know such things, Hume argued, without reasoning in a circle. And so there is no way to know such things at all. The present point is this: if one adopts an internalist account of epistemic justification, then Hume has all the premises he needs to mount his skeptical argument. Put another way, there will be nothing else to challenge in Hume's reasoning. Here is that reasoning, set out more formally:

*Skepticism about perception*

- 1 All our perceptual beliefs depend for their evidence on (a) sensory appearances, and (b) the assumption (R) that sensory appearances are a reliable indication of how things are in the world.
- 2 But (R) is itself an assumption about how things are in the world, and so ultimately depends for its evidence on perceptual beliefs involving sensory appearances.
- 3 Therefore, assumption (R) depends for its evidence on (R). (1, 2)
- 4 Circular reasoning cannot give rise to knowledge.
- 5 Therefore, (R) is not known. (3, 4)
- 6 All our perceptual beliefs depend for their evidence on an assumption that is not known. (1, 5)
- 7 Beliefs that depend for their evidence on an unknown assumption are themselves not known.
- 8 Therefore, no one has perceptual knowledge. (6, 7)

*Skepticism about induction*

- 1 All our inductive beliefs depend for their evidence on (a) past and/or present observations, and (b) the assumption (R') that observed cases are a reliable indication of unobserved cases.
- 2 But (R') is itself an assumption about something unobserved, and so ultimately depends for its evidence on induction from past and/or present observations.
- 3 Therefore, belief (R') depends for its evidence on (R'). (1, 2)
- 4 Circular reasoning cannot give rise to knowledge.
- 5 Therefore, (R') is not known. (3, 4)
- 6 All our beliefs about unobserved matters of fact depend for their evidence on an assumption that is not known. (1, 5)

- 7 Beliefs that depend for their evidence on an unknown assumption are themselves not known.
- 8 Therefore, no one knows anything about unobserved matters of fact. (6, 7)

What is wrong with Hume's arguments? In each case, the independent premises of the argument are 1, 2, 4, and 7. Also in each case, 2, 4, and 7 seem uncontroversial. That leaves premise 1 as the only thing left to challenge. But if internalism is true, then in each case premise 1 is true.

Once again, a closer look at a motivation for internalism ends up providing a motivation for externalism. Internalism, we have seen, makes it impossible to reply to Hume's skeptical arguments. On the contrary, we can hope to avoid Hume's skeptical conclusions only by adopting externalism.<sup>9</sup>

### 3 The General Argument against Internalism

The arguments in section 2 are directed against internalism about epistemic justification, or the sort of subjective justification that is required for knowledge. We have noted, however, that it is possible to be an internalist about other kinds of normative epistemic status. More specifically, one might think that there are other sorts of subjective appropriateness, which are independent of knowledge and epistemic justification, but which nevertheless correspond to interesting and important kinds of epistemic evaluation. Perhaps this sort of normativity is at issue when we evaluate the beliefs of Descartes's demon victim. There is no question about whether the victim has knowledge, and we might even agree that he lacks epistemic justification. Still, we might insist, there is *some* sense in which the victim's beliefs are intellectually respectable, or up to par, or at least not objectionable. In this section I will argue that there is no important or interesting normative property that is also internalist. To be clear, I do not mean to deny that there are important epistemic properties that are independent of knowledge and justification. I think there are. My point is rather that no such property is internalist.<sup>10</sup>

The argument begins by recalling the two broad kinds of evaluation noted above. We said that, broadly speaking, we can evaluate persons and their beliefs either from an "objective" point of view or from a "subjective" point of view. **The objective point of view concerns (roughly) whether a belief has good objective fit with the world. From this point of view, we ask such questions as whether a belief is accurate, or reliably formed, or appropriately causally related to the facts. The subjective point of view concerns (roughly) what is subjectively appropriate to believe. From this point of view, we ask such questions as whether a belief is subjectively plausible, or responsibly formed, or well motivated.** The argument then proceeds as follows. First, evaluations from the objective point of view are obviously externalist. Considerations concerning accuracy, reliability, and causal relations involve factors that are paradigmatically externalist. But, second, evaluations from the subjective point of view are also externalist. For example, **the considerations in section 2 show that epistemic responsibility, no less than reliability, is a function of etiology. But the etiology of a belief is an external matter – it concerns such things as the history of the belief and the reasons why it is held, and these are things that are typically external to one's perspective.**

And now for the final premise of the argument: there is no interesting or important kind of epistemic evaluation that does not concern either objective fit or subjective appropriateness. Therefore, no interesting or important kind of epistemic evaluation, and no corresponding sort of epistemic normativity, is internalist.

A corollary of this argument is that all interesting kinds of epistemic normativity depend on factors related to accuracy and/or etiology. Of course, one can *stipulate* a kind of evaluation that abstracts away from these entirely. For example, we can stipulate that Mary's belief about Dean Martin is "weakly blameless" in the following sense: S's belief *b* is weakly blameless just in case S is no more blameworthy at the moment for believing *b* than she was a moment before. The present point is that this sort of normativity will not be interesting.

One way to see why is to look at the purpose of epistemic evaluation. It has often been noted that knowledge is a social product with a practical value. We are social, highly interdependent, information-using, information-sharing beings. As such, it is essential to our form of life that we are able to identify good information and good sources of information. In this context, it is not surprising that we make evaluations concerning how beliefs are formed, their history in relation to other beliefs, why they are believed, and so on. In other words, it is not surprising that we make evaluations concerning whether beliefs are reliably and responsibly formed. But evaluations of these sorts involve considerations about accuracy and etiology. And, therefore, evaluations of these sorts are externalist evaluations.

This context also shows why judgments that abstract entirely away from external factors will be uninteresting. For example, why should we care that Mary is no more blameworthy for her belief at the moment than she was the moment before? We care about whether Mary is, in general, a responsible and reliable cognitive agent. We also care about whether, in this instance, Mary arrived at her belief in a reliable and responsible way. We also care, of course, about whether Mary's belief is true. These are important considerations about Mary and about her belief – considerations that are important from the point of view of information-using, information-sharing beings such as ourselves. On the other hand, "time-slice" evaluations that abstract away entirely from the formation of beliefs, their relation to the world, and the character of believers will not be very important. Of course, we often want to abstract away from *some* external factors – we often want to abstract away from some or others. The point here is that we never want to abstract away from *all of them at once*. In other words, we have no interest in epistemic evaluations that are (entirely) internalist.

Consider the analogy to moral evaluation. We care about which people are good and which actions are right. That is, we care whether, in general, a person is a reliable and responsible moral agent. And we care about whether, in a particular instance, a person acted in a responsible and reliable way. What we don't care about is artificial, time-slice evaluations such as that S is not more blameworthy at the moment for bringing about some state of affairs than she was the moment before. Neither do we care whether some action *A* is right relative to S's own moral norms, in abstraction from questions about how S did *A*, or why S did *A*, or whether S's norms are themselves any good. Of course, we often want to abstract away from *some* external considerations – we want to abstract away from some or others. The point is that we never want to abstract away from *all of them at once*. In other words, we have no interest in moral evaluations that are (entirely) internalist.

These last points can be illustrated by applying them to a particular version of internalism. A number of philosophers have wedded internalism to evidentialism (for example, see Bonjour, 1985; Conee and Feldman, 2001). The main idea behind evidentialism is that a belief has positive epistemic status (of some sort or another) if and only if it is appropriately related to good evidence. Put another way, S's belief *b* has positive epistemic status just in case *b* "fits" S's evidence. The internalist adds to this that notions such as "good evidence" and "fit" are to be understood along internalist lines (for example, see Conee and Feldman, 2001, 1985).

Consider a case where S has a belief *b* and evidence *E*. For example, let *E* be a set of observations together with relevant background beliefs, and let *b* be the belief that all crows are black. There are many dimensions along which S's belief can be evaluated. For example, we can ask (a) whether *E* is true, (b) whether *E* is objectively probable, (c) whether *E* was reliably formed, (d) whether *E* was responsibly formed, (e) whether *E* leaves important information out, (f) whether *E* is a reliable indication that *b* is true, (g) whether *b* is objectively probable on *E*, and (h) whether S believes *b* because S believes *E*. All of these are external matters, involving factors that are neither part of S's mental life nor something to which S has privileged access in the typical case. Now certainly, in some situations we want to abstract from some of these matters. For example, we might be interested to know whether S's evidence is responsibly formed, whether or not S's evidence is true. Alternatively, we might be interested to know whether S's evidence makes her belief probable, whether or not her evidence was responsibly formed. But is there any situation in which we are interested to abstract away from *all* these factors at once? In other words, is there any situation in which we are interested to abstract away from *all* external factors? It is hard to imagine that there is.

Let us consider what is perhaps the most plausible possibility along these lines. We are sometimes interested to know, it might be suggested, whether S's belief is justified in the following sense: S believes *b* and S has *E*, and believing *b* on the basis of *E* would be licensed by the norms of evidence that S accepts. Remember, however, that we are supposed to be abstracting away from all external considerations. There is no question, therefore, whether S's norms of evidence are in fact reliable, or whether *E* is itself probable, or whether *E* was arrived at through prior negligence, or whether S believes *b* because S believes *E*, and so on. Abstracting away from all of this, why would we be interested to know whether *b* is licensed by norms of evidence that S accepts? Why would this be an important evaluation to make? It would be analogous to asking whether S's action *A* is licensed by the *moral* norms that S accepts, but independently of any questions about the adequacy of S's moral norms, or prior negligence by S, or the probable or actual consequences of *A*, or S's motives in performing *A*. As in the epistemic case, it is hard to imagine a situation in which that sort of moral evaluation would be interesting or important. Both moral and epistemic evaluations, we may conclude, are more closely tied to the world than that. They concern not just what is internal to one's perspective, but how that perspective is related to things outside it.

One apparent drawback of this argument is that it is difficult to fight over what is "important" or "interesting." However, in my mind the dispute between internalists and externalists comes down largely to just this issue. Accordingly, I have tried to put the ball in the internalist court by (a) focusing the argument right there and (b) making it explicit that internalist evaluations abstract away from *all* externalist considerations.

When we focus the dispute in this way, I think many will agree that the sorts of evaluations left over – the sorts that count as internalist – are not very interesting. Not in life, because they do not serve the purposes of information-using, information-sharing cognitive agents. Not in philosophy, because an internalist reply to traditional skeptical arguments is impossible, and so cannot serve that philosophical purpose.

## 4 Conclusion

We may now take stock of the arguments presented against internalism and in favor of externalism. In section 2, we looked at three considerations that are commonly put forward as motivations for internalism about epistemic justification, or the kind of justification required for knowledge. In each case, we saw that the consideration in question failed to motivate internalism. In fact, each consideration motivated externalism about epistemic justification. In section 3, we considered a general argument against internalism in all its interesting varieties. The argument was that all interesting epistemic evaluations are made from the objective point of view or the subjective point of view – they concern questions about objective fit, or subjective appropriateness, or both. But all such evaluations involve considerations about the accuracy of beliefs and/or their etiology, and these are paradigmatically externalist factors. The conclusion is that there are no interesting internalist evaluations. Put another way, all interesting epistemic evaluations are externalist evaluations.

## Acknowledgments

Thanks to Richard Feldman, Stephen Grimm, and Ernest Sosa for helpful comments and discussion.

## Notes

- 1 This is what Alston calls “access internalism” (see Alston, 1985, pp. 57–89; 1986, pp. 179–221). See also Ginet (1975), Chisholm (1977), and Goldman (1980, pp. 27–51).
- 2 This is Earl Conee’s and Richard Feldman’s understanding (see Conee and Feldman, 2001, p. 233).
- 3 If we understand “internal to S’s perspective” in terms of privileged access, we get:

(I-PA) Whether *b* is subjectively appropriate for *S* is entirely a matter of factors to which *S* has a privileged epistemic access.

If we understand “internal to S’s perspective” in terms of what goes on in S’s mental life, we get:

(I-M) Whether *b* is subjectively appropriate for *S* is entirely a matter of factors that constitute part of S’s mental life.

In the remainder of the essay I will ignore this distinction, since it is not important to the arguments that follow.

- 4 For arguments along these lines see Ginet (1975) and BonJour (1985).
- 5 For an extended discussion of this distinction and its importance, see Audi (1993, esp. chapter 7).
- 6 For arguments along this line see Foley (1984) and Luper-Foy (1988, p. 361).



- 7 See Fumerton (1995, esp. chapter 6). BonJour (1985, pp. 36–37 and 56–57) hints at this sort of objection. See also Cohen (2002).
- 8 Many internalists believe that all knowledge requires evidential grounds, in which case the following considerations apply to knowledge in general.
- 9 For further discussion along these lines see Greco (1999, 2000). An alternative response to Hume might challenge premise 2 of each argument, claiming that the assumption in question is known because it is reliably formed or because it meets some other externalist criterion for justification and knowledge. It is hard to see how the internalist could pursue an analogous strategy, however. For further discussion of this alternative response to Hume, see Greco (2002, esp. section V).
- 10 Another possibility is that there is some interesting status that is a component of epistemic justification. The arguments below will count against this possibility as well.

## References

- Alston, W. (1985) Concepts of epistemic justification. *The Monist* 68: 57–89.
- Alston, W. (1986) Internalism and externalism in epistemology. *Philosophical Topics* 14: 179–221.
- Alston, W. (1989) *Epistemic Justification*. Ithaca, NY: Cornell University Press.
- Audi, R. (1993) *The Structure of Justification*. Cambridge: Cambridge University Press.
- BonJour, L. (1985) *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.
- Chisholm, R. (1977) *Theory of Knowledge*, 2nd edn. Englewood Cliffs, NJ: Prentice Hall.
- Cohen, S. (2002) Basic knowledge and the problem of easy knowledge. *Philosophy and Phenomenological Research* 65: 309–329.
- Conee, E. and Feldman, R. (2001) Internalism defended. In *Epistemology: Internalism and Externalism*, ed. H. Kornblith. Oxford: Blackwell.
- Conee, E. and Feldman, R. (1985) Evidentialism. *Philosophical Studies* 48: 15–34.
- Foley, R. (1984) Epistemic luck and the purely epistemic. *American Philosophical Quarterly* 21: 113–114.
- Fumerton, R. (1995) *Metaepistemology and Skepticism*. Lanham, MD: Rowman & Littlefield.
- Ginet, C. (1975) *Knowledge, Perception and Memory*. Dordrecht: Reidel.
- Goldman, A. (1980) The internalist conception of justification. *Midwest Studies in Philosophy* 5: 27–51.
- Greco, J. (1990) Internalism and epistemically responsible belief. *Synthese* 85: 245–277.
- Greco, J. (1999) Agent reliabilism. In *Philosophical Perspectives* 13, ed. J. Tomberlin. Atascadero, CA: Ridgeview Press.
- Greco, J. (2000) *Putting Skeptics in Their Place*. Cambridge: Cambridge University Press.
- Greco, J. (2002) How to Reid Moore. *Philosophical Quarterly* 52: 544–563. Reprinted in *The Philosophy of Thomas Reid* (2003), ed. J. Haldane and S. Read. Oxford: Blackwell.
- Luper-Foy, S. (1988) The knower, inside and out. *Synthese* 74: 361.

# Justification Is Internal

*Richard Feldman*

Internalism in epistemology is a view about what sorts of things determine or settle epistemic facts. As its name suggests, it holds that “internal” things determine epistemic

facts. To say that the internal things “determine” the epistemic facts is to say that if two things are alike with respect to internal factors, then they must also be alike in the relevant epistemic ways. But we are not clear about just what internalism amounts to until we are clear about what counts as an internal thing and which epistemic facts are in question. In this essay I will say a little about what counts as an internal thing and considerably more about which epistemic facts are, according to internalism, determined by them. I will argue that internal factors do determine some epistemic facts, and thus my conclusion will be in stark contrast to the thesis John Greco defends in his essay, “Justification Is Not Internal.”

## I Background

To understand and assess the current debate about internalism and externalism in epistemology, it is useful to look briefly at the way in which the debate arose. A plausible starting point for this examination is the traditional analysis of knowledge and its rivals.

### A The traditional analysis of knowledge

In Plato’s dialogue *Meno* Socrates says, “That there is a difference between right opinion and knowledge is not at all a conjecture with me but something I would particularly assert that I know.” Lucky guesses make vivid the difference to which Socrates refers. If I am a contestant on a game show, I might have the belief that the valuable prize is behind door 1 rather than door 2 or door 3. Suppose my belief turns out to be true. Still, it is clear that I did not know that the prize was behind door 1 until the door was opened and I saw it. This prompts the question, “What is the difference between mere true belief and knowledge?”

Many philosophers thought that knowledge required true belief plus good reasons or good evidence. Thus, Norman Malcolm (1963) wrote, “Whether we should say that you knew, depends in part on whether you had grounds for your assertion and on the strength of those grounds.” Roderick Chisholm (1957, p. 16) wrote that “‘S knows that *h* is true’ means: (i) S accepts *h*; (ii) S has adequate evidence for *h*; and (iii) *h* is true.” These remarks are entirely typical.

The two quotations just presented reveal a generally shared outlook on what is required for knowledge in addition to true belief. Philosophers put the point in a variety of ways. Some say that knowledge requires adequate evidence, others say that it requires good reasons, and others say that it requires strong grounds. And, more commonly, philosophers have expressed this idea by saying that a true belief must be *justified* in order to be knowledge. Thus, a standard way to formulate the traditional analysis of knowledge is by saying that knowledge is justified true belief. This, then, is our starting point. I will sometimes refer to the traditional analysis as “the good reasons” analysis.

There are difficult issues to resolve if the traditional analysis is to be spelled out in detail. For one thing, one must say how much evidence is adequate (or how good the reasons must be or how strong the grounds must be). This detail will not affect the discussion that follows. Another issue has to do with what counts as reasons (or evidence or grounds). This will matter in what follows. It will be discussed briefly in section C, after some rivals to the traditional analysis are introduced.

## B Rival accounts of knowledge

Some philosophers have proposed analyses of knowledge that contrast sharply with the “good reasons” type analyses just described. **According to one of these alternatives, knowledge requires true belief plus an appropriate type of causal connection to the fact that the belief is about.** The details of the causal theory need not concern us here. In the words of one of its leading proponents, a central fact about it is that it “flies in the face of a well established tradition in epistemology, the view that epistemological questions are questions of logic or justification, not causal or genetic questions” (Goldman, 1967, p. 372). David Armstrong (1973, p. 166) defended a similar view. He said that the difference between knowledge and true belief was that knowledge involved “a *law-like connection* between the state of affairs [of a subject’s believing that p] and the state of affairs that makes ‘p’ true such that, given the state of affairs of [the subject’s believing that p], it must be the case that p.” The idea, then, is that whether you have knowledge depends not on what your reasons are, but rather on what the cause of your belief is. People know about the world because there is a causal connection between states in the world and their beliefs. For example, a person can know that it is cold out because the person’s belief that it is cold out is caused by its being cold out.

Another theory that differs significantly from the traditional good reasons view is reliabilism. **According to a simple version of reliabilism, a person has knowledge when the person’s true belief is caused by a method of belief formation that reliably (i.e., regularly) leads to true beliefs** (for details, see Goldman, 1979). Thus, reliable processes such as perception and memory yield knowledge, but unreliable processes such as wishful thinking and guessing do not yield knowledge, even on those occasions that they result in true beliefs. **Reliabilism allows that one can have knowledge even if one does not have any reason to think that the process that causes one’s belief is a reliable one and even if one has no evidence for the proposition believed. What matters is just that the belief be caused in a reliable way.**

Although there are differences between the causal theory and the reliabilist theory, for present purposes their differences are far less significant than their similarities. In what follows, I will sometimes refer to all theories relevantly like these two as “causal” theories.

## C A crucial difference between good reasons analyses and causal analyses

The discussion so far suggests that there are two distinct categories of theory about knowledge. One category includes the good reasons theories and the other includes the causal theories. A preliminary way to see the internalism/externalism debate is as a debate between defenders of good reasons theories (the internalists) and the defenders of causal theories (the externalists). However, as is so often the case, things are far more complicated than this initial formulation suggests. Before we look at why things are more complicated, it will be useful to make more explicit the key difference between good reasons theories and causal theories. This will bring to light something about what counts as an internal factor.

Good reasons theorists initially thought that what needs to be added to true belief in order to get knowledge is evidence or reasons. Reasons can include other beliefs, perceptual experiences, apparent memories, and so on. Consider your belief that there is a maple tree outside your window. According to the good reasons theorists, your reason for this belief might be another belief, say the belief that the tree has leaves of a particular shape. Or your reason might include the way the tree looks – how it appears to you. While we might ordinarily say that your reason for thinking that the tree is a maple is that its leaves are a particular shape, the fact that the leaves are that shape is not part of your evidence. What you are going on in judging the tree to be a maple is your belief that it has leaves of particular shape, and perhaps ultimately you are going on how the tree looks to you (your perceptual experience). These are internal, mental states you are in. As another example, suppose you believe that it is warm outside now. Your reason for this belief is your feeling of warmth. And this is an internal factor, while the actual outside temperature is not an internal factor. A person's reasons for a belief are things that the person can, at least in the typical case, describe to someone else and cite in support of the belief. In some cases, however, people may have reasons that they cannot describe, perhaps because they lack the proper vocabulary.

The idea, then, is that a person's reasons are the things the person has to go on in forming beliefs, and they will include how things look and seem to the person, the person's apparent memories, and the person's other beliefs. These are mental things. And internalism, at least as it will be construed here, is the idea that these mental things determine certain crucial epistemic facts. As noted earlier, this implies that if two people are alike with respect to the mental factors, then they must also be alike in the relevant epistemic ways. Thus, if an internalist holds that justification is an internal epistemic matter, then that internalist is committed to the view that if two people are mentally alike in all ways that bear on the justification of a particular proposition, then either they are both justified in believing that proposition or they are both not justified in believing that proposition. Equivalently, if they differ with respect to justification, then there must be an internal or mental difference.

Notably, the things that the causal theorists emphasized in constructing their theories are external factors. The fact that one's belief is causally connected in some particular way to some state of the world is not a fact internal to one's mind. Nor is it, by itself, an evidential fact. (More will be said about this topic in section IIIB.) Of course, one can have evidence about causal connections, but that is different. Thus, the good reasons theorists said that knowledge required true belief plus the right kind of internal factors, namely good reasons or evidence to support the belief. The causal theorists said that knowledge required true belief plus the right kind of external factor, namely a causal connection of the right sort.

As a first pass, then, we might take internalism to include theories relevantly like the good reasons theories and externalism to be theories relevantly like the causal theories. One way to state the idea is that internalism is the idea that knowledge requires justification whereas externalism is that the idea that it requires the right kind of causal connection rather than justification. But framing the issue this way is adequate only if we are careful about several key terms and we set aside some issues that some philosophers have seen as central. In the next section, I turn to those matters.

## II Clarifications

### A Knowledge and justification

The discussion so far conforms to one way philosophers use the word “justified.” This usage necessarily associates justification with good reasons and evidence. On this way of speaking, a belief is justified only if one has good reasons for it. Accordingly, when the word is used this way, we can say that causal theorists hold that knowledge requires true belief plus the right kind of causal connection, and not true belief plus justification. The preliminary statement of the internalism/externalism theory works out well on this usage (setting aside complications to be described later).

However, it is possible for causal theorists to use the word “justified” in a different way. They can agree with traditionalists that knowledge requires justified true belief, but differ with traditionalists over what is required for justification. These causal theorists would say that a belief is justified when it is caused in the right way and deny that justification always requires good evidence. When the word is used in this way, the dispute between internalists and externalists cannot be formulated as dispute about whether knowledge requires justification, since all sides agree that it does.

There is no point in arguing about the proper use of the word “justified.” What does matter is to be clear about which way we are using the word here. It will be somewhat more convenient in what follows to follow this first way of describing things. On this alternative, then, the controversy between internalists and externalists is, in its initial formulation, over whether knowledge requires justification.

### B Justification and well-foundedness

Supporters of versions of the traditional good reasons views acknowledge a distinction between merely having good reasons for believing a proposition and believing that proposition on the basis of those good reasons. The difference emerges in cases in which a person has good reasons for believing some proposition, but ignores or misevaluates those reasons, yet believes the proposition out of wishful thinking or on the basis of some mistaken inference. This has a moral analogue: doing the right thing for the wrong reasons. In the belief case, we might say that the person believes the right thing on the wrong basis. **Some philosophers will say that the person is justified in believing the proposition (since she has good reasons) but that she does not believe it justifiably (since she bases her belief on something other than good reasons). Another way to put the latter point is to say that the belief is not well-founded** (see Feldman and Conee, 1985, pp. 23–25). I will use this terminology here.

Internalists will agree that whether a person is justified in believing a proposition is an internal matter. It is less clear what to say about whether a belief’s being well-founded depends only on internal factors. At this point, some lack of clarity about exactly what counts as internalism emerges. Suppose a person is mixed up about what one of his own beliefs is actually based on. He has some good reasons for believing some proposition and he thinks that he believes it on the basis of those reasons. In fact, it is vanity or wishful thinking rather than those reasons that actually causes him to have the belief. This shows that he is justified in believing the proposition but his belief in it is not well-founded. Perhaps this makes well-foundedness not a fully internalist

notion. This depends upon whether the fact that his belief is based on wishful thinking rather than his good reasons counts as a mental (or internal) fact. It is hard to say. This a fact about the causal relations among one's mental states. Perhaps no causal facts count as internal. On the other hand, one might also say that these causal facts are about mental causes, and thus are internal to the person's mind. It is unclear, therefore, whether internalists will count well-foundedness as an internal matter.

Rather than attempt to resolve what really counts as internal, it may be better for present purposes to grant that well-foundedness is not a purely internal concept. **Still, a belief is well-founded only if the believer is justified in believing the proposition.** And **an internalist will argue that whether a person is justified in believing a proposition is determined by reasons or evidence, where these are internal matters in even the most restrictive sense.** Thus, their view is that justification is an internal matter, and it is necessary for knowledge. Externalists, presumably, will deny this. They will say that justification, so construed, is not necessary for knowledge.

## C The Gettier problem

Internalists have long recognized that a well-founded true belief can fall short of knowledge. That is, they realized that one can base a true belief on excellent reasons yet it can still be not a case of knowledge. It is easiest to see how this can happen by noticing first that well-founded beliefs can be false. Consider anything that you do know. Take as an example the proposition that your neighbor owns a Ford. You have excellent reasons for this: you have seen him driving a Ford, he often talks about the Ford he owns, he has proudly shown you a title to the car with his name on it, and so on. You know that he owns a Ford. Now, imagine an odd but possible alternative example. You have exactly the same reasons, with exactly the same reasons to trust your previously honest neighbor. But, in this alternative case, he is faking Ford ownership. The car he drives is owned by his wealthy uncle, the ownership papers he has are forged, and so on. He has pulled off an elaborate hoax. We need not speculate on his motives for this. In this second example, you have a well-founded belief that he owns a Ford, but your belief is false. You don't know that he owns a Ford, and the traditional analysis gets this case right. It implies that you don't know because your belief is not true. Of course, you did have knowledge in the original, normal, case. And the traditional analysis gets this example right as well.

There is, however, an even odder variation on the case. This case makes life difficult, but interesting, for epistemologists. In this odder case, your belief that your neighbor owns a Ford is true, but not for the reasons you have. Suppose that, in addition to the Ford your neighbor drives, his wealthy uncle also has recently bought a Ford and had the ownership of this other Ford assigned to your neighbor. So it is true that he is a Ford owner, but not for the reasons you are aware of. In this situation, you have a well-founded true belief that he is a Ford owner, but you lack knowledge of this fact. Examples like this one are called "Gettier cases," after Edmund Gettier (1963), who first brought them to the attention of epistemologists.

The reason why Gettier cases are important for the present discussion is as follows. **Philosophers who are firmly in the good reasons tradition agree that knowledge is not justified (or even well-founded) true belief. They think that there is something else to be added as well.** We need not dwell on the details of the possible fourth conditions

such traditionalists will place on knowledge. It will suffice to examine briefly one line of thought. One idea is that knowledge requires, in addition to justified true belief, that there be no *defeater* for one's belief. According to one prominent account, a defeater is a true proposition such that, if the person were justified in believing it, then the person would not be justified in believing the target proposition. The idea is that this is a missing truth that ruins the person's knowledge. So, on this view, knowledge is undefeated well-founded true belief.

Whether a belief is defeated or not is by all accounts something external to the mind of the believer. That is, things could be exactly the same from the perspective of two believers, yet one might be subject to a defeater and the other not. Two people could be internally alike, yet one is the victim of a Gettier case and the other is not. The examples about your neighbor's Ford illustrate this. Thus, good reasons theorists do not think that what must be added to true belief in order to get knowledge is merely something internal. **Instead, what they think is that what must be added is good reasons (which are internal) plus two other conditions: that the belief is based on those good reasons, and that there be no defeaters. The latter is definitely an external condition and the former may be, depending upon just how "internal" and "external" are understood.**

In modifying the traditional theory in these ways, good reasons theorists are not abandoning their original approach and accepting something like a causal theory. Good reasons remain essential for knowledge. We can, however, continue to interpret the internalism/externalism debate as the debate between the modified good reasons theories on the internalist side and the causal theories on the externalist side. The causal theorists may also have to say something to deal with the Gettier problem. Or, they may think that the causal connection condition replaces the combination of the justification condition and the condition internalists use to deal with the Gettier cases. In any case, the plausible internalist thesis is that **justification is an internal epistemic matter, while knowledge requires something external, such as lack of defeat**, to deal with the Gettier problem.

## D Deontology

Some philosophers identify internalism with the view that justification involves fulfilling obligations or duties. Similarly, some say that justified belief is epistemically responsible belief. For example, Alvin Plantinga (1993, p. 19) has written that one central internalist idea is that "epistemic justification is *deontological* justification.... All that it requires is that I do my subjective duty, act in such a way that I am blameless." Exactly how one connects this conception of justification to internalism is a matter of some controversy. I will not pursue it here. What is crucial to note is that the view that epistemic facts depend upon internal facts is logically distinct from the view that epistemic facts are in some sense matters of duty (for discussion, see Conee and Feldman, 2001). That is, it is one thing to say that justification is a matter of doing one's duty or doing what one ought to do. It is another thing to say that the relevant epistemic duty is determined by internal facts (such as evidence). It is possible to emphasize either the duty fulfillment aspect of the view or the internal factors aspect of the view when one characterizes what internalism is. The present discussion takes the second approach. In what follows it is not assumed that internalists are committed to a deontological approach.

As I and Earl Conee have argued elsewhere (Conee and Feldman, 2001, section I), any decision about what aspect of the traditional theories to emphasize in characterizing internalism is somewhat arbitrary. Internalism is the name given to the traditional theories. But those theories were stated prior to the names “internalism” and “externalism” being used to designate epistemological theories. Armstrong introduced the term “externalism” to characterize his causal theory and contrast it with the traditional theories. One could think that what **was crucial to the traditional theories was the fact that they made good reasons necessary for knowledge**. One could think that what was crucial was that they made justification a deontological matter. One could think that one of these aspects – say the deontology – was supposed to support the other. In this essay, I am identifying internalism with good reasons theories and their descendants, and setting aside the deontologism.

### III Internalism Vindicated

Before turning to my general defense of internalism, I want to respond to John Greco’s radical claim, in his contribution to this debate, that “all interesting epistemic evaluations are externalist evaluations.”

#### A Some internalist epistemic evaluations

It is, of course, difficult to argue effectively about which evaluations are “interesting,” since what is of interest to one person may not be of interest to another. In this section I will describe some evaluations that are internalist, at least in the way characterized earlier. I believe that these count as interesting evaluations, but I do not contend that they are among the necessary conditions for knowledge.

**1 Consistency.** A person’s overall set of beliefs can be evaluated for *consistency* and each individual belief can be evaluated in terms of whether or not it is consistent with the rest of the person’s beliefs. Although it is unclear exactly what counts as an epistemic evaluation, this would seem to be one. And it is internalist: it depends only upon the relations among the person’s beliefs. Of course, whether the beliefs are consistent is a logical matter, and it may be that people are in some cases ignorant of the logical relations among their beliefs. This shows that people may, for example, fail to realize that a new belief is inconsistent with some previously held beliefs. Still, whether the belief is inconsistent with another belief is an internal matter in at least the following sense: two people who are internally alike, in that they have the same mental states, cannot differ with respect to this property. If one has a belief that is inconsistent with the rest of her beliefs, then the other does as well.

**2 Identifiable good reasons.** A person might at times think about something he believes and ask whether he has any good reason for that belief. The ability to identify (or think of) such a reason is a mental fact about the person. Again, two people who are mentally alike in that they have the same beliefs and are able to identify the same reasons are alike with respect to this epistemic evaluation. And, once again, it is not assumed here that people are infallible with respect to whether they can identify a good reason. They may be mistaken about what counts as a good reason. Nevertheless, whether a person can identify such a reason is an internal mental fact about the



person. Thus, this is a second internalist epistemic evaluation. By and large, the ability to identify a good reason one has will coincide with the more familiar property of being able to state a good reason. But these two evaluations will diverge in rare cases in which a person has a reason but, somehow, is unable to say what it is. In assessing someone else's beliefs, we may often be interested in whether they can state reasons for them. But it is difficult to see why the ability to identify such reasons should be ruled out as "not interesting." We might wonder, for example, whether a person can think of any good reasons for some unusual beliefs, even if we know that the person is unable to communicate those reasons.

There are, then, some internal epistemic evaluations that are of interest. This refutes Greco's extreme thesis. But it leaves open the question of whether there are any internal evaluations that are directly relevant to knowledge. I turn next to that.

## B Knowledge without reasons

Some ordinary uses of the word "knows" clearly fit an externalist picture. We sometimes say such things as "The thermostat knows that it is time to turn on the furnace" or "The plants know that winter is over." Presumably, the truth about the thermostat and the plants involves only facts about the reliability of their responses to certain environmental factors. These uses of "knows" do not imply that the thermostat and the plants have any true beliefs, nor do they imply that they have good reasons for believing anything. One might liken ordinary attributions of knowledge to what is said in these cases, and affirm externalism as a result.

Arguably, however, these uses of "knows" are non-literal (see Conee and Feldman, 2004). We often use psychological terms in ascribing properties to things that do not actually have psychological states. We might say of a car that only runs on premium gas that it "does not like standard grades of gas" even though cars are not capable of literally liking or disliking anything. What we say is just a colorful way of saying that the car does not run well on standard grades of gas. Similarly, then, it may be that the knowledge attributions to the thermostat and the plants are not literally true. The thermostat does not *know* that it is time to turn on the furnace. It just is set up to turn on the furnace at the desired time. Similarly, the plants do not *know* that winter is over. They just come out of winter dormancy when winter is over. If this is right, then these uses of "knows" provide no support for externalism as we are currently understanding it.

Perhaps some epistemologists think that the uses of "knows" currently under discussion are literal and that they refute the traditional analysis. Perhaps they think that this resolves the debate in favor of externalism. However, at most it shows that not all kinds of knowledge require reasons, and perhaps that there is no particular kind of internal state that is necessary for all knowledge. (See Ernest Sosa's, 1997, discussion of animal knowledge.) This leaves open the possibility that we can plausibly divide all cases of knowledge into two kinds. One kind includes only the cases that are mere regular reactions to environmental stimuli. The other kind might be termed "discursive knowledge" or "reflective knowledge." It does seem to require good reasons, and thus to support a kind of internalism. So, it is open to internalists to argue that there is a kind of knowledge for which there is an interesting internalist necessary condition.

Finally, even if one does not say that there is a kind of knowledge that requires the possession of reasons, it is difficult to see why having reasons is not a matter of considerable epistemological interest in its own right. Arguably, this is what the traditional debate about skepticism has concerned (see Conee and Feldman, 2004, for discussion). That is, skeptics have asserted, and their critics have denied, that we have good reasons to believe many of the things we ordinarily believe. If we concede knowledge to the externalists, then skepticism is best transformed into a debate directly about the quality of our reasons. Since skepticism, so transformed, is interesting, this epistemic evaluation continues to be of interest.

## C Knowledge with reasons

While discussing an argument for internalism based on deontological considerations, Greco points out that “we can make a distinction between (a) *having* good reasons for what one believes, and (b) believing *for* good reasons” (p. 329). This is the same as the distinction drawn earlier between a proposition being justified for a person and the person having a well-founded belief in that proposition. Greco argues that for a belief to be responsibly formed, and thus justified according to a deontological view, it must be that it comes about in the right way. That is, it must be held for good reasons. And this, he assumes, is not an internal matter. I will grant that this is correct.

Greco makes similar claims about a view closer to the non-deontological internalism defended here. He considers the idea that two believers who are internally alike are equally well justified. He replies that

The considerations about epistemic responsibility above suffice to counter this line of reasoning, however. The problem is that two believers might be alike internally, yet different regarding the causal genesis of their beliefs. And once again, etiology matters. Suppose that two persons arrive at the same perspective, but that one does so in a way that is epistemically responsible, whereas the other does [... not.] The two persons will not be alike with respect to epistemic justification, even though they share the same internal perspective. (p. 330)

Here, “epistemic justification” must be used to mean “well-foundedness.” Given that, the conclusion can be granted. However, this fails to show that the two believers will not be alike with respect to having good reasons, the epistemic evaluation to which Greco himself called our attention. Indeed, it seems clear that if they are internally alike, then they are alike with respect to having good reasons, and thus internally alike with respect to this epistemic evaluation. Furthermore, having good reasons is necessary for knowledge. Hence, it follows that there is an internal necessary condition for knowledge – having good reasons, or justification.

This does not quite clinch the case for internalism, if the issue is whether there is an “important” or “interesting” internal necessary condition for knowledge. Since Greco himself calls our attention to the property of having good reasons, one might conclude that he thinks that it is interesting. However, I can think of two reasons for denying that justification is an interesting or important necessary condition for knowledge. Greco explicitly discusses only one of them.

Consider first the reason Greco does not discuss. An externalist might argue as follows. Consider two rival analyses of knowledge:

Analysis 1: Knowledge is undefeated justified true belief.

Analysis 2: Knowledge is undefeated well-founded true belief.

Philosophers who support anything along the lines of the traditional analysis will prefer Analysis 2 to Analysis 1. Thus, the interesting necessary condition for knowledge, the one that shows up in the best analysis, is well-foundedness rather than justification. It follows that well-foundedness, rather than justification, is the interesting epistemic evaluation in this vicinity, and it is an externalist evaluation. Justification, while internal, is not interesting.

This argument fails. Not all interesting necessary conditions for knowledge show up in a properly spelled out statement of the conditions for knowledge. Note that having good reasons, or justification, is necessary for well-foundedness. So having good reasons is necessary for knowledge. From the fact that justification is not listed as a separate or independent element in Analysis 2, it does not follow that justification is not an interesting genuinely necessary condition for knowledge. What shows up as an “independent” necessary condition of knowledge depends upon inconsequential details about how we happen to write out our analysis of knowledge. Here is a way to rewrite Analysis 2:

Analysis 3: Knowledge is undefeated justified true belief in which the belief is based on justifying reasons.

Given that well-founded beliefs are justified beliefs that are based on the justifying reasons, Analysis 2 and Analysis 3 are equivalent. Analysis 2 does not explicitly include a specifically internalist component. But this is because it makes use of well-foundedness. Analysis 3 spells out what this element depends on, and in doing so it does make use of an explicitly internalist element. Thus, the (presumed) adequacy of Analysis 2 does not show that knowledge does not have an important necessary condition that is internalist. It only shows that there is a way to write down this analysis that does not make explicit appeal to this element.

There is not the slightest reason to think that justification (or good reasons) is not “interesting” or not “important” just because we can restate the analysis of knowledge in terms that mask its presence. It is not that Analysis 2 makes having good reasons not necessary for knowledge. It just does not make that factor explicit. It puts it inside the “well-foundedness” requirement.

Greco presents a different reason for thinking that justification is not interesting in section 3 of his essay. He argues that all interesting epistemic evaluations are either objective or subjective. The objective evaluations have to do with truth or “fit with the world” and are not internal. The subjective evaluations are the ones that have the best chance of being internalist. But he argues that all such evaluations involve responsibility and this involves etiology and is therefore external as well. However, as I pointed out earlier, an evaluation in terms of whether one has a good reason – what I have been calling “justification” – does not depend upon either responsibility or etiology. I see no reason to reject this evaluation as “uninteresting.” Consider some beliefs of importance, say religious beliefs or beliefs that figure centrally in your views about morality. A question you might ask yourself is whether you have any good reason to think that these beliefs are true. The claim that the evaluation that results is uninteresting is without merit.

## D Externalist accounts of good reasons

The argument of the preceding section depends upon an assumption about reasons that should be made explicit. It is that what counts as a good reason is not an external matter. This is not to say that everyone always knows what counts as a good reason. It is just to say that external factors cannot make one thing a good reason in one case but not in another. External factors cannot make it the case that two believers who are internally alike differ in what they have good reasons to believe. Some externalists, such as Christopher Hill (1999), reject this assumption.

We can use an example Greco describes to develop this point. Suppose that a person learns the history of the country from a source that is unreliable but which he has every reason to trust. Greco says that there is a clear sense in which this person's historical beliefs are "subjectively appropriate." He goes on to argue that subjective appropriateness depends upon past factors. But the key thing to notice here is that in this example, the beliefs are subjectively appropriate when they are formed (since the person has every reason to trust his source) and they remain that way later on. This is just what an internalist would say about the case.

To reject the idea that this internalist evaluation is an "interesting" epistemic evaluation, one would have to say that the belief is not subjectively appropriate, either when formed or later on, simply because the source was not in fact reliable. Now, that is not a judgment Greco is prepared to make. However, it is possible that some externalists will say that the person has a good reason only if the source actually is reliable, no matter what information the person happens to have about its reliability. But this response is surely implausible. There is, of course, a difference between relying on a reliable source that one has reason to trust and relying on an unreliable source that one has equally good reason to trust. Yet there surely is something common to the cases, something favorable to be said of the person who does accept the word of both sources. The person who accepts what one such source says and rejects what the other says might, by chance, hit upon the truth. But such a person would surely not have reason on his side.

## E Forgotten evidence examples

Examples in which people forget their original basis for a belief provide a key element of some objections to internalism. I examine them in this section.

Greco describes an example of a person, Maria, who has a clear apparent memory that Dean Martin is Italian and no current reason against the proposition that he is Italian. However, she initially formed this belief irresponsibly and unjustifiably, relying on the testimony of someone she knew to be untrustworthy. She has, however, forgotten that this is her source. Greco's example resembles one put forward by Alvin Goldman (1999, section III) in a critical discussion of internalism.

Greco says that Maria is not blameless in holding the belief that Dean Martin is Italian. His point is that whether a current belief is blameless depends upon its history, not just the believer's current situation. This case comes up as part of his response to an argument for internalism that makes use of premises about blameless belief. I do not endorse any such argument and will not dispute this part of his evaluation of the case.

However, I do want to examine more carefully whether Maria's belief is justified. I will consider two possible responses open to internalists.

**Reply 1.** Maria's belief is justified, given that her current evidence does clearly support the proposition that Dean Martin is Italian. There is an important assumption that gives this response credibility, and it is worth making this assumption explicit. The assumption is that there is some attitude toward this proposition that it is reasonable for Maria to take (and, more generally, for any person to take toward any proposition the person considers). Suppose Maria considers the proposition that Dean Martin is Italian and wonders what attitude to take toward it. She has a clear memory of learning of this, and has good reason to trust her memory. She has, as the statement of the example makes explicit, no reason to think otherwise. It would be absurd for her to think, in spite of all this, that he is not Italian. So, disbelieving the proposition is clearly not a reasonable option, given the situation she is in. Perhaps a critic thinks that she would be most reasonable to suspend judgment. But this, too, is quite implausible. She has reasons to think he is Italian and no reason to think otherwise. Nothing competes with her reasons in favor. She might appeal to some general skeptical worries – one's memory can always lead one astray – but this is not relevant here. Thus, of the options open to her – believing, disbelieving, suspending judgment – believing is the only sensible option. Her belief is justified after all. This justification is determined by internal factors.

As noted, an assumption behind this argument is that some attitude or other is the reasonable one for her to take. This is a plausible assumption. How could it be that no option is reasonable, that whatever attitude she takes, it would be epistemically bad? Of course, it could be that no attitude will give her knowledge. She might not be in a situation in which she can know the truth of the matter. Still, some attitude or other must be reasonable for her. It is very hard to see how any attitude other than belief could be the one that is epistemically best in her situation. It is true that she made a mistake earlier. And perhaps this mistake would prevent her from having knowledge even if her belief is true. That is, if her belief is true, then, as Conee and I (2001) argue, this is a Gettier case.

I conclude that, whatever we say about blamelessness, Maria is currently justified in believing that Dean Martin is Italian, just as an internalist view suggests. Perhaps some readers will not be convinced. Perhaps they will insist that Maria's belief is a bad one, epistemically speaking. Another response suggests an additional point about the resources open to internalists.

**Reply 2.** Notice that as Greco describes the case, Maria previously had reason to distrust her source for this belief. Internalism implies that her belief was not justified when she still had this reason. Greco's view is that having this origin makes her belief not justified at the later time, after all information about its bad origin is forgotten. Suppose we go along with this. What exactly this would show about internalism is far from clear. Much turns on just what counts as internalism. **At most, what this argument shows is that past internal states matter. So suppose one holds that justification is a matter of one's history of internal states, not just one's current internal states. This presents a puzzling question: is such a view a kind of internalism?**

As I see it, there is no definitive answer to this, since what counts as internalism is at least in part a matter of stipulation. But it is worth noting that such a view does not vindicate anything very close to the causal theories. It differs from current state internalism only in that it makes past internal states matter. Reasons still matter. It is just that one's history of reasons matters, as well as one's current reasons.

## IV Conclusion

I conclude that the case for internalism in epistemology is very strong. The internalism in question is the view that certain interesting and important epistemic evaluations depend entirely on internal factors, namely reasons or evidence. There are, of course, epistemic evaluations that are not internalist. These include knowledge, being defeated (or undefeated), and, perhaps, well-foundedness. Still, justification, construed along traditional lines, remains an important necessary condition for knowledge. And this suffices to vindicate internalism.

## References

- Armstrong, D. (1973) *Belief, Truth and Knowledge*. Cambridge: Cambridge University Press.
- Chisholm, R.M. (1957) *Perceiving: A Philosophical Study*. Ithaca, NY: Cornell University Press.
- Conee, E. and Feldman, R. (2001) Internalism defended. In *Epistemology: Internalism and Externalism*, ed. H. Kornblith. Malden, MA: Blackwell.
- Conee, E. and Feldman, R. (2004) Making sense of skepticism. In *Evidentialism*, pp. 277–306. Oxford: Oxford University Press.
- Feldman, R. and Conee, E. (1985) Evidentialism. *Philosophical Studies* 48: 15–34.
- Gettier, E. (1963) Is justified true belief knowledge? *Analysis* 23: 121–123.
- Goldman, A. (1967) A causal theory of knowing. *Journal of Philosophy* 64: 357–372.
- Goldman, A. (1979) What is justified belief? In *Justification and Knowledge*, ed. G.S. Pappas. Dordrecht: Reidel.
- Goldman, A. (1999) Internalism exposed. *Journal of Philosophy* 96: 271–293.
- Hill, C. (1999) Process reliabilism and Cartesian skepticism. In *Skepticism: A Contemporary Reader*, ed. K. DeRose and T.A. Warfield. Oxford: Oxford University Press.
- Malcolm, N. (1963) Knowledge and belief. In *Knowledge and Certainty: Essays and Lectures*. Englewood Cliffs, NJ: Prentice Hall.
- Plantinga, A. (1993) *Warrant: The Current Debate*. Oxford: Oxford University Press.
- Sosa, E. (1997) Reflective knowledge in the best circles. *Journal of Philosophy* 94: 410–430.

## Chapter Fourteen

# Is Truth the Primary Epistemic Goal?

Justification can take many forms: epistemic, prudential, moral. How can we distinguish between them? The standard answer is that epistemic justification is related to truth in a way in which prudential and moral justification are not. But how is epistemic justification related to truth? One answer says that epistemic justification is a means to achieving the goal of believing what is true and not believing what is false. In his defense of truth as the primary epistemic goal, Marian David examines a variety of goals: the truth goal, the knowledge goal, and the justification goal. He argues that the truth goal is more basic than the justification goal, but cautions against conceiving of the connection between truth and justification as a means-end relation. And, he suggests, if we reconceive the truth goal as the goal of having beliefs that are non-accidentally true, then we may view truth as the ultimate epistemic goal.

In his opposing essay, Jonathan Kvanvig defends the view that truth is not the primary epistemic value. In addition to truth, there are, according to Kvanvig, many other epistemic values: knowledge, understanding, wisdom, justification, and making sense of something. Among these, truth is not primary but just one goal among others. Truth becomes the primary goal only when we adopt a conception of epistemology as the analysis of knowledge. From Kvanvig's point of view, this is unduly reductive. Epistemology, Kvanvig suggests, is more than that: it is the study of successful cognition. Although he rejects the view that truth is the primary epistemic goal, in the end he endorses an account that is not so different from David's. The fundamental function of cognition, Kvanvig proposes, is that of determining, for any  $p$ , whether  $p$ . Nevertheless, he insists, there are epistemic goals which are not subservient to the truth goal. One such goal is that of epistemic responsibility, a goal that, according to Kvanvig, we can conceive of without invoking truth.

# Truth Is Not the Primary Epistemic Goal

*Jonathan L. Kvanvig*

The question before us concerns the epistemic goal, standardly taken in epistemology over the past 50 years or so to be that of getting to the truth and avoiding error. In order to assess the plausibility of any answer to this question, it will be useful to begin by thinking about the question itself to make sure that it is properly understood. Asking about goals is asking about values or goods. To ask what one's goals are in, say, going to college, we inquire about the perceived values or goods that are being pursued. So the first thing to note about our question is that it asks about the values or goods that are epistemic in character.

The second point to note about the question is that it can be addressed from two quite different perspectives. One perspective is that of the theoretician. From this perspective, the question concerns what goods or values are central or primary for the theoretical task undertaken by the epistemologist, whatever that task may be. There is also another perspective, however, and that perspective is the point of view of those organisms about whose cognitive activity the epistemologist is theorizing. From this perspective, the question concerns the values or goods involved in the type of states and activities investigated by epistemologists.

It is important to keep the difference between these perspectives in mind when addressing this question, for there is no reason to assume that the answer to the question will be the same from either perspective. For example, it might be the case that truth is the primary good that defines the theoretical project of epistemology, yet it might also be the case that cognitive systems aim at a variety of values different from truth. Perhaps, for instance, they typically value well-being, or survival, or perhaps even reproductive success, with truth never playing much of a role at all.

Our question arises primarily from the perspective of the theoretical project of epistemology, and I will address the question from that point of view. We will see that the perspective of the cognitive system itself plays a role in this investigation, but only an ancillary one. In addressing the question before us I will be arguing for a negative answer to it. I will be arguing, instead, that there is a plurality of epistemic values and goals, and that though truth is an important epistemic goal, it has no claim to being the primary such value or goal. In order to argue for this plurality view, it is important to begin with a general account of the subject matter of epistemology, for the narrower one's conception of epistemology, the easier it is to defend the idea that truth is the primary epistemic goal. After explaining the appropriate domain of epistemological theorizing, I will take up the task of defending the pluralistic view.

## **What Is Epistemology?**

Epistemology is often taken to be the theory of knowledge, but that conception is too narrow. At the most general level of characterization, epistemology is the study of certain aspects of our cognitive endeavors. In particular, it aims to investigate



*successful* cognition. Within its purview, then, are various kinds of cognizing, including processes such as thinking, inquiring, and reasoning; events such as changes in one's world view or the adoption of a different perspective on things; and states such as beliefs, assumptions, presuppositions, tenets, working hypotheses, and the like. Also within its purview is the variety of cognitive successes, including true beliefs and opinions, viewpoints that make sense of the course of experience, tenets that are empirically adequate, knowledge, understanding, theoretical wisdom, rational presuppositions, justified assumptions, working hypotheses likely to be true, responsible inquiry, and the like.

Two notes of caution are in order here. First, not just any kind of inquiry into kinds of cognizing and varieties of cognitive success counts as epistemology. Presumably, one can investigate these issues scientifically as well as philosophically, and I assume that they can be investigated in other ways as well (religiously, politically, morally, aesthetically, etc.). Epistemology is the result when these issues are investigated philosophically, but I will not here attempt any general characterization of the difference between philosophical and other forms of investigation.

The second point to note is that what kinds of success in cognition are relevant to epistemology is somewhat controversial. Beliefs that contribute to the well-being of the organism are successful in some sense of that term, and yet some will hold that such success is not the kind of success within the purview of the discipline of epistemology. Notice that success of a practical sort will advert to the causal consequences of holding the beliefs in question, and a common view is that epistemology is more concerned with intrinsic features of cognition, the kind reflected in talk of inquiry for its own sake. When we engage in inquiry for its own sake, successful results will partake of a kind of success that is independent of any causal contribution to well-being or other practical concerns. When epistemologists reflect on the nature of successful contribution and the extent to which an organism achieves it, the predominant approach has been to reflect on a kind of success that abstracts from the consequences of cognition, whether those consequences are practical, moral, religious, political, or social. I am inclined here to make a terminological restriction that what I mean by the use of "epistemology" and related terms is just this study of success which abstracts from the consequences of cognition, but I do not wish to be understood to denigrate more pragmatic approaches that claim to be epistemological yet deny that there is any value in, or value in thinking about, inquiry for its own sake. In part, my decision to limit a discussion of epistemology to one involving reflection on success of this rather rarefied sort rests on an attempt to consider the best case available to those who defend a positive answer to the question of whether truth is the primary epistemic goal. On these pragmatic approaches, such a positive answer has nothing to recommend it, but I will ignore such approaches here.

In slogan form, my characterization of epistemology is that it is the study of purely theoretical cognitive success, where the notion of what is purely theoretical is understood as above in terms of abstraction from the causal consequences of the success in question. As already noted, I grant that this characterization has a mild stipulative dimension in that it refuses to count as epistemology certain types of pragmatic approaches to the study of successful cognition. Even with this mild stipulation, however, the point I am trying to bring across is the breadth of this

conception of epistemology, breadth with ramifications for the question before us. Once we notice this breadth, we will be struck by the strong reductionist flavor of the claim that truth is the primary epistemic goal. One's first inclination should be to maintain that each independent kind of cognitive success within the purview of epistemology identifies a cognitive goal in its own right. From this viewpoint, epistemic goals include knowledge, understanding, wisdom, rationality, justification, sense-making, and empirically adequate theories in addition to getting to the truth and avoiding error. Once we have seen the variety of cognitive successes, the proper answer would seem to be that the class of epistemic goods is manifold, as wide as the class of cognitive successes. To answer otherwise would seem to engage in Procrusteanism involved in mild humor that begins with "there are two kinds of people in the world," except that this particular truncation would begin by trimming the number from two to one. Given this initial variety, how could the reductionist viewpoint be sustained?

## Epistemic Values and Goals

One place to turn for a defense of the reductionist thesis is to the needs of epistemology itself. Perhaps in theorizing about cognitive success, there is a need for appeal to the goal of getting to the truth and avoiding error. I think something along these lines underlies the standard view that the epistemic goal is that of getting to the truth and avoiding error. The *Reader's Digest* condensed version of this approach goes like this: epistemology is the theory of knowledge, and knowledge includes, but is something more than, true belief. What more is required provides a connection between truth and belief. For example, it is well accepted that knowledge requires that the connection between belief and truth is *non-accidental*. Furthermore, for those who think that knowledge has a normative ingredient such as justification, rationality, responsibility, or the like, there will be a theoretical need to distinguish the sense of such terms from senses which have no place in an account of knowledge, such as practical senses. Perhaps that epistemic sense of the term should be characterized in terms of the goal, or good, of truth itself. If so, then perhaps the best way to think of the conditions needed for knowledge in addition to belief and truth are conditions best conceived in terms of some connection to truth.

Marian David provides an account of the matter along these lines. He says:

Knowledge, the epistemic concept par excellence, is usually defined in terms of belief, truth, and some other epistemic concept, say, justification: S knows p iff p is true, S belief p, and S is justified in believing p in a manner that meets a suitable anti-Gettier condition. Belief and truth, although fundamental to epistemology, are not themselves epistemic concepts. They are the nonepistemic ingredients in knowledge. This means that epistemology is not responsible for them; that is, as far as epistemology is concerned, belief and truth are *given* and can be invoked to account for epistemic concepts. The distinctly epistemic ingredient in knowledge is justification: the concept of S's *being justified* in believing p.... Epistemology is certainly responsible for this concept. Indeed, once an account of knowledge is at hand, the task of epistemology pretty much reduces to the task of giving a theory of justification. (David, 2001)

David here endorses several of the claims noted earlier. First, his conclusion that epistemology's task reduces to that of giving a theory of justification requires the assumption that epistemology is the theory of knowledge. So David endorses

- (1) Epistemology is the theory of knowledge.

Second, David endorses the strong claim that

- (2) The only epistemic concept in an account of knowledge is justification.

David argues for (2) by claiming that belief and truth are not epistemic concepts, though he does not explain why they are not. Perhaps he is thinking as follows. Truth is a *semantic* notion and belief a *psychological* notion. As such, they should be investigated by semanticists and psychologists, not epistemologists.

Even if we accept this explanation, David's claim is a bit strong since it ignores the condition needed to solve the Gettier problem. (2) denies that attempts to solve the Gettier problem fall within the domain of epistemology, and that claim is mistaken. This point is more important than it might seem initially to be, since David's approach to showing that truth is the primary epistemic goal requires focusing exclusively on the theory of justification. Once we alter (2) to accommodate this point about the Gettier problem, no claims about the theory of justification can establish on their own the claim that truth is the primary epistemic goal.

I want to stress that we should not overestimate this difficulty, however. One approach to the Gettier problem, the defeasibility approach, explains the condition needed to handle that problem as a function on truth and justification. The simplest example of this approach says that having knowledge requires, in addition to justified true belief, there being no true information which, if learned, would result in the person in question no longer being justified in believing the claim in question. Though complications arise for this overly simple approach,<sup>1</sup> the point to note here is that the central concepts in this approach – truth and justification – already appear in the account of knowledge. If David's point can be sustained that justification is the only epistemic concept among the concepts of justification, truth, and belief, then this approach to the Gettier problem will be quite useful in the attempt to argue that truth is the primary epistemic goal by focusing on its role in the theory of justification. So even though David's account cannot succeed without some discussion of the Gettier problem, there is some hope for the idea that the Gettier problem will not force an alteration of his view that providing a theory of justification is the central task of epistemology. Perhaps David finds the defeasibility approach to the Gettier problem attractive, and it is this attraction that leads him to bypass any discussion of the Gettier problem in his defense of the claim that truth is the primary epistemic goal.

The defense of (2) is still troublesome even if we allow this explanation of why the Gettier problem does not undermine it. One problem with this explanation is that it proves too much. Justification is a property not only of beliefs but also of other things (such as actions) not within the domain of epistemology. According to the above explanation, if a concept is suitable for investigation by disciplines other than epistemology (such as semantics or psychology), then that concept is not an epistemic concept. Insofar as justification is also a property of actions, it is suitable for

investigation by action theorists, a subclass among philosophers of mind. So it appears that the above defense of (2) implies that there are *no* epistemic concepts in the proper account of the nature of knowledge!

One might reply that there is a distinctive concept of justification that is at home only in epistemology, and that concepts of justification which apply to other things such as actions are simply different concepts. In this way, the concept of justification needed in an account of knowledge is at home only in epistemology and is thus the only epistemic concept in an account of knowledge.

Such a view faces serious difficulties. Consider a simple juridical example such as the O.J. Simpson case. Some who watched the case closely say that contrary to the actual finding of not guilty, the evidence justified both their position that Simpson was guilty (beyond a reasonable doubt) and a decision by the judge to send him to prison for life. This claim predicates justification of two things: the first is a position of the speaker, that is, a belief, and the second is an action by a legal authority. On the view proposed in support of (2) above, we would need to treat this juridical pronouncement as semantically awkward, in the way it is semantically awkward to say that both sides of rivers and certain financial institutions are banks.

Let us put these difficulties aside for the moment, however, to see how (2) is supposed to take us to the idea that truth is the primary epistemic goal. On this issue, David provides us with two quite different answers. His first answer appeals to the idea of supervenience:

However, it is usually held that, at some point, the theory of justification has to “break out of the circle” of epistemic concepts and provide a nonepistemic “anchor” for justification by connecting it in some significant manner with nonepistemic concepts. ... It is not hard to see how the truth-goal fits into this picture. It promises to provide a connection between the concept of justification and the concept of true belief, tying together the different ingredients of knowledge. (David, 2001)

This attempt to find a central place for the truth goal in epistemology is unsuccessful. The need to anchor epistemic concepts in the non-epistemic is more plausibly taken as a demand that evaluative concepts supervene on non-evaluative concepts, and there is no reason to identify truth with the supervenience base of evaluative concepts in epistemology, any more than there is reason to identify the concept of the good as the supervenience base of evaluative concepts in ethics. For example, one answer to the demand for a supervenience base is broadly empiricist: justification supervenes on experiential states and logical (or quasi-logical) relations between propositions. In this view, what *makes* a belief justified is that it stands in the right logical or quasi-logical relationship to another belief or to (the propositional content of) an experience. One may be tempted to maintain that the concept of truth will still enter into this picture when we try to characterize logical relationships, but that reply is a red herring. First, logical relationships can be characterized syntactically as well as semantically. Second, and more important, even if truth is an ingredient in our account of this supervenience base, it is not present in the manner of a *goal* or anything of the sort. To argue that because truth is needed to explain logic, it must be that a primary epistemic goal would also commit one to the view that because the concept of a sensation is necessary to explain experience, sensation itself must be a primary epistemic goal according to empiricism.

There is a particular kind of approach to the nature of justification that fits David's picture a bit better than the empiricist example. Alvin Goldman's reliabilism of 1979 begins with the same demand to break out of the circle of epistemic concepts in order to clarify the nature of justification, and the way in which Goldman prefers to break out of the circle is in terms of the concept of reliability. The central non-epistemic concept in his theory is that of the reliability of a process or method of belief formation, the percentage of times that process or method generates a true belief. Given that it is percentage of truth over error by which we favor some processes and methods over others, we might wish to characterize our theoretical preference for some processes or methods over others in terms of the goal of truth over error, thereby securing the idea that truth is the primary epistemic goal.

This way of defending the claim that truth is the primary epistemic goal thus requires the subconclusion that reliabilism is the correct approach to the nature of justification, a position that is far from trouble free.<sup>2</sup> But the particular problems one might cite for reliabilism are not the central difficulty. The deeper problem is that the question we are attempting to answer arises at a different point in epistemological inquiry than the question of the adequacy of some particular theory of justification. The question whether truth is the primary epistemic goal is a meta-epistemological question, whereas the question of the adequacy of reliabilism is not. When we ask about the primary epistemic goal, we are asking whether there is a way of defending the idea that truth is the primary epistemic goal that is neutral between competing epistemological theories, and given that interpretation of the question, the approach that relies on the adequacy of a reliabilist approach in epistemology must be judged to be unsuccessful.

David seems to recognize the meta-epistemological character of the issue, for he provides another and quite different answer to the question of the relationship between justification and truth. He says:

Why is truth typically cast as a goal ...? Alston provides the reason. It is generally agreed that being justified is an evaluative concept of some sort: To say that believing *p* is justified or unjustified is to evaluate belief *p*, in some sense, as a good thing or as a bad thing, as having some positive status or some negative status. The suggestion is that this type of evaluation, epistemic evaluation, is most naturally understood along broadly teleological lines, as evaluating beliefs relative to the standard, or goal, of believing truth and avoiding error. (David, 2001)

David here endorses Alston's claim that the evaluative character of justification should be understood teleologically, and this explanation of the relationship between truth and justification is a suitably meta-epistemological account in contrast to the previous account which seems to depend on the particular epistemological theory of reliabilism.

What we need to know, however, is how this point about the teleological character of justification helps to show that truth is the primary epistemic goal. David's argument proceeds by contrasting the idea that truth is the goal with the idea that knowledge is the goal:

Although knowledge is certainly no less desirable than true belief, the knowledge-goal is at a disadvantage here because it does not fit into this picture in any helpful manner.

Invoking the knowledge-goal would insert the concept of knowledge right into the specification of the goal, which would then no longer provide an independent anchor for understanding epistemic concepts. In particular, any attempt to understand justification relative to the knowledge-goal would invert the explanatory direction and would make the whole approach circular and entirely unilluminating. After all, knowledge was supposed to be explained in terms of justification and not the other way round. This does not mean that it is wrong in general to talk of knowledge as a goal, nor does it mean that epistemologists do not desire to have knowledge. However, it does mean that it is bad epistemology to invoke the knowledge-goal as part of the theory of knowledge because it is quite useless for theoretical purposes: The knowledge-goal has no theoretical role to play *within* the theory of knowledge. (David, 2001)

David's argument thus proceeds as follows. We have already seen his claims that (1) epistemology is the theory of knowledge and (2) justification is the only epistemic concept within the theory of knowledge. These points imply that the central epistemic concepts are knowledge and justification, so the primary epistemic goal will have to be one related to these concepts. David's teleological conception of justification makes truth a contender for being the primary epistemic goal, and if the only epistemic concepts are justification and knowledge, the central competitor of the idea that truth is the primary epistemic goal would be that knowledge itself is the goal. Once we see the implication from (1) and (2) to the point that the central task of epistemology is to construct a theory of justification, the knowledge goal cannot be the primary epistemic goal on pain of rendering our epistemology viciously circular. Only the truth goal can be central if our theory has any hope of adequacy, so truth must be the primary epistemic goal.

The form of argument here bears scrutiny since the particular argument in question relies on the false assumption that epistemology is the theory of knowledge. Given this assumption, the competitor to the idea that truth is the primary epistemic goal is the idea that knowledge is also an epistemic goal. Given the more general conception of epistemology outlined earlier and the variety of cognitive successes it implies, we will need a generalization of David's argument in order to conclude that truth is the primary epistemic goal. Such a generalization could begin by allowing that any purely theoretical cognitive success is of value and hence a suitable epistemic goal. But for each such goal, it has no theoretical goal to play within the project of theorizing about it, for such an account would make the theory "circular and entirely unilluminating." Moreover, it would have to be claimed, any goal that would render a theory circular and unilluminating in this way cannot be an epistemic goal, leaving truth as the only standing candidate for that role.

So, on this view, nothing of epistemic value can be an epistemic goal. Instead, only values that fall outside the domain of cognitive successes can legitimately be cited when explaining epistemic success. It also follows from this point that truth and belief cannot themselves be epistemically valuable, for otherwise getting to the truth cannot itself be an epistemic goal.

Contrast this result with an alternative picture. For simplicity, let us grant David's claims that epistemology is the theory of knowledge and that knowledge is to be understood in terms of justified true belief plus a Gettier condition. The alternative picture holds that each element in the theory of knowledge should isolate some epistemic good (something whose value contributes to and helps explain the value of knowledge). So, for example, truth itself must be an epistemic good, implying that it

has value exceeding that of mere empirical adequacy. In a similar vein, belief must be an epistemic good, implying that Pyrrhonian counsel to abandon beliefs in favor of acquiescing to the appearances cannot, in any literal sense, accord with the facts about what is important from a purely theoretical point of view. Similar points could be voiced about the other two conditions as well, and the accumulated result of these points is that if truth is an epistemic goal so are the other values just mentioned. As a result, truth is not the primary epistemic goal, but rather one among several epistemic values that have equal claim to being epistemic goals. This alternative picture strikes me as much more plausible than the idea that nothing that is epistemically valuable can itself be an epistemic goal or good.

The generalization of David's argument above has a further problem. This generalization requires that for any epistemic value *V* other than truth itself, the fundamental goal in terms of which a theory of *V* is constructed is truth itself. Once we put the view in these terms, I think it is fairly easy to see that it is false, though it may be more plausible in some parts of epistemology than in others. Perhaps, for example, it has some plausibility with respect to the value of belief itself. One might maintain that the value of belief is in some way dependent on the value of truth in that cognitive activity in general is prompted by the *desire* to get to the truth, and that belief is the result of such alethic motivation. Alternative alethic accounts are also available, ones which posit something sub-intentional as the driving force behind such activity, perhaps a need or interest or drive, or perhaps even an instinct, to get to the truth. One might even hold that it is the *function* of cognition to get to the truth and avoid error, whether the organism in question can be properly characterized by any of the intentional or sub-intentional descriptions just noted.

Such a view is not without difficulties, however, for there are alternatives to this truth-based account with some plausibility. To the extent that an organism has explicitly intentional motivation for cognitive activity, it may be more plausible to characterize those motivations in terms of a desire for knowledge or understanding rather than in terms of a desire to get to the truth, for we typically view mere true belief as something less than what we really want. Moreover, this point about fully intentional cognitive activity carries over to subintentional contexts as well. If the function of cognition is to be given a factive characterization – one which implies truth, as opposed to one which requires only adequate coping with an environment for purposes of survival and reproduction – the factives of knowledge and understanding have at least as much going for them as truth itself. That is, it may be more plausible to hold that the function of cognition is to produce knowledge or understanding of one's surroundings than it is to maintain that the function of cognition is to get to the truth about one's surroundings.

I do not wish to rest my negative answer to the question of whether truth is the primary epistemic goal on a rejection of truth-based accounts of cognition, however. My own view is that something close to a truth-based account is correct. Since my answer to the question before us does not depend on my view of this matter, I will present only the briefest account of it here. In my view, the fundamental function of cognition and the fundamental intentional attitudes involved in cognitive activity can be characterized without appeal to the concept of truth. The function in question is that of determining, for any claim *p*, whether *p*: for example, whether it is raining, or whether a particular object is a predator. Such a characterization appeals to no factive concepts stronger than truth, however. Moreover, even though the affective states

underlying cognitive activity often involve factives such as knowledge or understanding, such states are not universal – not everyone wants knowledge, for example, and not everyone is motivated by a concern for understanding. What is fundamental to explicitly intentional cognitive activity as such is a weaker state, for even small children and animals engage in fully intentional cognitive activity. We characterize curiosity as the desire to know, but small children, lacking the concept of knowledge, display curiosity nonetheless. The conclusion to draw from these considerations is that an adequate characterization of cognition as such will appeal to no motivational structure, either intentional or non-intentional, involving a factive element stronger than truth. To the extent, then, that we are persuaded that such a motivational structure must involve facticity of some sort, we should have no objection to characterizing cognition in terms of the goal of truth.<sup>3</sup>

Let us return to the main issue, arising from the fact that the argument we are considering requires that for any epistemic value *V* other than truth itself, the fundamental goal in terms of which a theory of *V* is constructed is truth itself.

Even if we grant that truth is in some way fundamental to our understanding of cognition itself, much more is needed to sustain this generalization. In particular, one would have to be able to argue that truth is fundamental to any of the kinds of purely theoretical cognitive successes noted earlier. Some such successes clearly involve the concept of truth, concepts such as knowledge, understanding, wisdom, likely hypotheses, and justified beliefs; and if these concepts involve the concept of truth, there is some hope that the goal of truth is somehow fundamental to an adequate account of these concepts. There are also cognitive successes that are not obviously truth related, such as the concepts of making sense of the course of experience and having found an empirically adequate theory. Both of these concepts can be explained without recourse to the goal of truth. An empirically adequate theory is one that will never be refuted by the course of experience, and one makes sense of the course of experience by developing a classification system for experiences together with a theory of explanation of how the various categories are explanatorily related.

Another example of a cognitive success that is not obviously truth related is that of responsible inquiry. Explaining why will take a bit of work, and we can begin by distinguishing deontic concepts from evaluative ones. Most ethical theories are teleological, explaining deontic notions such as obligation, permission, and forbiddenness in terms of evaluative notions such as the good. Not all do, however. Deontological theories refuse such definitions, explaining deontic concepts in other ways, the paradigm example being a Kantian theory that explains forbiddenness in terms of internal contradictions between universal maxims.

Epistemology has seen defenders of what is termed deontology, and here Chisholm is the paradigm example. Epistemic deontologists have not been deontologists in the sense defined in the previous paragraph, however, for they have defined epistemic duties in terms of epistemic values such as truth. Chisholm, for example, says our epistemic duties arise from the fundamental duty to (do one's best to) believe the truth and avoid error (Chisholm, 1977).

What would real deontology look like in epistemology? A good example of such can be developed along Bayesian lines. According to one kind of Bayesianism,<sup>4</sup> each of us has a complete theory of evidence in our heads, encoded in the form of conditional



probabilities. Such a theory can explain the forbiddenness of a (degree of) belief in terms of logical notions such as logical inconsistency, rather than in terms of value terms. A degree of belief is forbidden when, in the face of new experience, it becomes logically inconsistent with the relevant conditional probabilities.

Other examples are available as well. Though Richard Foley<sup>5</sup> wholeheartedly endorses the standard epistemic goal of getting to the truth (now) and avoiding error (now), his theory doesn't require this teleological aspect. On his theory, a belief is justified if and only if it conforms to one's deepest epistemic standards. These standards are epistemic principles one would endorse given as much time to reflect as is needed to reach a stable point of view.<sup>6</sup> Once we have such internal standards playing a theoretical role, real deontology can emerge, for it is in conforming to these standards that one's epistemic duties are satisfied.

Teleology in epistemology is needed only when the epistemic principles are licensed in some other, non-subjective way. If one's theory is truly subjective, then consistency with an internalized theory of evidence is all that is needed for defining key epistemic concepts such as justification. If, however, the theory of evidence is objective, then one needs some standard against which to assess its adequacy. It is here that epistemic teleology has its natural home.

Given this background, let us return again to the concept of responsible inquiry. The concept of epistemic responsibility can have strong subjective overtones, though it need not. A theory of epistemic responsibility may be modeled on the ethical notion of responsibility, which is to be contrasted with the notions of praiseworthiness and blameworthiness. In the moral sphere, one can be blameless in failing to live up to one's responsibility, and if we analogize to epistemic responsibility, we will be willing to say that one's inquiry might be irresponsible but blameless. If we refuse to distinguish epistemic responsibility from epistemic blamelessness, our theory of epistemic responsibility will be much more subjective, and the possibility of a non-teleological account of it will emerge.

Not much turns on this issue in the present context, however, for even in the face of a compelling argument that responsibility should be identified with blamelessness, there will still be an epistemic concept of value that can be understood in non-teleological terms – namely, intellectually blameless inquiry. So not only are there the notions of sense-making and empirical adequacy that do not presume the goal of truth, there are other, normative notions that can equally be clarified without appeal to the goal of truth.

## **Conclusion**

The slogan that truth is the epistemic goal has influenced much of epistemology in recent history, and we have seen that there is some truth in this idea. In particular, the idea of truth as a goal will play a significant and fundamental role in our understanding of cognition itself. The claim that truth is the primary epistemic goal, however, goes well beyond such a claim. It implies that the concept of truth plays a fundamental role in any adequate theory of any cognitive success of a purely theoretical sort. The view I have defended is that such a position has too narrow a view of the variety of epistemic values and goods, and that once we appreciate this variety and the broadened

conception of the domain of epistemological inquiry, we will go no further than to embrace the idea that truth is an important and central epistemic value and goal.

## Notes

- 1 For discussion of these complications, see Klein (1981).
- 2 To my mind, the most serious such problem is that reliabilism cannot explain adequately the concept of propositional justification, the kind of justification one might have for a proposition one does not believe or which one disbelieves (believes the opposite). I have argued this point in several places (e.g., Kvanvig, 1990, 1992, 1996, 2000, 2003).
- 3 For a more complete investigation of these issues, see the chapter on the value of truth in my 2003.
- 4 For an intuitive account of Bayesianism and its commitments, see van Fraassen (1990).
- 5 The view I discuss here is from Foley (1986).
- 6 This “operational” definition of the concept of one’s deep standards leads to various objections to Foley’s view, of the sort that usually plague counterfactual accounts. For example, if one’s deepest standards are chary of reflection of any kind, what results from Foley-reflection will be self-stultifying. To avoid these problems, one could develop a Foley-like view where the counterfactuals in question were taken to be evidentially related to the question of what one’s deep standards are, but deny that they are related definitionally.

## References

- Chisholm, R. (1977) *Theory of Knowledge*, 2nd edn. Englewood Cliffs, NJ: Prentice Hall.
- David, M. (2001) Truth as the epistemic goal. In *Knowledge, Truth, and Duty. Essays on Epistemic Justification, Virtue, and Responsibility*, ed. M. Steup. Oxford: Oxford University Press.
- Foley, R. (1986) *The Theory of Epistemic Rationality*. Cambridge, MA: Harvard University Press.
- Goldman, A. (1979) What is justified belief? In *Justification and Knowledge*, ed. G. Pappas. Dordrecht: Reidel.
- Klein, P. (1981) *Certainty*. Minneapolis: University of Minnesota Press.
- Kvanvig, J. (1990) The basic notion of justification. Christopher Menzel, co-author. *Philosophical Studies* 59: 235–261.
- Kvanvig, J. (1992) *The Intellectual Virtues and the Life of the Mind: On the Place of the Virtues in Contemporary Epistemology*. Savage, MD: Rowman & Littlefield.
- Kvanvig, J. (1996) Plantinga’s proper function theory of warrant. In *Warrant and Contemporary Epistemology*, ed. J. Kvanvig. Savage, MD: Rowman & Littlefield.
- Kvanvig, J. (2000) Zagzebski on justification. *Philosophy and Phenomenological Research* 60: 191–196.
- Kvanvig, J. (2003) Propositionalism and the perspectival character of justification. *American Philosophical Quarterly* 40 (1): 3–18.
- Kvanvig, J. (2007) *The Value of Knowledge and the Pursuit of Understanding*. New York: Cambridge University Press.
- Pappas, G. (ed.) (1979) *Justification and Knowledge*. Dordrecht: Reidel.
- Steup, M. (ed.) (2001) *Knowledge, Truth, and Duty. Essays on Epistemic Justification, Virtue, and Responsibility*. Oxford: Oxford University Press.
- van Fraassen, B. (1990) *Laws and Symmetry*. Oxford: Oxford University Press.

# Truth as the Primary Epistemic Goal: A Working Hypothesis

Marian David

## I Possessing Truth

Let us first consider what it means to talk of truth as a goal, without worrying about its being the primary epistemic goal. “I want truth” is a bit like “I want fruit.” Like fruit, truth comes in pieces. We have separate words for the different sorts of pieces fruit comes in. The pieces truth comes in we can just call *truths*. Philosophers often call them true *propositions*. The term is convenient because it also fits the pieces falsehood comes in. Say you are looking at the monitor and assert that the flight is leaving. I am looking too and I also assert that the flight is leaving. We both assert the same thing, the same proposition: the proposition that the flight is leaving. The proposition is either true or false. If it is true, we are both right; if it is false, we are both wrong. So, “I want truth” says that I want true propositions. “I want *the* truth,” taken literally, says that I want exactly the truth, the whole truth and nothing but the truth; that is, taken literally, it says that I want all the true propositions and only the true propositions, no false ones. Of course, normally one would not take it literally; one would assume that what I really wanted is all and only the true propositions relevant to some salient subject matter not explicitly specified.

“I want truth” and “I want the truth” do not mention what I want with the truth. But it is clear that I want at least to have it, to possess it. How does one possess truth? We can possess truth by way of *believing* it, by believing true propositions. We can possess truth by way of *knowing* it, by knowing true propositions. “I want truth” can be understood either way. So the broad idea of truth being a goal comes in two versions, corresponding to the two ways of possessing truth. Let us call them the *true-belief goal* and the *knowledge goal* respectively. They are both truth goals.

On the believing-way of possessing truth, “I want truth” comes out as: “I want to believe true propositions.” The more colloquial alternative, “I want to have true beliefs,” is fine too as long as we remember that using the word “belief” for the things possessed can be awkward if believing is the mode of possessing them. We would not want to say: “I want to believe true beliefs.” The noun “belief” is, as one puts it, act/object ambiguous. It might be used to refer to the act, or attitude, of believing a proposition. It might also be used to refer to the object, or content, of the attitude, to the proposition believed. If you believe that the flight is leaving, and someone says that your belief is true, then the noun “belief” refers to the proposition you believe. If someone were to say that your belief was silly, the noun would refer to your attitude of believing the proposition. The desire for truth relevant to the true-belief goal is a desire directed more at the attitude than the propositions; it is a desire for believing what is true, rather than for the truth of what one believes.

Propositional knowledge entails truth. One cannot know a proposition unless it is true: you cannot know that the flight is leaving at eight, unless it is leaving at eight

(of course, you might *think* you know it is leaving at eight even though it is not). Since propositional knowledge entails truth, we can express the knowledge goal simply as “the goal of having knowledge,” without mentioning truth. It is a truth goal nevertheless: it is the goal of possessing truths by way of knowing them.

The true-belief goal and the knowledge goal are both truth goals. Are there really two distinct goals here? Propositional knowledge entails true belief: one cannot know a proposition unless one believes it. (Sure enough, we sometimes say things like: “I don’t believe it; I know it.” But this is short for: “I don’t *merely* believe it; I know it.” We wouldn’t say: “I know it, but I don’t believe it.”) Since propositional knowledge entails true belief, the goal of having knowledge and the goal of having true beliefs are not entirely distinct. They are at least “overlapping” goals in this sense: if you want to have knowledge, you want something you cannot have without having true beliefs.

Epistemologists like to emphasize that knowledge requires something in addition to true belief: justified or warranted or rational belief, or belief based on good reasons or on adequate evidence (and they usually add that more is required still). This seems right for many typical uses of the word “know.” There also seem to be uses of “know” that do not require more than true belief. Alvin Goldman (1999) calls this the *weak* sense of “knowledge”: when an unfaithful husband greets his lover with the words “She knows,” he is not worried about whether his wife is justified in believing that he is cheating on her; all he is worried about is that she believes it and that it is true. The goal of having knowledge in this weak sense coincides with the goal of having true beliefs. But let us reserve the term “knowledge” for the *strong* sense, the sense in which it requires more than true belief.

We can say, then, that the knowledge goal and the true-belief goal are different, albeit overlapping, truth goals.

Invocations of truth as a goal, as something desired or aimed at, show up fairly frequently, especially in religious, philosophical, and broadly scientific contexts: phrases like “the search for truth” and “the pursuit of truth” are almost commonplaces there. Such invocations usually leave open which of the two truth goals is intended. In general, any reference to truth as a goal might be taken to refer to the goal of believing truths or to the goal of knowing truths – our title is a case in point – and even when the word “knowledge” is used, it is often difficult to tell whether it is intended in the weak or the strong sense. Compare William James’s (1911) commandment, “We must know the truth; and we must avoid error”; which he rephrases almost immediately as “Believe truth! Shun error!”

## II Truth Goals

The goal of having true beliefs, the goal of believing true propositions, is an indefinite or indeterminate goal: it is vague about how much of the truth is being aimed at. The most ambitious determinate goal would be the one corresponding to “I want to believe *the* truth”; it would be the goal of believing all the true propositions and not believing any false ones. Using “ $\forall p$ ” to abbreviate the universal quantifier “for all propositions  $p$ ,” it might be represented like this:

- (1)  $G(\forall p)(Tp \rightarrow Bp \ \& \ Bp \rightarrow Tp),$

but we must keep in mind that the relevant desire is supposed to be directed at the attitude rather than the propositions. So the goal represented in (1) should be understood in terms of wanting to believe all and only true propositions, or wanting to be such that one believes all and only true propositions, rather than wanting all propositions to be such that they are true if and only if one believes them.

It is fairly clear that we don't actually have this goal. The first part, the part about believing all the truths, looks like an absurd thing to want – partly because it is too obvious that there must be way too many truths for it to be humanly possible to believe them all, partly because there are vast numbers of unimportant and boring propositions, which, it seems, we wouldn't particularly want to believe even if they were true (and the truths are so redundant: any conjunction of true propositions is itself a true proposition). The second part, though, believing only truths, not believing any falsehoods, does look like something we want – though we might say “I would like that,” rather than “I want that,” signaling that we don't think the chances of reaching this state are at all good. Still, because of its first part, (1) looks like an overly ambitious goal – verbal “commitment” to it (“Tell me the truth!”) would normally be neither taken nor meant literally. One would automatically assume that it was meant to be restricted to all the relevant truths about some salient subject matter not explicitly specified. (But we should also remember that theists typically attribute omniscience to God, which involves believing all and only the truths. This indicates that we do tend to regard this state as something valuable, as some sort of ideal, albeit a remote one.)

Explicitly restricting the first part of (1) to “important and interesting” propositions would yield a goal that might be ascribed to us with some plausibility.

The indefinite goal mentioned earlier, the goal of “having true beliefs,” can be regarded as a tacitly as well as indeterminately restricted version of (1). Indeed, this indefinite goal may well be the one most plausibly ascribed to us; it looks like a goal we actually have – maybe just because it is so vague. It is naturally understood with a suppressed rider about falsehood. Say like this: it is the goal of having true beliefs and few, or no, false ones (or maybe the goal of having a large stock of true beliefs containing few, or no, false ones; or maybe simply the goal of having true beliefs rather than false ones).

There might be some propositions that are so discomforting (painful, upsetting, disgusting) to believe that we wouldn't want to believe them even if they be true, and some so comforting to believe that we would want to believe them even if they be false. Would this show that we don't have the indefinite goal? No. It would not even show that we don't have unrestricted (1) as a goal. We have various goals. In some cases different goals come into conflict and one “loses out.” In some cases a goal can be “neutralized” by contrary emotions. This does not show that we don't have the beaten or neutralized goal; it merely shows that the goal is not an absolute one.

The goals considered so far are “collective” goals. They refer not to individual propositions but to whole collections of propositions: for example, unrestricted (1) refers to the collection containing all and only the true propositions; the indefinite goal refers to some “fuzzy” collection, or fuzzily refers to a collection. But talk about wanting to have true beliefs can also be intended “distributively.” Consider the generalization:

$$(2) \quad (\forall p) \ G \ (Tp \rightarrow Bp \ \& \ Bp \rightarrow Tp),$$

which says that, for every proposition  $p$ , one has the goal (wants) to believe  $p$  if and only if  $p$  is true. Whereas (1) represents a single goal referring to a collection of propositions, (2) itself does not actually purport to represent any goal at all; rather, it represents a (false) statement ascribing to us a huge number of individual relativized goals, one for each proposition  $p$ . It is clear that we can't have all these goals: there are way too many propositions.

But we can at least say this. For each proposition  $p$  such that we are seriously asking whether  $p$  is true, we will have the relativized true-belief goal with respect to  $p$ ; that is, for each such proposition  $p$ , we have the goal (want) to believe  $p$  if and only if  $p$  is true: the very fact that we are seriously asking (ourselves or others) whether  $p$  is true shows that this is one of our goals with respect to  $p$ , at least at the time at which we are asking. Actually, what we want with respect to any such  $p$  is more completely captured like this: to believe  $p$ , if and only if  $p$  is true, and to believe  $\neg p$ , if and only if  $\neg p$  is true, where " $\neg p$ " refers to the negation of  $p$ . (Unlike Williams (1978, p. 38) and Sosa (2003), I take the "only if" parts to be important. Assume that, having asked whether  $p$  is true, you are being told that  $p$  is true and also, maybe by someone else, that  $\neg p$  is true. Since one of them must be true, you are being told the truth with respect to  $p/\neg p$ . But this is not what you wanted. You wanted *only* the truth with respect to  $p/\neg p$ : you did not want to get both, the truth and the falsehood.)

Note that there is something funny about how collective true-belief goals relate to goals concerning individual propositions. Say you are looking through a list of bargain CDs, wanting to buy all but only the ones by Eminem. You have reached the next item; it is called *Slim Shady*. You will want to buy *Slim Shady* if and only if it is by Eminem. Now assume you come to believe that *Slim Shady* is by Eminem: then you will (probably) want to buy *Slim Shady* (unless you believe that getting it would interfere with a more important goal, like not spending more than a certain amount). Here the collective goal relates fairly straightforwardly to a non-conditional goal with respect to one particular thing via your belief that the thing belongs to the collection of things you want. Now assume you are going through a list of propositions, wanting to believe all and only the ones that are true. You have reached the next item: the proposition  $p$ . Like before, we can say that you will want to believe  $p$  if and only if it is true. But now assume you come to believe that  $p$  is true. Can we again go on and say, non-conditionally, that you will (probably) want to believe  $p$ ? This seems odd at best. Having come to believe that  $p$  is true, you already believe  $p$  (it is hard to see how you could believe that  $p$  is true without believing  $p$ ); and once you believe  $p$ , it is at best odd to say that you *want* to believe  $p$ . Sosa (2003) suggests this shows that collective true-belief goals can play no role in "guiding" us about what to believe – whereas the collective Eminem goal can play a role in guiding you about what record to buy.

Let us briefly consider the knowledge goal too. Begin with the ambitious goal of having all knowledge; that is, the goal of knowing everything. To know everything would be to know everything there is to know. What is there to know? Truths, that is, true propositions, but no falsehoods: a false proposition cannot possibly be known (although it can be known to be false, but that would be knowing a truth about a falsehood). So part of the ambitious goal can be represented as " $(\forall p)(Tp \rightarrow Kp)$ ," which brings out that the knowledge goal is directed at possessing truths, just like the true-belief goal. What about a second, negative, part for the knowledge goal, a part corresponding to the  $(Bp \rightarrow Tp)$ -part of (1)? I am not quite sure. There seem to be two options:

- (3.1)  $G (\forall p) (Tp \rightarrow Kp \ \& \ Bp \rightarrow Tp)$ ,  
 (3.2)  $G (\forall p) (Tp \rightarrow Kp \ \& \ Bp \rightarrow Kp)$ .

According to (3.1), the ambitious knowledge goal would be the goal of knowing all the truths and not believing any falsehoods (avoiding error). But one might hold that the second part of the goal should be stronger than the second part of (1); that it should be about not believing something unless one knows it. Accordingly, (3.2) describes the goal in terms of wanting to be such that one knows all the truths and believes only what one knows. Note that, just as the condition  $(Tp \rightarrow Kp)$  subsumes the condition  $(Tp \rightarrow Bp)$ , due to knowledge requiring true belief, so the condition  $(Bp \rightarrow Kp)$  subsumes the condition  $(Bp \rightarrow Tp)$ , and for the same reason.

Problems about the plausibility, or rather implausibility, of thinking that the (3)s describe goals we actually have arise here in just the same way as they did with respect to unrestricted (1). To get goals more plausibly ascribed to us, the first parts of the (3)s would have to be restricted to important and interesting propositions. The collective knowledge goal most plausibly ascribed to us may again be an indefinite one, say, the goal of having knowledge (and avoiding error, or not believing things unless one knows them); which can again be regarded as some sort of tacit fuzzy weakening of an ambitious goal. And again collective goals should be distinguished from relativized goals:

- (4.1)  $(\forall p) G (Tp \rightarrow Kp \ \& \ Bp \rightarrow Tp)$ ,  
 (4.2)  $(\forall p) G (Tp \rightarrow Kp \ \& \ Bp \rightarrow Kp)$ .

These statements ascribe goals that are relativized to individual propositions  $p$ . Of course, no one can have all these goals, one for each proposition  $p$ . But we can say that, for each proposition  $p$  such that we want to know whether  $p$  is true, we will have such a relativized goal with respect to  $p$  – again, the full goals we have with respect to such  $p$ 's should add clauses about wanting to know  $\neg p$ , if  $\neg p$  is true.

### III Epistemic Goals

Among the many goals we have, or might have, only some appear to be directly relevant to the subject matter of epistemology. Let us call them *epistemic* goals. The primary epistemic goal would then be the goal that is the most important or most fundamental of the goals directly relevant to the subject matter of epistemology. What are the candidates? Epistemology is the theory of knowledge. So the main candidates should be the goal of having knowledge, the goal of having true beliefs, and some goal relating to the other main requirement for knowledge, say *having justified beliefs* (or warranted or rational beliefs, or beliefs based on good reasons or on adequate evidence – for convenience, I will often use the concept of justified belief as representative of this group).

It is difficult to be precise about the goal of having justified beliefs – in part that is because there is little agreement about what justification actually amounts to. Maybe the following will suffice for the indefinite version of the goal: it is the goal of believing a proposition, if one has justification for believing it, and only if one has justification

for believing it (but we should not assume that one *has* a justification only if one has *given* one's justification). Alternatively, we might say that it is the goal of believing a proposition, if one has adequate evidence for it, and only if one has adequate evidence for it. Again we should distinguish between the single collective goal of having justified beliefs and the relativized goals we have, or might have, with respect to individual propositions *p*, to believe *p* if and only if we have justification for believing *p*.

The goal of having knowledge and the goal of having true beliefs are both truth goals. What about the goal of having justified beliefs? Is it also a truth goal? No: having justified belief is not a way of possessing truth. This is because, unlike true belief and knowledge, justification does not entail truth: a belief can be justified even though it is false. This does not mean that there aren't some forms of justification that do entail truth – justification by mathematical proof might be one. Such forms of justification are called “infallible” justification. But most forms of justification are fallible: they don't entail (guarantee) truth, e.g. justification by induction, or by the evidence of our senses. So, in general, having a justified belief does not entail that the justified belief is true.

We have identified three main epistemic goals: having justified beliefs, having true beliefs, and having knowledge. Which one, if any, can be regarded as the primary epistemic goal?

Many of our goals are connected. Often we want something because we want something else and because we realize that if we get the former we will (probably) get the latter, or because we realize that without getting the former we can't get the latter. Chains of goals we have because of other goals we have terminate in “basic” or “intrinsic goals,” things we want, or value, just for their own sake and not for the sake of anything else. So, considering our epistemic goals, we might ask whether our epistemic goals are connected in some such way. If there is a single one at the bottom of such a chain, then that would be the primary epistemic goal.

Note that to claim that something is the primary *epistemic* goal does not imply that it is a basic goal all things considered, that we want it only for its own sake, that it is an “absolutely” intrinsic goal. Take the idea that possessing truths (by believing them? by knowing them?) is the primary epistemic goal. We want to possess truths in part for their own sake (we tend to be simply curious about things) and in part, or maybe mostly, because we think that having truths will increase our chances for securing our other, non-epistemic goals (in particular, we want truths about the most effective means for securing our other goals). The idea that possessing truths is the primary epistemic goal does not commit itself on this issue beyond the claim that there are no other *epistemic* goals for the sake of which we want truths – never mind any non-epistemic goals for the sake of which we might want truths. It only commits itself to the claim that possessing truths is a “relatively” intrinsic goal, that it behaves like an intrinsic goal within the domain of epistemology.

Actually, it commits itself to even less, or rather, to something slightly different. The idea that something is the primary epistemic goal has more the force of an “ought” than the force of an “is.” It says that, to the extent that we do have epistemic goals, we *ought* to have them because we have the one singled out as primary. Obviously, it is rather risky to make claims about what we actually want and about why we want what we want, without asking or studying many of “us.” Even if all of us do have the epistemic goals discussed earlier (or at least closely related ones), and even if these goals



are in fact connected in the manner mentioned above, some of us might not realize that their epistemic goals are so connected. It would then be wrong to say that *we* have the other goals because we have the goal that is in fact the primary one; all that can be said is that we ought to. One can then go on and say that, to the extent that we realize this, we will tend to have the other epistemic goals because we have the primary one.

#### **IV The True-belief Goal and the Justified-belief Goal**

The goal of having justified beliefs is not a truth goal: having justified beliefs is not a way of possessing truths. Still, it is very plausible to think that it is a truth-oriented goal in the sense that it derives from, or depends on, a truth goal. The truth goal most often mentioned in this connection is the goal of having true beliefs. Consider the following claims:

- A1 If you want to have TBs you ought to have JB.
- A2 We want to have JB because we want to have TBs.
- B1 If you want to have JB you ought to have TBs.
- B2 We want to have TBs because we want to have JB.

Assume we want to have true beliefs and want to have justified beliefs. A1 would give us a reason for connecting these goals in the manner described in A2: if we believe what A1 says, we ought to and probably will connect these goals in the manner described in A2. If, on the other hand, we believed what B1 says, we ought to and probably would connect these goals in the manner described in B2.

Comparing these pairs, it is obvious that the A's are way more plausible than the B's. Indeed, initially one may even think that the B's have nothing going for them at all, that they are just false. That is not quite right, however. True beliefs are generally useful; hence, they can be instruments for acquiring new justified beliefs as well as for acquiring justifications for old beliefs, provided they are true beliefs of the right sort: namely, true beliefs about how evidence, old or newly acquired, bears on propositions we don't yet believe; true beliefs about how the evidence we already possess bears on our old beliefs; true beliefs about where to find new evidence for or against old beliefs; and so on. (If you come to believe, correctly, that some evidence *e* you have just acquired adequately supports some proposition *p*, and you then come to believe *p* in part because of your true belief about how *e* bears on *p*, then this belief contributed causally to the fact that you now hold the justified belief *p* – this is not the typical case though: more typically, acquiring *e* will just make you believe *p*, without going through the higher-level belief about *e*'s bearing on *p*). True beliefs of this special sort – roughly, true beliefs about epistemic matters – have a certain measure of instrumental value relative to the JB goal. They have “epistemic utility” (Firth, 1998b) insofar as they tend to have causal consequences (namely the acquisition of justified beliefs) that contribute to the realization of one of our epistemic goals. So, the Bs are not simply wrong. Still, compared to the As, they look tenuous. They apply only to true beliefs of a special sort, whereas the As apply quite generally to justified beliefs of all sorts.

Moreover, the rather weak reason B1 provides for wanting to have true beliefs because we want to have justified beliefs seems itself to take us back to the desire for true beliefs: to some extent we want true beliefs (namely true beliefs about epistemic matters) because we want justified beliefs (about many things); and we want justified beliefs (about many things) because we want to have true beliefs (about many things). In short: we want true beliefs about epistemic matters because we want true beliefs about many things. The As have much more going for them than the Bs.

The claims above are put in terms of collective epistemic goals. The B's are not entirely wrong because, collectively speaking, having true beliefs tends to have some epistemic utility relative to the JB goal: to the extent that a collection of true beliefs contains special true beliefs about epistemic matters, the collection will tend to promote the acquisition of justified beliefs. The As, on the other hand, are much more plausible than the Bs, because they stay plausible even when taken as claims about particular goals relativized to individual propositions  $p$  – read “iff” as “if and only if”:

- A3 For any  $p$ : if you want to <believe  $p$  iff  $p$  is true>, then you ought to <believe  $p$  iff you have justification for believing  $p$ >.

The corresponding claim for B1, the claim that results from exchanging the parentheses, has no plausibility at all. Note that A3 does not talk about the instrumental value of having justified belief in  $p$  relative to the goal of having true beliefs. It does not talk about the causal consequences justified belief in  $p$  has, or tends to have, for acquiring other beliefs that are true: it does not talk about any causal consequences of having justified belief in  $p$  at all. Where, then, does A3 “come from,” as it were, if not from the epistemic utility of having justified beliefs vis-à-vis the goal of getting true beliefs? The intuitively most plausible answer to this question is surely something like:

- A4 You ought to have justified beliefs rather than unjustified beliefs, given that you want to have true beliefs rather than false ones, because justified beliefs are likely to be true, at least considerably more likely to be true than unjustified ones.

The intuitive plausibility of A1 is grounded largely in A3, which seems to rest on A4. Note again that A4 does not talk about the causal consequences of having justified beliefs. It says that justified beliefs *themselves* are likely to be true. This claim is much stronger than the claim that, collectively speaking, having justified beliefs tends to make it likely that one has true beliefs. The latter would hold even if having justified beliefs merely tended to promote the subsequent acquisition of true beliefs. Although, somewhat confusingly, having justified beliefs does tend to do that too. In addition to A4 (justified beliefs being themselves likely to be true), there is also an “epistemic utility”-aspect to having justified beliefs, but this aspect depends on A4. Justification is transferable: to the extent that you acquire new beliefs properly based on justified beliefs, your new beliefs will themselves be justified; hence, because of A4, they will themselves be likely to be true. So, because of A4, having justified beliefs tends to have a measure of general epistemic utility relative to the TB goal: it tends to promote the growth of true beliefs. This epistemic utility of having justified beliefs, which itself depends in part on A4, adds to the plausibility of A1.

It is tempting to talk of justification as if it were essentially a *means* to the truth (e.g., BonJour, 1985, pp. 7–8). But note that such talk can be seriously misleading. It connects with the *less* important aspect, the “epistemic utility”-aspect, underlying A1. Though it is correct to say that having justified beliefs, collectively speaking, tends to lead to true beliefs (via the transfer of justification), this instrumental value that accrues to justified beliefs in virtue of their tendency to have desirable consequences itself depends on the point that (A4) justified beliefs themselves are likely to be true – and this latter relation between justified beliefs and true beliefs is not like the relation between means and the (desired) consequences they tend to produce; it is not a causal “productive” relation at all.

If one really wanted to say that the epistemic status of a belief as justified or unjustified derived from its power as a means to truth in the ordinary sense of “means,” that is, from the tendency of the belief to produce true beliefs, one would be in serious trouble. A belief might causally contribute to the acquisition of any number of false beliefs even though it is justified. Say you are about to watch a news-show because you believe it to be a good source of information: your belief may well be justified even though it will soon lead you to acquire lots of false beliefs through watching a show that is in fact a source of massive misinformation. Conversely, a belief might causally contribute to the acquisition of any number of true beliefs even though it is unjustified. Say you are about to watch a news-show because you believe it to be a good source of information: your belief may well be unjustified even if it will soon lead you to acquire lots of true beliefs through watching a show that is in fact a good source of information.

I should mention a problem that arises with respect to A4. On the one hand, A4 is highly plausible; on the other hand, it is unclear whether it can remain plausible while keeping the promise to connect justification to truth. The problem is this. What does it mean to say that a justified belief is “likely” to be true? Justified beliefs are based on grounds, typically referred to as evidence: mainly other beliefs, memories, perceptual experiences, and introspective experiences. To say that a justified belief is likely to be true can mean that its truth is likely *relative* to its grounds. This seems to fit with our concept of justification. Unfortunately, a belief that is likely relative to some grounds may be unlikely relative to other grounds, and may not be likely at all in any absolute sense. So, on this construal, it is not quite clear whether A4 really manages to connect the JB goal to the goal of having true beliefs. Alternatively, one might propose that justified beliefs are likely in some *absolute* sense, which would seem to secure the connection to the TB goal. Unfortunately, this may not fit well with our ordinary concept of justification. Suppose, as Descartes did, that a powerful evil demon leaves our perceptual experiences and memories just as they are but makes all the perceptual beliefs we base on them false. In this scenario, our perceptual beliefs would not be likely to be true in the absolute sense. So they are all counted as unjustified, even though they would still be based on the same grounds as our beliefs are actually based. This goes against widely shared, though not universal, intuitions about the concept of justification.

The problem is a tangled one, especially since it is not even clear whether it really is a problem. How much weight should really be given to intuitions about far-fetched scenarios like demon worlds? Why isn’t relative likelihood of justified beliefs enough to secure the connection between the JB goal and the TB goal? It turns out that these

issues are rather complex, and I will have to set them aside for now. It should be remembered, in any case, that they pertain to A4 rather than A1 and A3. To acknowledge that the plausibility of A1 rests largely on A3 is one thing. To uncover difficulties with the additional idea that the plausibility of A3 rests on A4 is another thing.

## V The True-belief Goal and the Knowledge Goal

Given the considerations of the previous section, we can say that we (ought to) want to have justified beliefs because we want to have true beliefs, rather than the other way round. Let us conclude from this, at least as a working hypothesis, that the goal of having true beliefs is more basic than the goal of having justified beliefs; that the latter can be derived from the former. Even if this is correct, it does not follow that the goal of having true beliefs is the most basic epistemic goal. We still have to ask how it relates to the goal of having knowledge. Consider, then, the following claims:

- C1 If you want to have TBs you ought to have K.
- C2 We want to have K because we want to have TBs.
- D1 If you want to have K you ought to have TBs.
- D2 We want to have TBs because we want to have K.

If the C's were clearly more plausible than the D's, we could say that we (ought to) want to have justified beliefs *and* (ought to) want to have knowledge because we want to have true beliefs, but not the other way round. This would give us a reason for thinking that the TB goal is the basic epistemic goal.

However, there is something odd about the Cs. Consider C1. It is quite OK when taken collectively, as a claim about the general epistemic utility of having knowledge vis-à-vis the goal of having true beliefs. For knowledge itself constitutes evidence (Williamson, 2000): to the extent that you acquire beliefs properly based on knowledge you already possess, your new beliefs will tend to be true. But taken as a general claim about arbitrary individual propositions *p*, C1 seems deviant. The reason is not hard to find. Since knowledge requires true belief, part of what C1 tells you is that, if you want to believe *p* iff it is true, then you ought to believe *p* iff it is true.

It looks like the Ds are much more plausible. Consider D1. Quite independent of any instrumental value true beliefs might tend to have for acquiring knowledge (e.g., special true beliefs about how to acquire knowledge or evidence), D1 is plausible because true belief is a necessary condition for knowledge. It derives from a general claim about relativized goals with respect to individual propositions *p*: for any *p*, if you want to <know *p* if it is true, and believe *p* only if you know it> then you ought to <believe *p* if it is true, and believe *p* only if it is true>. On the face of it, it looks like we should say that we (ought to) want to have true beliefs because we want knowledge, rather than the other way round. This would give us a reason for thinking that the K goal is a more basic goal than the TB goal. We should, then, consider the competing hypothesis that the K goal is the basic epistemic goal: we (ought to) want justified beliefs because we want true beliefs, and (ought to) want true beliefs because we want knowledge.

The knowledge goal and the true-belief goal are both truth goals. So our two hypotheses are but two versions of the idea that truth is the primary epistemic goal.

Nevertheless, they would seem to be competing hypotheses. Most philosophers who have advanced or considered the idea that truth is the primary epistemic goal have taken this to mean that the true-belief goal is the primary one. Only a few – Williamson (2000) being the most recent – have explicitly advanced the view that it is rather the knowledge goal that is the primary one. Which view is the more plausible one?

Consider this thought: having true belief is a simpler, and in this sense, more basic state than having knowledge; having true belief is also a goal; hence, having true belief is a more basic goal than having knowledge. If “X is a more basic goal than Y” just means that X is somehow a more basic thing and is also a goal, then the true-belief goal may indeed be “more basic” than the knowledge goal. But this is not what one would normally mean when saying that X is a more basic goal than Y. Rather, what one would normally mean is that X is more basic *as a goal*; that we want, or ought to want, Y for the sake of X, rather than the other way round; or that the goal of having Y is in some other way derivable from the goal of having X. So, even if the state of having true belief is somehow more basic than the state of having knowledge, this does not show that it is more basic as a goal.

Let us try a different line of thought. The concept of knowledge is but the concept of justified true belief. So to want knowledge is not to want anything more than justified true beliefs. We have already seen that we want justified beliefs because we want true beliefs, rather than the other way round. Hence, we want justified true beliefs because we want true beliefs. The TB goal is the most basic goal after all. The impression that the K goal might be more basic than the TB goal is an illusion created by the complexity of the concept of knowledge.

This line of reasoning immediately faces the objection that, as Gettier (1963) shows, knowledge isn’t merely justified true belief. Assume you believe that one of your friends owns a Corvette because you have adequate but nevertheless misleading evidence for the belief that your friend Smith owns a Corvette. Your friend Smith does not own a Corvette; but your friend Jones does. Then your belief that one of your friends owns a Corvette is both justified and true, but it isn’t knowledge. Intuitively, your belief falls short of knowledge because it is only accidentally true.

Maybe we can make constructive use of this objection by turning it into a suggestion for how best to conceive of the goal of having true beliefs. Consider an initial diagnosis of Gettier cases: they show that knowledge requires *non-accidentally* true belief. Maybe this indicates that the true-belief goal is best conceived as the goal of having non-accidentally true belief. The goal might then be characterized by using subjunctives: we want to be such that we would believe p, if p were true, and would not believe p, if p were not true. Or, as Sosa (2003) might rather put it: it is the goal of being such that one would believe p if and only if p were true. (As a proposal for an analysis of the nature of knowledge that is responsive to Gettier cases the non-accidentality condition seems little more than a first step. However, it may well be good enough for paraphrasing our ordinary concept of knowledge and – more pertinent to the concerns of this paper – for describing the content of the true-belief goal). With the goal reconceived in this manner, the line of thought given above can be revived – now starting from the claim that to want knowledge is to want nothing more than justified non-accidentally true belief. As a result we get a defense of the idea that the true-belief goal is the primary epistemic goal in the slightly revised (improved?) version that the goal of having non-accidentally true beliefs is the primary epistemic goal.

## VI Problems

The view that the goal of having true beliefs is the primary epistemic goal appears to be committed to the claim that we *ought* to want knowledge and justification *only* because, and insofar as, we want true beliefs. In other words, the view seems committed to a version of epistemic “value monism” (DePaul, 2000): the sole basic epistemic value, or good, is true belief; and the sole basic epistemic disvalue, or bad, is false belief. The other epistemic goods, knowledge and justified belief, are derived goods; they are valuable only because true belief is valuable; they are extrinsic epistemic goods, whereas true belief is the sole intrinsic good, epistemically speaking.

This view has to face Meno’s question: “It makes me wonder, Socrates, this being the case, why knowledge is prized far more highly than right opinion” (Plato, *Meno*, 97d). Meno has observed that intuition speaks in favor of value judgment (a) (below). One can make additional observations concerning intuitive judgments about epistemic value that will give rise to similar questions:

- (a) Knowledge seems better than mere true belief.
- (b) Justified true belief seems better than unjustified true belief.
- (c) Unjustified false belief seems worse than justified false belief.
- (d) Unjustified true belief versus justified false belief? Intuition hesitates.

If knowledge and justification derive their whole value from the epistemically basic value of true belief, then their presence shouldn’t add (and their absence shouldn’t subtract) any value from states already containing true belief or false belief. But then, why (a)? And why doesn’t intuition tell us that the states mentioned in (b) and (c) are equally good and bad respectively? And why does intuition waver about (d)? Why doesn’t it speak firmly in favor of true belief, however unjustified?

The value monist might ward off Meno’s original question, the one concerning (a), by citing the revised true-belief goal, the goal of having non-accidentally true belief: knowledge is better than mere true belief, because knowledge contains non-accidentally true belief. Of course, this will lead to another difficult question: Why is non-accidentally true belief better than mere true belief? (Timothy Williamson, who is not a true-belief monist, holds a view reminiscent of Plato’s own view: knowledge is better than mere true belief because it is more stable, less vulnerable to rational defeat by future evidence; cf. Plato’s *Meno*, 97–8, and Williamson, 2000, pp. 62, 78, 86). In any case, the revised (improved?) true-belief goal doesn’t seem to be of much help to the true-belief monist when it comes to the questions concerning (b), (c), and (d): these questions remain, even if the true-belief goal is reconceived.

Consider the other version of epistemic value monism, the one according to which knowledge is the sole basic epistemic good with true belief and justified belief as derived goods, valuable only because they are required for having knowledge. In a way, this view seems to fit with (a) to (d). Knowledge is better than mere true belief, because the whole value of the latter derives from the value of knowledge. Items (b) to (d) would be accounted for in terms of how many goods required for knowledge the states mentioned there contain or lack. But in another way, this version of monism seems deeply strange. How does knowledge get its basic value if true belief and justified belief don’t have any

basic value themselves? This seems like adding 0 to 0 to get 1. Does it get its basic value from the non-accidentality of true belief? But then that wouldn't be value *monism*: non-accidentally true belief would have to be a second basic epistemic value in addition to knowledge (cf. DePaul, 2002, p. 181). Or is the view simply that knowledge boils down to nothing more than non-accidentally true belief?

Now consider a pluralistic view about basic epistemic value. There are two basic epistemic goods, together with their accompanying bads: true belief and justified belief; false belief and unjustified belief. Knowledge, on this view, is a derived good, valuable only because it contains the basic goods. Evidently, this picture makes for the most natural fit with (a) to (d). Knowledge and justified true belief are better than mere true belief because they contain all the goods. Unjustified false belief is worst because it lacks all the goods. Intuition wavers about unjustified true belief versus justified false belief, because each contains one of the two goods and one of the two bads, and we can't decide, at least not right away, whether one of the two should weigh more heavily than the other.

Of course, on this pluralistic picture truth is not the primary epistemic good or goal after all. Instead, having justified beliefs is an additional epistemic value independent from true belief; it is at least as basic an epistemic good as having true beliefs. What, then, about the earlier considerations from section IV, supporting the view that the goal of having true beliefs is more basic than the goal of having justified beliefs? The pluralistic picture needs an alternative diagnosis of the plausibility of A1 – the claim that you ought to have justified beliefs if you want to have true beliefs. The pluralist, it seems, has to say this: A1 is indeed highly plausible; but not because of A3 and A4. Rather, A1 is plausible because it is trivial. You ought to have justified beliefs anyway, never mind whether you want to have true beliefs or want to have false beliefs or neither: you just ought to have justified beliefs, period. You ought to want justified beliefs for their own sake.

This alternative diagnosis of A1 is off the mark. It confuses epistemology with ethics (Firth, 1998a). If there is a proper place for absolute *oughts* – for *oughts* that have a hold on us no matter what – then their place must surely be in ethics. (Pointing out that the *ought* in question is the “epistemic” *ought* does not help; for this just leads to a reformulation of the very issue under discussion: What is the epistemic *ought*, if not the one that has a hold on us insofar as we want to possess truth?) Moreover, the position that justified belief is valuable for its own sake leads straight into a mystery: Why is it that, when we do attempt to justify our beliefs, we produce evidence we take to be indicative of their truth? Why do we regard something as evidence for a proposition, only if we take it to indicate that the proposition is true, or at least likely to be true?

So let us return once more to the idea that having true beliefs is the sole basic epistemic good and the primary epistemic goal. Bernard Williams (1978, pp. 37–46) attempts to give an account of how a “pure inquirer,” an inquirer who starts out with the basic desire for having true belief about some issue, would on reflection be led to want knowledge. Much adapted to our present framework, the account goes roughly like this. Assume you want true belief with respect to a certain question. Assume you are aware that (A1) if you want to have true beliefs, you ought to have justified beliefs. This gives you a reason for wanting justified belief on the issue in question. You now have two desires (“wants”): you want to have true belief on the issue, and you want to

have justified belief on the issue. If you are minimally reflective, you will notice this: having both your desires satisfied requires being in the state of having justified true belief on the issue in question. This gives you a derivative reason for having justified true belief on the issue, which gets us close to an explanation of why it is that you might want knowledge because you want true belief (C2). But wait, knowledge isn't merely justified true belief. Fair enough, so we should give a slightly revised account starting from the goal of having non-accidentally true belief.

The account offers the true-belief monist a fairly natural interpretation of C2: the claim that we want knowledge because we want true beliefs. If you want (non-accidentally) true belief and are aware of A1, then a desire for knowledge will tend to come along with the ride. The desire for knowledge turns out to be an offshoot of the desire for true belief.

Does the account also help the true-belief monist with the *Meno*-problems raised by the intuitions catalogued under (a) to (d)? Here is a proposal (see Sosa, 2002, for a different proposal). If you want to have true belief with respect to some question and are aware of A1, then you will have *two* desires: a desire for true belief on the issue and a desire for justified belief on the issue. Note that both are real desires. Even though one desire, the desire for justified belief, is directed at a derived good without basic epistemic value, it is nevertheless a real desire, just as real as the desire for true belief – desires for derived goods are no less real *as desires* than desires for basic goods. But we usually want our desires to be satisfied; and we prefer it if more of our desires are satisfied rather than fewer. This – I am inclined to propose – is the source of the intuitions catalogued under (a) to (d). The intuitions arise due to a confusion of sorts. They do not reflect any bonus of intrinsic value accruing to knowledge over and above (non-accidentally) true belief, nor do they reflect any intrinsic value accruing to justified belief that would be independent from the value of (non-accidentally) true belief; rather, they reflect our desire to have our desires satisfied.

## References

- BonJour L. (1985) *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.
- DePaul, M. (2001) Value monism in epistemology. M. Steup (ed.), *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, ed. M. Steup. Oxford: Oxford University Press.
- Firth, R. (1998a) Are epistemic concepts reducible to ethical concepts? In *In Defense of Radical Empiricism: Essays and Lectures by Roderick Firth*, ed. J. Troyer. Lanham: Rowman & Littlefield.
- Firth, R. (1998b) Epistemic utility. In *In Defense of Radical Empiricism: Essays and Lectures by Roderick Firth*, ed. J. Troyer. Lanham, MD: Rowman & Littlefield.
- Gettier, E. (1963) Is justified true belief knowledge? *Analysis* 23: 121–123.
- Goldman, A. (1999) *Knowledge in a Social World*. Oxford: Clarendon Press.
- James, W. (1911) The will to believe. In *The Will to Believe and Other Essays in Popular Philosophy*. New York: David McKay.
- Plato (1976) *Meno*. Indianapolis: Hackett.
- Sosa, E. (2001) For the love of truth. In *Virtue Epistemology: Essays on Epistemic Virtue and Responsibility*, ed. A. Fairweather and L. Zagzebski. Oxford: Oxford University Press.



- Sosa, E. (2003) The place of truth in epistemology. In *Intellectual Virtue: Perspectives from Ethics and Epistemology*, ed. M. DePaul and L. Zagzebski. Oxford: Clarendon Press.
- Williams, B. (1978) *Descartes: The Project of Pure Inquiry*. London: Penguin.
- Williamson, T. (2000) *Knowledge and Its Limits*. Oxford: Oxford University Press.

## Further reading

- Alston, W.P. (1989) Concepts of epistemic justification. In *Epistemic Justification: Essays in the Theory of Knowledge*. Ithaca, NY: Cornell University Press.
- Alston, W.P. (1993) Epistemic desiderata. *Philosophy and Phenomenological Research* 53: 527–551.
- Chisholm, R.M. (1957) *Perceiving: A Philosophical Study*. Ithaca, NY: Cornell University Press.
- David, M. (2001) Truth as the epistemic goal. In *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, ed. M. Steup. Oxford: Oxford University Press.
- Feldman, R. (1988) Epistemic obligations. *Philosophical Perspectives* 2: 235–256.
- Firth, R. (1998) Epistemic merit, intrinsic and instrumental. In *In Defense of Radical Empiricism: Essays and Lectures by Roderick Firth*, ed. J. Troyer. Lanham, MD: Rowman & Littlefield.
- Foley, R. (1993) *Working without a Net: A Study of Egocentric Epistemology*. Oxford: Oxford University Press.
- Fumerton, R. (2001) Epistemic justification and normativity. In *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, ed. M. Steup. Oxford: Oxford University Press.
- Goldman, A. (1986) *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.
- Goldman, A. (1992) What is justified belief? In *Liaison: Philosophy Meets the Cognitive and Social Sciences*. Cambridge, MA: MIT Press.
- Haack, S. (2001) “The ethics of belief” reconsidered. In *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, ed. M. Steup. Oxford: Oxford University Press.
- Kvanvig, J.L. (1998) Why should inquiring minds want to know? *Meno* problems and epistemological axiology. *Monist* 81: 426–451.
- Kelly, T. (2003) Epistemic rationality as instrumental rationality: a critique. *Philosophy and Phenomenological Research* 96: 612–640.
- Maitzen, S. (1995) Our errant epistemic aim. *Philosophy and Phenomenological Research* 55: 869–876.
- Nietzsche, F. (1989) *Beyond Good and Evil: Prelude to a Philosophy of the Future* (trans. W. Kaufmann). New York: Vintage Books (original work published 1886).
- Steup, M. (1996) *An Introduction to Contemporary Epistemology*. Upper Saddle River, NJ: Prentice Hall.
- Velleman, D. (2000) On the aim of belief. In *The Possibility of Practical Reason*. Oxford: Oxford University Press.
- Wedgwood, R. (2002) The aim of belief. *Philosophical Perspectives* 16: 267–297.
- Williams, B. (1973) Deciding to believe. In *Problems of the Self*. Cambridge: Cambridge University Press.
- Williams, B. (2002) *Truth and Truthfulness: An Essay in Genealogy*. Princeton, NJ: Princeton University Press.
- Zagzebski, L.T. (2003) Intellectual motivation and the good of truth. In *Intellectual Virtue: Perspectives from Ethics and Epistemology*, ed. M. DePaul and L. Zagzebski. Oxford: Clarendon Press.
- Zimmerman, M.J. (2002) Intrinsic vs extrinsic value. In *The Stanford Encyclopedia of Philosophy*, ed. E.N. Zalta. <http://plato.stanford.edu>.

# Index

- a posteriori, 284–287, 294
- a priori
  - deduction, 43
  - justification, 116, 228–230, 238, 287
  - knowledge, 54, 127, 135, 177–201, 227, 273, 284, 331
- abduction, 185, 188, 190, 193, 198
- abilities, 158, 160–162
- abominable conjunctions, 31–32, 38, 41
- action
  - excusing, 89, 96–97, 99–100
  - rational, 100–101, 104, *see also* practical rationality
- alethic motivation, 359
- analysis of knowledge, 1, 7, 11–12, 33–34, 48, 96, 138, 225, 338–340, 346–347, 351
- analytic–synthetic distinction, 6
- analyticity, 183, 191, 192, 254
- animal knowledge, 11, 138, 142
- anti-luck
  - intuition, 152, 158, 161–162, 172
  - condition on knowledge, 152, 154, 156
- anti-skepticism, 69–70, 110, 112
- appropriate belief, 184, 203, 205, 214–217, 252, 302–303
- Aquinas, T., 21n, 141–142
- arbitrary switching objection to permissivism, 305, 318–322
- assertion, 7, 27, 46, 62
- assertive states, 207, 209–210, 213, 215, 217, 219n, 223–224
- attributer relativity, 61–62
- auditory experience, 111
- autonomous warrant, 276–277, 292
- availability of reasons, 215, 288–289
- basis relative safety, 165
- Bayesian
  - confirmation theory, 55n
  - deontology, 360–361
  - epistemology, 5, 8–19, 14–15, 17, 23, 25, 54
  - objective, 25
  - subjective, 3, 298, 304–305
- belief
  - clusters, 253
  - forming process, 95, 127, 146–147, 151, 165, 168–169
  - revision, 188–189, 303
- Berkeley, G., 30, 130
- Boole, G., 259–260
- bootstrapping, 227
- brain in a vat, 2, 5, 30–33, 35, 37, 40, 46, 48–49, 54, 111, 121
- bridge principle for permissiveness, 306

Carnap, R., 23, 309  
 causal  
     history of belief, 293, 324, 330, 333  
     relation, 6, 112, 117n, 178, 192, 327, 333, 342  
     theory of knowledge, 339, 344  
 causation, 117n  
 Chisholm, R. 13–15, 19, 23, 129, 162, 213, 255, 266n, 273n, 338, 360  
 circularity, 2, 20, 184, 189–190, 195–198, 200, 256, 275–276, 278–279, 293, 332, 358  
 Clifford, W., 84–86, 91  
 Clifford's Link, 87–91, 93–94  
 cognitive  
     abilities *see* cognitive faculties  
     architecture, 17  
     faculties, 4, 16, 99, 137, 147, 151, 158, 212, 330  
     goals, 302–303, 307, 310n, 354, monitoring, 146–147  
     success, 158, 160, 162, 244, 353–354, 358, 360–361  
 coherence and truth, 141, 246, 248–250, 253, 259–261, 266n, 273  
 coherentism, 202, 206–207, 209, 212, 222–223, 235, 253, 256, 261–265, 267–270, 275–279, 293–294  
 collective experience, 19  
 common sense, 15, 22, 69–70, 77, 82, 122, 130, 326  
 comparatives, 67  
 competent deduction, 27, 43, 45, 54, 63, 225–226, 229–232  
 competent speakers, 64, 70–71, 73–75  
 competitor propositions, 108–112, 122–123, 130  
 concepts, 12, 15, 68, 191, 205, 207, 212, 218–219, 281, 287–288, 356–357, 360  
 conceptual  
     analysis, 15, 165, 191  
     competence, 191, 211  
     content, 207, 211–212, 219–220n  
 conclusive reasons, 34, 38–39, 46, 48–51, 55n, 57–58  
 condemnation of belief, 86–87  
 congruence, 258–259, 261  
 conjunction of truth and belief, 169–170  
 conjunctivity of adequate justification, 271  
 conspiracy theory, 236–237, 239–241  
 content requirement on justifiers, 208–210, 212, 220n  
 context of attribution, 70, 94  
 contextualism  
     about indexicals, 61  
     about knowledge, 52–53, 58, 60–79  
     about quantifiers, 62  
     about sentences, 61  
     about truth conditions, 61, 64, 66–67, 68, 70, 72–74  
 contingent general truths, 272  
 conversational context, 44, 62 104–105  
 correlation, 19, 124–125, 130, 193  
 counterexamples, 2, 27, 38, 84, 99, 101, 118, 147, 151, 159, 226, 274  
 counterfactuals, 34, 38, 48, 50–51, 55n, 57–58, 100–101, 105–106, 144, 151, 229  
 curiosity, 136–137, 360  
  
 Davidson, D., 208, 210, 235  
 deceiver argument, 108–111, 114, 119n  
 deductive  
     reasoning, 42–43, 46, 128, 278, 291  
     validity, 126, 291  
 defeasibility of justification, 205, 223, 227–228, 232, 291, 355  
 deflationary theory, 74  
 degrees  
     of belief, 2, 5, 145, 272, 300, 302–303, 361  
     of knowledge, 104–105, 145  
 deliberation, 102–103, 105, 318–320  
 deliverances, 200, 270, 272  
 delusions, 270  
 deontological justification, 343–344, 346, 360–361  
 Descartes, R., 128, 136, 268, 272, 294, 326–327, 329, 371  
 desire for truth, 133–134, 142–146, 150, 359, 363–364, 370, 375–376  
 desires as propositional attitudes, 208–209, 214  
 direct realism, 13, 19  
 direction of fit, 158  
 disjunction, 122  
 Distribution principle, 45–46  
 dogmatism, 227, 235, 275, 280, 301  
 domestic vs. exotic skepticism, 109, 114, 116, 120, 128–129  
 doxastic vs. propositional justification, 223, 228–232  
 Duhem-Quine thesis, 183, 185–186, 190

- easy knowledge, 144–147
- easy knowledge problem, 227–228
- empirical regularity, 113
- empiricism, 13, 182–184, 251, 356
- entailment principle, 226–227, 230–232, 235–243
- enthymematic reasoning, 125, 128
- epistemic
  - access, 4, 7–9
  - anxiety, 53–54
  - circularity, 141, 143,
  - conscientiousness, 141, 143–148
  - deontologists, 360–361
  - difference-making, 241
  - evaluation, 126, 325–327, 333–336, 344–348, 357
  - merit, 108–110, 122
  - normativity, 3, 325, 327, 333–334
  - rationality, 110, 283–284, 292, 312
  - risk, 51, 102–103, 157
  - utility, 369–372
  - value, 24, 143, 351–354, 358–362, 374–376
  - virtues *see* virtue epistemology
- epistemology of peer disagreement, 312–313
- equivalence principle, 45
- error theory, 73
- ethics of belief, 87
- everyday contexts, 64, 67, 70–71, 73, 75, 82–83
- evidential
  - probability, 5, 8–9
  - system, 187–190
  - transmission *see* transmissibility of reasons
  - warrant, 28
- evidentialism, 335
- evil demon, 111, 371
- explanatory success, 78, 110–114
- externalism, 182, 261–262, 325, 327, 329–330, 333, 336, 338–341, 343, 345
- fake barn cases, 38, 105
- fallibilism, 102, 104, 170–171
  - see also* infallibilism
- fallibilism and safety, 170–171
- finite mind objection to infinitism, 275, 281
- first-person perspective, 125, 175n
- first-person pronoun, 61–62
- forgotten evidence, 317, 328, 348–349
- formal epistemology, 25
- formal models and methods, 7–10, 14–15, 23–26
- foundationalism, 123, 128, 138, 207, 212, 244, 254–258, 261–263, 265, 267–271, 275–279, 284, 287
- foundherentism, 257, 265
- Frege, G., 183
- Gettier
  - cases, 89, 153, 155–156, 162, 174, 342–343, 349, 373
  - problem, 10, 15, 66, 72, 78, 105–106, 143, 342–343, 354–355, 358
- Given Theory, the, 206–207, 219n
- global skepticism, 120–121
- Grice, P., 24, 44
- hallucination, 50, 121, 326
- heavyweight propositions, 30–31, 33–37, 40–41, 46–51, 53, 58–59
- high stakes, 87, 89, 92–94
- highest common factor for safety, 170
- holism, 119n, 177, 183–184, 246, 250, 252, 254
- holistic evidence, 177, 186, 190, 193, 195–198
- holistic naturalism, 177, 183–184
- honesty, 134, 136, 148
- Hume, D., 112, 130, 182, 332–333, 337n
- hypothetical scenarios, 87–89
- idealism, 30, 130
- illusion, 180, 285, 327
- indexical expressions, 61, 67, 74, 281, 322
- indicators, 19, 22, 36
- indiscriminability, 2–3, 8–9, 13, 24
- induction, 116–117, 195–196, 198, 210, 223, 261, 309, 330–332, 368
  - by enumeration, 114–115, 128–130, 188
- inductive
  - confirmation, 114–115
  - generalization, 64
- inductivism, 225–226, 230, 232–233, 235–239
- infallibilism, 170–171, 204
- inference to the best explanation, 107, 110, 114–115, 119n
- inferential
  - internalism, 107, 123–129
  - justification, 115–116, 123–124, 131, 204, 210–214, 223, 284–291, 294, 296–297
  - semantics, 6

- infinite regress, 107, 256, 265
- infinetism precluding assent, 280
- informal epistemology, 23, 25–26
- information, 7–8, 34, 36–39, 72, 89, 137, 193, 241–242, 246–248, 258, 299–300, 334–336, 371
- inquiry, 1, 14, 22–23, 316, 318, 353, 357, 360–362
- instrumental value, 290, 294, 369–372
- intellectual character, 133–151, 326, 333
- intellectualism, 95
- intentional cognitive activity, 359–360
- intentional state, 5–6, 359
- intentions, 10–11, 104
- interpersonal vs. intrapersonal slack, 298, 303–306, 308
- internal to subject's perspective, 325–329, 335–336
- internalism, 143, 215, 265
  - about justification, 324–349
- interpretation, 12, 17, 63–64, 126–127, 169
- intersubjectivity, 18–19, 22
- introspection, 8–9, 20, 254, 267
- introspective
  - experience, 1, 3, 178–179, 219n, 371
  - knowledge, 135
- intuitions, 17, 32, 43, 51, 54, 57, 69–70, 72, 76, 80–82, 152, 157–162, 172, 179, 185–186, 299, 329, 371, 374–376
- invariantism, 60, 62, 64, 70, 77–82
- irrational behavior, 102–103, 189, 318–319
- isomorphic skeptical hypothesis, 107, 111–114, 118
- James, W., 301–304, 310n, 364
- Jeffrey conditionalization, 9, 14, 23
- justification and epistemic responsibility, 13, 328–330
- justification makers, 203, 206, 214–216
- Kant, I., 182, 360
- Keynes, J., 127–128, 131
- knowledge
  - attributions or ascriptions, 32, 43, 52–53, 58, 65–68, 70–74, 77, 80, 83, 94, 96, 101, 104, 345
  - in children, 135–136, 151, 360
  - denials, 65, 67, 71, 96
  - of the future, 57, 121–122
  - and normativity, 2–3, 12, 74, 188, 325, 327, 333–334, 354
  - knowledge-reasons principle, 98
  - Kripke, S., 42
  - Kuhn, T., 186
  - law of non-contradiction, 264, 266n, 269, 272
  - lemmas, 226–227, 230–231, 233, 235, 239, 243n
  - Lewis, C.I., 245, 257–262, 266n
  - Lewis, D., 50, 80
  - lightweight implications, 30, 33
  - linguistic data, 89
  - local skepticism, 120–121
  - locational properties, 112–113
  - logic, 19, 28, 31–32, 179, 183, 185–186, 189, 190–191, 195, 199, 264, 269, 272, 361
    - epistemic, 7–9, 14, 25
  - logical atomism, 12
  - logical consequence, 42–43
  - logical relations amongst propositions, 208–209, 213, 225, 233, 258, 265, 307, 311, 344, 356
  - lottery cases, 39n, 55n, 57, 159
  - lottery problem, 155–157, 163
  - low stakes, 87, 92, 98
  - lucky guess, 153–154, 338
  - material world, 28, 33, 35–37
  - mathematical knowledge, 185–187, 190, 193, 253
  - memory, 20, 22, 35, 39n, 67, 99, 101, 129–130, 257–262, 264, 272, 286, 339, 348–349
  - mentally alike individuals, 340, 344
  - meta-epistemology, 123, 357
  - meta-judgment, 71, 73, 82
  - metaphysical essentialism, 167
  - metaphysics, 113, 136, 182, 186–187, 269
    - of knowledge, 17
  - methodological conservatism, 110
  - mind-independent objects, 28, 30–31, 50
  - minimal skeptical hypothesis, 110–111, 113
  - misrepresentation, 36
  - modal configurations, 113, 118
  - modal predication fallacy, 168
  - modality, 113, 118, 154, 157, 160, 163, 165–168
  - moderate empiricism, 182–183
  - modus ponens*, 28, 46, 131, 179, 187–189
  - Moore, G.E., 78, 136, 227

- moral
  - analogy, 150, 334, 341
  - evaluation, 63–64, 334–335
  - justification, 351
  - relativism, 63–64
  - responsibility, 328
  - virtues, 134, 144
- mundane propositions, 109–110
- naturalism, 185–187, 190, 198
- necessary conditions, 47–49, 96, 122, 158, 208, 293, 344–347
- necessary truth, 42, 113, 127, 165, 179, 272
- necessity of origins principle, 42–43
- neo-rationalism, 227–230, 232–233, 238
- Neurath, O., 186, 189, 198
- no-self theories, 61–62
- non-arbitrary rejection of a proposition, 108, 122
- non-basic knowledge, 331
- non-deductive reasoning, 126, 128
- non-inferential justification, 116, 123–124, 284–290, 294, 296
- non-propositional insight, 179, 199
- normative status of belief, 85
- norms of reasoning, 141, 143–145, 150
- Nozick, R., 34, 38, 47–49, 225, 230
- objectivity, 14, 23, 326–327, 333–336
- open-mindedness, 133, 143, 150
- ordinary conversation, 44, 96
- paraconsistent logic, 269, 272
- paradox, 69–76, 78–83, 116
- perception, 4, 19, 29–30, 34–36, 121, 129, 227, 235, 249, 250–252, 264, 330–332
- perceptual
  - experiences *see* sensory experience
  - faculties, 41, 47, 50, 142
  - justification, 212, 235, 238
  - knowledge, 135, 227
  - reasons, 30
- phenomenal conservatism, 20, 263–264
- Plato, 136, 142, 338, 374
- plausibility judgments, 68
- pluralism, 20, 352, 375
- positism, 256
- possible worlds, 49, 50, 314, 316, 154, 156–158, 160, 161, 165–168, 170–172, 180, 187
- practical
  - deliberation, 103, 319–320
  - interests, 84–106
  - rationality, 96–104
- pragmatic encroachment, 84, 88, 95–104
- praiseworthiness, 135–136, 328–329, 361
- pre-theoretical, 138, 148, 150, 299
- premise principle, 210–218, 220n, 223–224
- premise vs. rule circularity, 189, 196
- principle
  - of avoiding arbitrariness, 279, 286, 293
  - of avoiding circularity, 279, 293
  - of reflection, 317
- privileged access, 191, 326–327, 335–336
- pro tanto, 91
- probability, 5, 8, 9, 14, 23, 43, 67, 115, 124–129, 214, 231, 258–261, 267–268, 272, 305–310
- problem of misleading evidence, 115–116
- propositional content in justifiers, 207–210, 214
- protoknowledge, 29
- pseudo-locations, 107, 112–113
- pseudo-objects, 113
- purism, 95
- quantifiers, 62–63, 364
- Quine, W.V.O., 183–187, 190, 194n, 264–265, 269, 272
- radical empiricism, 183–184
- rational reflection, 47, 135
- rationalism, 192, 198
- rationality, 96–104, 211, 283–285, 292, 300–301, 309–310, 315–316
- real world hypothesis, 107, 111–114
- reality of the past, 28
- reason-knowledge principle, 96–97, 99–101, 103–105
- reasoning and the rational credibility of a proposition, 275–277, 297
- reasons
  - for action, 91–93, 206,
  - for belief, 29–31, 37, 48–50, 57, 91, 93–94, 129, 178, 214–216, 223, 288–289, 292, 296, 328–329, 333, 340–350, 364, 367
  - weighing, 92–93, 98, 102
- reflective equilibrium, 41, 254–255, 263–264
- reflective vs. unreflective knowledge, 11, 138, 142, 345
- regress, 107, 123–124, 128, 179, 199, 256, 265, 273, 331–332

- regress argument for immediate justification, 205–206, 209, 223
- regress problem, 274–295
- regret, 27–28, 93
- Reid, T., 20–21, 26
- relative frequency interpretation of probability, 126–127
- relativity, 61–62, 274–275
- relevant
  - alternatives, 33, 48, 58
  - evidence, 34, 100, 110, 240
  - knowledge, 24, 96
- reliabilism, 127, 227, 244, 249, 251, 261, 280, 333–335, 339, 357
- reliable informant, 94, 154, 247
- representative realism, 130
- risk of error, 102, 157
- rules of thumb, 81–82
  
- safe reasons, 98, 102–103
- safety condition on knowledge, 152, 156–175
- schema, 7–8, 41
- scientific
  - evidence, 19, 23, 26, 251
  - inquiry, 3
  - instruments, 251
  - knowledge, 66, 72
  - laws, 186, 195, 262
  - realism, 188
- self
  - deception, 270
  - monitoring *see* cognitive monitoring
- semantics, 6–7, 10, 24, 355–356
  - contextualism, 60–64, 67–68, 77, 80–81
  - models, 81
- sensitivity, 45, 154–157, 162, 173, 193
- sensory
  - deception, 108–110, 114
  - evidence, 30, 33, 331–332
  - experience, 3, 13, 20, 26, 30, 33, 36, 108–110, 130, 178, 208, 212–213, 214, 251, 327, 330–332, 356
- set inclusion, 10, 24
- Sextus Empiricus, 274, 280
- simplicity, 114, 118, 130, 151, 184, 188, 260
- simplification, 7, 14, 256, 261
- skeptical
  - hypotheses, 41, 46, 69, 107, 109–114, 237–238
  - paradox, 69–76, 78, 80–83
  - skepticism, 2–4, 6, 8, 30–34, 37–39, 44, 48, 57–58, 65–75, 78–83, 107–132, 232, 255–256, 268, 272, 324, 330–333, 346
  - skepticism about
    - external world, 34, 37, 66
    - induction, 114–116, 332–333
    - inference to the best explanation, 114–115
    - perception, 112, 332–333
  - skepticism objection to externalism about
    - justification, 330–333
  - social epistemology, 26, 334
  - solipsism, 35
  - sorites argument, 2, 13, 36
  - Sosa, E., 11, 138, 142, 165, 366, 373
  - speaker intentions, 64, 104
  - standards of judgment, 64, 76–79, 81–82
  - stereotype, 77, 81–82, 246
  - structural
    - equivalence, 79, 111–112
    - realism, 118
  - subject relativity, 61–62, 96
  - subjective
    - Bayesianism *see* Bayesian
    - duty, 343
    - experience, 19, 121, 287
    - vs. objective, 14, 18, 23, 326–327, 333, 336
    - probability, 5, 320
  - supervenience, 2–3, 6, 264, 280, 282n, 356
  
  - tacit
    - belief, 289–290, 294
    - theory, 191
  - teleological, 10, 17, 357–361
  - testimony, 26, 35–36, 52, 57, 246, 251–252, 259, 268, 272, 326, 328, 348
  - theory construction, 111, 151
  - three-place vs. two-place relation of evidential support, 308–309
  - threshold
    - conception of knowledge, 42, 104, 268, 271
    - for belief, 302–303
  - tolerance of epistemic risk, 157
  - total body of evidence, 116
  - transfer of epistemic warrant, 276–280, 292, 294, 370–371
  - transitivity, 2, 8–9, 15, 24, 44–45
  - transmissibility of reasons, 28–30, 230–231
  - true-belief monism, 376

- trust
  - in cognitive capacities, 140–141, 248, 250–252, 349
  - pre-reflective, 141
  - in self, 143–145, 147, 151
- truth
  - conditions, 6–7, 60–68, 70–81, 126, 281
  - directed epistemology, 96, 104–105, 134, 137, 141–151, 351–377
  - and fiction, 248
  - goals, 363–364, 368, 372
  - maker, 125, 131, 203
- under-determination principle, 108–110, 114, 119–120, 122
- understanding, 11–12, 17, 24–25, 141, 251, 292, 296, 351, 353–354, 359–361
- uniqueness principle, 298–305, 307–308, 310n, 312–313
- unsafe belief-forming processes, 166–170, 173–175
- vagueness, 13–14, 24, 76, 80, 145, 174, 300
- value monism, 374–375
- veridical perception, 12, 30, 121, 123, 141, 292
- vicious motivation, 144, 147–148, 151
- virtue epistemology, 133–151
- virtuous motivation, 133–151
- visual experience, 29–30, 34, 50, 111, 121, 207, 251, 269, 285–289, 292, 295
- warrant-emergent vs. warrant transfer
  - coherentism, 276–279, 282n
- web of belief, 183–184, 186–187, 195, 268–269, 272
- weighing statements, 92–93
- well-foundedness, 316, 318, 324, 341–343, 346–347, 350
- wisdom, 11, 17, 322, 351, 353–354, 360