OXFORD

Internalism and Externalism in Semantics and Epistemology

edited by

SANFORD C. GOLDBERG

INTERNALISM AND EXTERNALISM IN SEMANTICS AND EPISTEMOLOGY

This page intentionally left blank

Internalism and Externalism in Semantics and Epistemology

Edited by
SANFORD C. GOLDBERG

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

© the several contributors 2007

The moral rights of the authors have been asserted
Database right Oxford University Press (maker)

First published 2007

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,
without the prior permission in writing of Oxford University Press,
or as expressly permitted by law, or under terms agreed with the appropriate
reprographics rights organization. Enquiries concerning reproduction
outside the scope of the above should be sent to the Rights Department,
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover
and you must impose the same condition on any acquirer

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Data available

Typeset by Laserwords Private Limited, Chennai, India
Printed in Great Britain
on acid-free paper by
Biddles Ltd., King's Lynn, Norfolk

ISBN 978–0–19–927575–5

1 3 5 7 9 10 8 6 4 2

Acknowledgements

All of the papers in this volume are new, and were written expressly for the volume. Most contributors presented their papers at a conference on this subject held in April 2005 at the University of Kentucky (in Lexington). For underwriting the cost of that conference I would like to acknowledge the Major Research Award given to me by the University of Kentucky's College of Arts and Sciences. In addition, I would like to acknowledge the Lady Davis Fellowship, whose support while I was at the Philosophy Department at Hebrew University in Jerusalem, Israel (Fall 2005), enabled me to complete work on this volume. Finally, I would like to thank Peter Momtchiloff of Oxford University Press, for seeing this volume through to publication; and the rest of the helpful staff at OUP (including Jacqueline Baker and Eva Nyika).

I dedicate this book to the memory of my teacher Sidney Morgenbesser.

This page intentionally left blank

Contents

Introduction 1
Sanford C. Goldberg

1. Externalism in Mind and Epistemology 13
Jessica Brown
2. What and About What is Internalism? 35
Richard Fumerton
3. Externally Enhanced Internalism 51
Earl Conee
4. How to be a Neo-Moorean 68
Duncan Pritchard
5. Some Ins and Outs of Transglobal Reliabilism 100
David Henderson and Terry Horgan
6. Entitlement, Opacity, and Connection 131
Brad Majors and Sarah Sawyer
7. Content Externalism, Entitlement, and Reasons 160
Anthony Brueckner
8. What's Wrong with McKinsey-Style Reasoning? 177
James Pryor
9. A Priority and Externalism 201
John Hawthorne
10. The Inference that Leaves Something to Chance 219
David Sosa
11. Semantic Externalism and Epistemic Illusions 235
Sanford C. Goldberg
12. Psychological Externalism and the Role of Belief in the Analysis of Knowledge 253
Joseph Owens

Further Reading 273
Index 295

This page intentionally left blank

Introduction

Sanford C. Goldberg

1. IDEOLOGICAL INTRODUCTION

Almost forty years have elapsed since the publication of the papers and books initiating the internalism/externalism debate in epistemology,[1] and it has been a bit more than thirty years since the publication of the papers initiating the internalism/externalism debate in the philosophy of mind and language.[2] Until very recently little attention has been paid to whether there are any interesting connections between these two debates.[3] (Until recently, what discussion there was on the topic of connections between the two I/E debates focused primarily on semantic externalism's implications for self-knowledge and skepticism.)[4] Perhaps this is because there are no interesting connections; or perhaps it is just a historical accident, owed perhaps to the fact that in these days of specialization few philosophers have such disjunctive interests. Whatever the explanation, the present volume aims to address this lack. If successful, it will contribute to the very recent trend among those who seek to determine what connections exist between these two I/E debates.

While we cannot tell in advance what the fruits of such examinations will be, there are reasons to think that they are worth pursuing.

A first, introductory point that can be made in this regard is this. Both debates have attached a great deal of significance (within their respective domains) to the distinction between 'internal' and 'external' matters, and it would be interesting to determine whether this reflects anything more than coincidence. To be sure, the distinction itself is drawn differently in the two domains. In the philosophy

[1] These include Goldman 1967 and Armstrong 1973.

[2] See Kripke 1972; Putnam 1976; Burge 1979.

[3] For just a few of the recent discussions, see e.g. McDowell 1994; Sturgeon 1997; Williamson 1998; Brewer 2002; Chase 2001; Brueckner 2002; Burge 2003; Vahid 2003; Brown 2004; Peacocke 2004a and 2004b; Pritchard and Kallestrup 2004; Schantz 2004; Goldberg 2007a.

[4] See for example the various contributions to Ludlow and Martin 1997; Wright, Smith, and MacDonald 1998; Nucettelli 2004; Schantz 2004.

of mind and language, the internal/external divide is standardly taken to be the divide between states and properties that supervene on the intrinsic, non-relational states and properties of the individual subject, and those that do not so supervene. In epistemology, the internal/external divide is standardly taken to be the divide between states and processes that are available to searching reflection, and those that are not.[5] Importantly, however, both debates construe their respective internal/external divide as marking an important choice-point in theory-construction, regarding the supervenience base of the properties at the heart of the two respective domains. So, for example, the debate in the philosophy of mind and language concerns the supervenience base of properties regarding linguistic meaning, mental content, and the propositional attitudes. Hence the traditional formulation of psychological internalism is as the thesis that psychological properties supervene on the intrinsic, non-relational features of the physical body of the subject instantiating those properties; and the traditional formulation of psychological externalism, as the denial of this supervenience thesis. In epistemology, the debate (in the first instance) is over the nature of epistemic justification, although connections can be made with the properties of knowledge and rationality as well. Hence a traditional first-pass formulation of epistemic internalism is as the thesis that epistemic justification supervenes on what is reflectively accessible, whereas epistemic externalism would be the denial of this. (But as we will see, this is just a first pass.) An interesting question is thus suggested: what would we discover if we compared the motivation for distinguishing 'internal' from 'external' matters in these two distinct domains? Does the fact that such a distinction is made in both reflect some underlying commonality between central properties investigated in their respective domains?

Relatedly, and second, both debates concern the core (semantic or epistemic) relations between mind and world, and it would be interesting to determine whether one's choice as to how the semantic relations are to be conceived will affect one's choices as to how the epistemic relations are to be conceived (or vice versa). The debate in the philosophy of mind and language concerns the materials needed in order for the mind to *represent, refer to*, and *conceptualize* the world as it does—whether rightly or wrongly.[6] Given that the mind represents or conceptualizes the world as being a certain way, it is then natural to go on and ask whether its representations (conceptualizations) amount to *knowledge* or *justified belief* regarding the features of the world; and the debate in epistemology

[5] But see Conee and Feldman 2004, and Conee's contribution to the present volume, for a formulation of the epistemic internal/external divide that is more in line with that in the philosophy of mind and language.

[6] Indeed, it is for this reason that I employ 'semantic' internalism and externalism as an expression covering the variety of particular internalisms and externalisms one finds in the philosophy of mind and language—including those pertaining to (the individuation of) linguistic meaning and mental content.

concerns the materials needed in order to answer these questions. The further question thus arises whether the position that one takes regarding one of these debates forecloses on options one can take on the other. Such a question, in turn, raises a host of broadly methodological issues. Should we impose epistemic constraints on our account of semantic content (mental representation, linguistic reference, etc.)? Alternatively, should we give priority to offering an account of semantic content first, and let the epistemic chips fall where they may? Or should these issues be treated together, aiming at something like a 'reflective equilibrium' in semantics and epistemology? Although these may be perennial questions, it is worth exploring whether the perspective provided by the two I/E debates provides new insights into them, or at least new tools with which to address them.

There is a third, metaphilosophical reason to think that it is worth exploring the possibility of connections between the two I/E debates. Both of the I/E debates have important implications for the relation between philosophy and related empirical disciplines such as psychology, linguistics, and (more generally) cognitive science. It is worth asking whether positions on one or both of the I/E debates have any bearing on metaphilosophical views regarding the relation between philosophy and the related empirical disciplines. A variety of interesting questions emerge. Do one's metaphilosophical views predict one's views on either or both of the I/E debates? (If so, does this reflect anything more than current fashion in philosophy?) Does a move away from the view of philosophy as conceptual analysis give us a reason to prefer one over the other side in either of these debates? Does a settled opinion on a position on one (or both) of the debates, or alternatively a settled opinion on the sort of argument that it would take to settle one (or both) of the debates, suggest how philosophy relates to the empirical disciplines? Although these questions can be considered in connection with either one of the debates taken in isolation from the other, raising them in connection with both allows us to see whether similar metaphilosophical winds blow in both metaphysics and epistemology.

Finally, there is a fourth reason to think that it is worthwhile pursuing the possibility of connections between the two I/E debates. Both debates seem to give a prominent role to skepticism. In the debate in the philosophy of mind and language, some have motivated semantic internalism by appeal to semantic externalism's alleged anti-skeptical implications, whereas semantic externalists have responded either by denying the alleged implications, or else by acknowledging them but going on to argue that this is to be counted as something in its favor.[7] In the debate in epistemology, some epistemic externalists have thought to advertise its anti-skeptical implications as a virtue

[7] There is a cottage industry devoted to determining precisely what these implications are, and what lessons we should draw from the fact that semantic externalism has these implications. The paper that started this discussion is McKinsey 1991. For subsequent discussions, see the various

of the position; whereas some epistemic internalists have responded by saying that such a motivation is patently question-begging—concluding further that epistemic externalism has simply changed the subject in epistemology.[8] With this in mind, we might well hope that a careful comparison of each debate's dialectic with skepticism might reveal whether semantically and epistemically externalist responses, and the internalist reactions to these, permit any generalizations about our engagements with skepticism: about why many find it seductive in the first place, and why, even so, many continue to hold that no fully satisfactory response has yet been offered.

Some (though not all) of these issues are explored in the contributions to this volume. Among the topics explored several emerge as dominant. These are: semantic externalism's implications for epistemic internalism (a theme in the chapters by Brown, Fumerton, and Conee); semantic externalism's implications regarding the epistemic character of reflection and/or reasoning (Pryor, Hawthorne, Sosa, and Goldberg); the connection, if any, between the individuation conditions proposed by semantic externalism, and issues pertaining to the justification of e.g. empirical belief (Brueckner, and Sawyer and Majors); and the prospects for semantically externalist and/or epistemically externalist responses to skepticism (in Pritchard, Henderson and Horgan, and Pryor).

2. INDIVIDUAL CONTRIBUTIONS TO THIS VOLUME

Jessica Brown's contribution, 'Externalism in Mind and Epistemology,' attempts to determine whether Externalism in the Philosophy of Mind poses any problem for Internalism in Epistemology, and she concludes with a qualified affirmative answer. She begins by distinguishing three forms of Epistemic Internalism: Supervenience on the Accessible (SA), according to which 'whether a thinker is justified in believing *p* supervenes on those states to which she has special access;' Access Internalism (AI), according to which 'one has special access to one's justificatory status;' and Supervenience on the Mental (SM), according to which 'whether a thinker is justified in believing p supervenes on that thinker's occurrent and dispositional mental states.' Next, she argues that SA and AI are challenged by Semantic Externalism, but that MI is not. However, Brown goes on to argue that MI is not well motivated. (Compare Earl Conee's contribution to the volume, where MI's connection to Semantic Externalism is explored by a proponent of MI.) One interesting result Brown pursues in the course of making out her argument is this: even if (as is widely conceded) Semantic Externalism is

papers in the volumes edited by Ludlow and Martin 1997; Wright, Smith, and MacDonald 1998; Nucettelli 2004; and Schantz 2004.

[8] For other discussions on this topic, see e.g. Sosa 1994, 2004; Stroud 1994; Greco 2000; Pritchard 2004; Bonjour 2003.

incompatible with Introspective Access to Sameness and Differences of Content, even so Semantic Externalism is still compatible with SA—so long, that is, as Semantic Externalism is compatible with introspective knowledge of content.

Richard Fumerton's chapter, 'What and About What is Internalism?,' raises the question of how best to understand the two 'internalist' positions (in epistemology, and in mind and language). He is anxious to avoid so construing epistemic internalism that a commitment to an externalist position in the philosophy of mind will *ipso facto* make one an epistemic externalist. He suggests that epistemic internalists will do best to construe their thesis as pertaining to the notion of *having a justification to believe*, rather than to the notion of *being justified in believing*. (The latter would involve the epistemic internalist in disputes about the basing relation, and from there to disputes about the nature of causation—disputes that Fumerton regards as better left to the side when formulating a workable epistemically internalist position.) Moving on to the internalism/externalism debate in metaphysics and the philosophy of mind, Fumerton suggests that any externalist position will have to regard the *de re/de dicto* distinction, as it applies to the attitudes, as an ontological distinction, and will have to hold that at least some propositional attitudes are 'in some important sense' *de re*—where this appears to mean that some propositional attitudes have as among the constituents of their propositional contents the very *res* in question. He then notes the various ways open to semantic internalists to resist such a view. A crucial question facing the epistemic internalist (according to Fumerton) is then whether psychological states are among one's internal states: if not, then epistemic internalism reduces to an uninteresting thesis. He goes on to suggest that metaphysical internalism is best construed as a thesis about the sort of state that is left once you 'scrape off' the external conditions that obtain in a case involving a propositional attitude.

Earl Conee's contribution, 'Externally Enhanced Internalism,' aims to show that, while there is a *prima facie* tension between epistemic internalism and semantic (or 'content') externalism, this tension disappears if epistemic internalism is construed as mentalism. According to mentalism, justification supervenes on a subject's mental states. Conee's idea is that if this version of internalism is combined with content externalism, the result is an expansion in the supervenience base for epistemic justification: such a base will now include any 'external' conditions that are relevant to the individuation of the attitudes (and their contents). Conee recognizes that mentalism is not the most popular form of epistemic internalism, but he aims to show that (1) mentalism captures the core motivating intuitions behind epistemic internalism better than more traditional internalist positions do, and (2) in any case mentalism avoids some problems that might be thought to arise, on the assumption of content externalism, for traditional epistemic internalist positions. Conee brings out the latter claim in connection with a case involving incomplete or partial understanding of a sort familiar to proponents of content externalism. Conee argues that such cases

appear to pose a problem for epistemic internalism, and that mentalism has a better account of what is going on in such cases than do more traditional forms of epistemic internalism. (It is worth noting that the problem Conee points to appears to be a cousin of the sort of problem raised in both Sosa's contribution and Goldberg's contribution, in connection with the epistemology of reasoning-based belief.)

Duncan Pritchard's chapter, 'How to be a Neo-Moorean,' examines the dialectic with the skeptic, in order to argue in favor of what he calls a 'Neo-Moorean' position. This position is like the traditional Moorean position in that it holds that we know the truth of anti-skeptical hypotheses, but it is unlike the traditional Moorean position in that it does not think that one can respond to the skeptic simply by asserting the known anti-skeptical proposition in question. (The neo-Moorean holds that it would be conversationally inappropriate to do so in such a context.) The burden of the chapter is to show that the neo-Moorean position is in a dialectically stronger position *vis-à-vis* skepticism than is any one of its main competitors: standard Mooreanism, contrastivism, contextualism, closure-denial, and skepticism itself. In the course of making his case, Pritchard: distinguishes two kinds of skeptical argument, noting that the stronger version (for employing the weaker epistemic principle) is the underdetermination argument; suggests that the existence of the underdetermination argument, together with the fact that this argument is not met by denying closure, undermines most of whatever motive there is for the move to deny closure; argues that the neo-Moorean position is unique among positions in being able to acknowledge what is wrong with Moorean assertion, without abandoning the core intuition behind an anti-luck epistemology; and suggests that, at least when it comes to anti-skeptical hypotheses, discriminability considerations are best seen as bearing on knowledge *ascriptions*, rather than as bearing on *the conditions for possessing knowledge* in the first place. Although Pritchard appears to favor a version of neo-Mooreanism that endorses epistemic externalism, he does not settle whether such a version of neo-Mooreanism is to be preferred to a McDowell-inspired version—one which combines a disjunctivist conception of (the content of) perceptual experience with a commitment to epistemic internalism.

In 'Some Ins and Outs of Transglobal Reliabilism,' David Henderson and Terry Horgan offer an extended defense of 'transglobal reliabilism,' a particular version of reliabilism about justification. As with other forms of reliabilism about justification, transglobal reliabilism holds that justified belief is reliable belief; the difference with standard versions is primarily over how to understand the relevant notion of reliability. Henderson and Horgan motivate their proposal by considering the two main notions of reliability in the literature. One of these is local reliability, that property a belief has when it is formed (and sustained) through belief-forming (and -sustaining) processes that yield a suitably high percentage of true beliefs in scenarios similar to the actual one. The other is the property of global reliability, that property a belief has when it is formed

and sustained through belief-forming and -sustaining processes that yield a suitably high percentage of true beliefs in the variety of situation-types in which those belief-forming and -sustaining processes are likely to be employed (in the actual world). Transglobal reliabilism aims to formulate yet another notion of reliability—transglobal reliability—that captures virtues of each of the above notions, without having any of their drawbacks. The ideal they aim to capture is that of a cognitive system in which reliance on particular belief-forming processes is 'suitably modulated', that is, where the subject relies on a given belief-forming process in all and only those circumstances in which the process itself is likely to be locally reliable—so that the system as a whole is globally reliable. They go on to designate as 'neoclassical reliabilism' the view that a belief is justified when it is the result of a process that is 'globally reliable under suitable modulational control.' The problem that they find with 'neoclassical' reliabilism is that it regards justification as turning on the peculiar aspects of the agent's global environment—with the result that the view yields the wrong verdicts in e.g. New Evil Demon cases. They propose a fix according to which we expand the class of relevant scenarios in which to determine the reliability of a belief-forming process, to include all scenarios that are 'experientially possible'—all scenarios involving experiences that a subject with our cognitive make-up could undergo. A belief-forming process is then judged 'transglobally reliable' if it is 'reliable with respect to the class of experientially relevant possible global environments;' and a belief can be said to be transglobally reliable, and hence objectively justified, just in case it is the result of 'a process that is transglobally reliable under suitable modulational control.' The main advertised virtues of transglobal reliabilism are two. First, the view yields the intuitively correct verdicts in cases in which standard forms of reliabilism do not. And second, unlike other versions of reliabilism, transglobal reliabilism can acknowledge that psychological duplicates are duplicates justification-wise as well—thereby undermining one of the standard epistemically internalist criticisms of reliabilism. Henderson and Horgan conclude by noting that one but not all of the arguments that motivate transglobal reliabilism depends on semantically internalist assumptions.

In their chapter, 'Entitlement, Connection, and Opacity,' Sarah Sawyer and Brad Majors argue that both internalist and externalist accounts in epistemology fail for being incomplete. They contend that the epistemic internalist correctly acknowledges the relevance of a subject's perspective to the epistemic status of her beliefs, but the internalist fails to accommodate the constitutive link between justification and truth; whereas the epistemic externalist accommodates the constitutive link between justification and truth, but fails to acknowledge the relevance of a subject's perspective to the epistemic status of her beliefs. They propose that an adequate account will marry the insights of both epistemic internalism and externalism; and they go on to suggest that such an account is already present in certain anti-individualistic (semantically externalist) views regarding mental content. The core insight of such views, Sawyer and Majors

suggest, is this: the mental contents of a subject's empirical beliefs are individuated in those worlds in which the subject's experiences do in fact provide truth-linked justification to the empirical beliefs she forms on the basis of those experiences. As Sawyer and Majors themselves put it, 'external individuation conditions of a subject's experiential and psychological states serve to delineate a set of environments relative to which reliability yields justification or entitlement.'

Tony Brueckner's chapter, 'Content Externalism, Entitlement, and Reasons,' argues against two recent attempts to suggest a connection between semantic externalism and issues in the justification of perception. Brueckner's targets are Burge (2003) and Brewer (2002). The aim of Burge (2003) was to elucidate the connection between anti-individualism (a version of externalism) and perceptual entitlement; his thesis was that anti-individualism about perceptual content helps to explain why we are epistemically entitled to hold our perceptual beliefs. After characterizing Burge's argument, Brueckner argues that its success depends on an unargued-for assumption regarding the conditions under which reliability considerations make for epistemic entitlement. The aim of Brewer (2002) was to appeal to content externalism as part of an argument—really two distinct arguments, according to Brueckner—whose conclusion was that perceptual experience provides reasons for empirical belief. After characterizing what he takes the two arguments to be, Brueckner develops various criticisms of each. Of the various criticisms I highlight one. Brueckner argues that, while one of Brewer's arguments depends on the premiss that subjects of perceptual experience recognize themselves as confronted by how things are in the non-mental world, this sort of argument faces a fatal dilemma. Either this use of 'recognize' is an epistemic use, or it is not. If it is, then the argument is question-begging, since in that case the claim that a subject so recognizes herself appears to assume what it is that we want shown, namely, that perceptual experiences provide reasons for empirical belief. But if the use of 'recognize' is not epistemic, then the point, that subjects so recognize themselves, would (even if true) be irrelevant to the epistemic conclusion Brewer seeks to draw. Brueckner concludes that to date no sound argument has been given connecting semantic externalism with doctrines pertaining to perceptual justification. (In this respect it is interesting to wonder what Brueckner's attitude would be to the Sawyer–Majors contribution to this volume, which aims to offer an argument for just such a connection.)

Jim Pryor's chapter, 'What's Wrong with McKinsey-Style Reasoning?,' considers and rebuts a McKinsey-style *reductio* argument against Semantic Externalism. According to this argument, Semantic Externalism, together with the thesis that a subject has *a priori* (or at least reflective) knowledge of her occurrent thoughts, yields the implausible conclusion that a subject can know *a priori* (or at least by reflection) details about the empirical nature of the world. Pryor's case against this sort of argument is that even if a thinker can know *a priori* that she is thinking that water puts out fire, she cannot know *a priori*, of the concept that is expressed by her use of 'water,' that it succeeds in being

true of an actual natural kind. Pryor then defends this sort of move against Boghossian's more recent appeal to the possibility of Dry Earth, where not only is there no water, but (more radical still) there is no watery liquid at all (subjects are given to systematic and pervasive illusions on this score). Where Boghossian had argued that the externalist must say that in such Dry Earth cases there is no concept expressed by the empty would-be natural kind term, Pryor holds that the externalist has many other options: perhaps the subject expresses a 'fall-back' descriptive concept, or a concept individuated by a necessarily uninstantiated property. Either way, though, Boghossian is wrong to suppose that the externalist must say that in such Dry Earth cases there is no concept expressed by the empty term. Pryor goes on to offer an analogous suggestion for cases involving empty names, arguing that even if the Evans/McDowell hypothesis, that such cases involve 'illusions of thought,' is correct, this result will not help one to formulate a sound McKinsey-style argument against Semantic Externalism. Pryor concludes by suggesting what his argument tells us about the nature of reflective justification.

In his chapter, 'A Priority and Externalism,' John Hawthorne argues that once we accept semantic externalism, there would appear to be no way to draw an interesting distinction between the *a priori* and the *a posteriori*. In particular, given semantic externalism, several traditional ways of drawing out the relevant notion of the 'empirical independence' of the *a priori* are called into question. Using a safety-theoretical conception of knowledge, Hawthorne glosses *a posteriori* knowledge as the sort of knowledge that depends on the environment's playing the role of a 'safe haven,' and suggests that *a priori* knowledge would then be the sort of knowledge that is not in any need of the environment's playing any such 'safe haven' role. But if this is what the distinction amounts to, Hawthorne argues, it does not withstand serious scrutiny. He characterizes three types of scenario (two of which turn on the assumption of semantic externalism) on which a would-be paradigmatic case of *a priori* knowledge depends on the environment's having to play the role of a safe haven after all. The first scenario is an analogue, in the domain of (would-be) *a priori* knowledge, of the fake barn case: there is an environmental factor that makes the subject's *a priori* judgement luckily true. The second and third scenarios turn on implications of (certain versions of) semantic externalism—the danger of empty thoughts, and the social nature of meaning—both of which will have the implication (Hawthorne argues) that most or all of what is commonly taken to be *a priori* knowledge is actually *a posteriori*. These results turn on a certain conception of the 'environmental independence' conception of the *a priori*; but Hawthorne goes on to suggest that even those conceptions of the *a priori* that conceive of it in terms of 'experiential independence' will fail to mark any epistemologically interesting distinction. Hawthorne concludes by noting that, while there are those who would try to understand the *a priori* as a kind of intellection-generated source of epistemic justification (with states of 'intellectual seemings' playing the role of justifiers),

this view is not without its difficulties. He concludes that the *a priori/a posteriori* distinction does not mark an epistemologically interesting distinction.

David Sosa's chapter, 'The Inference that Leaves Something to Chance,' aims to establish that externalism is incompatible with the conjunction of two theses about inference. These he labels *internalism about inference* and *ignorance is insufficient for incoherence*. According to the first, internal duplicates are inferential duplicates: if one draws an inference then so does the other. According to the latter, inferring subjects are in principle in a position to avoid invalidity, no matter their state of knowledge. The chapter turns on a problem that Boghossian (1992) raised for semantic externalism, in which (owing to considerations pertaining to the external determinants of meaning) a subject's reasoning is guilty of the fallacy of vacillation, yet she is not in a position to discern this *a priori*. In Sosa's terms, Boghossian's argument appears to establish that semantic externalism is incompatible with the ignorance principle above. The semantic externalist can respond to this, Sosa notes, if but only if she surrenders the internalism about inference thesis above. Either way, an independently plausible doctrine regarding inference is incompatible with semantic externalism.

Like Sosa's chapter, Sandy Goldberg's chapter, 'Semantic Externalism and Epistemic Illusions,' also focuses semantic externalism's implications regarding the epistemology of reasoning (or inference-based belief). It, too, does so in connection with Boghossian's 1992 paper. But where Sosa's chapter casts its focus broadly, to include what we might call the metaphysics of inferential belief—its causal basis and so forth—Goldberg's chapter focuses more narrowly on the epistemology of inference-based belief. Goldberg treats Boghossian's argument as establishing that 'brutely external conditions'—conditions whose obtaining does not affect the proper cognitive functioning of the subject in question—are sometimes relevant to the epistemic assessment of reasoning-based belief. Goldberg concludes by noting that this need not undermine one's confidence in the truth of semantic externalism; instead, he suggests, the proper conclusion is to acknowledge yet another case—involving the epistemic internalist's favored epistemic tool, explicit reasoning—in which what is available from the armchair is less epistemically robust than has been previously thought.

The thrust of Joseph Owens's chapter, 'Psychological Externalism and the Role of Belief in the Analysis of Knowledge,' is to establish an incompability claim: Owens argues that psychological externalism is incompatible with the doctrine that beliefs causally explain a subject's behavior. The burden of his argument is to show that, while twins in twin earth cases satisfy different propositional-attitude predicates—so Alf satisfies 'x believes that cans are made of aluminum,' whereas twin-Alf satisfies 'x believes that cans are made of twalum'—the nominalization of these proposition-attitude predicates, in terms of an expression such as 'Alf's belief that cans are made of aluminum,' does not refer to any state that plays a causal role in the explanation of Alf's behavior. The argument for this conclusion is a *reductio*: those who hold that the nominalized expressions do refer to

causally relevant states face a dilemma, according to whether they accept or reject that the content of the state in question is essential to the state itself. Owens argues that both horns end in an unacceptable position: endorsing the former horn is unattractive on its face, whereas endorsing the latter horn undermines the thesis (which many philosophers of mind will want to retain) that belief-desire explanation is causal explanation. Owens concludes that the upshot of psychological externalism is that belief is not a mental state.

REFERENCES

Armstrong, D. 1973: *Belief, Truth, and Knowledge* (Cambridge: Cambridge University Press).

Boghossian, P. 1992: 'Externalism and Inference.' In E. Villanueva (ed.), *Philosophical Issues* 2: 11–28.

Bonjour, L. 2003: 'A Version of Internalist Foundationalism.' In L. Bonjour and E. Sosa, *Epistemic Justification* (Oxford: Blackwell).

Brewer, B. 2002: *Perception and Reason* (Oxford: Oxford University Press).

Brown, J. 2004: *Anti-individualism and Knowledge* (Cambridge, Mass.: MIT Press).

Brueckner, A. 2002: 'The Consistency of Content Externalism and Justification Internalism.' *Australasian Journal of Philosophy* 80: 512–15.

Burge, T. 1979: 'Individualism and the Mental.' Reprinted in D. Rosenthal (ed.), *The Nature of Mind* (Oxford: Oxford University Press, 1991).

____ 2003: "Perceptual Entitlement." *Philosophy and Phenomenological Research* 67.3: 503–48.

Chase, J. 2001: 'Is Externalism about Content Inconsistent with Internalism about Justification?' *Australasian Journal of Philosophy* 79: 227–46.

Conee, E. and Feldman, R. 2004: *Evidentialism* (Oxford: Oxford University Press).

Goldberg, S. 2007a: *Anti-Individualism: Mind and Language, Knowledge and Justification* (Cambridge: Cambridge University Press).

Goldman, A. 1967: 'A Causal Theory of Knowing.' *Journal of Philosophy* 64: 357–72.

Greco, J. 2000: *Putting Skeptics in their Place* (Cambridge: Cambridge University Press).

Kripke, S. 1972: *Naming and Necessity* (Cambridge, Mass.: Harvard University Press).

Ludlow, P. and Martin, N. (eds.) 1997: *Externalism and Self-Knowledge* (Palo Alto, Calif. CSLI Publications).

McDowell, J. 1994: *Mind and World* (Cambridge, Mass.: Harvard University Press).

McKinsey, M. 1991: 'Anti-Individualism and Privileged Access.' *Analysis* 54: 124–8.

Nuccetelli, S. (ed.) 2004: *New Essays on Semantic Externalism and Self-Knowledge* (Cambridge, Mass.: MIT Press).

Peacocke, C. 2004a: 'Explaining Perceptual Entitlement.' In R. Schantz (ed.), *The Externalist Challenge* (Berlin: Walter de Gruyter).

____ C. 2004b: *The Realm of Reason* (Oxford: Oxford University Press).

Pritchard, D. 2004: *Epistemic Luck* (Oxford: Oxford University Press).

____ and Kallestrup, J. 2004: 'An Argument for the Inconsistency of Content Externalism and Epistemic Internalism.' *Philosophia* 31: 345–54.

Putnam, H. 1975: 'The Meaning of "Meaning".' In *Mind, Language and Reality: Philosophical Papers* (Cambridge: Cambridge University Press).

Schantz, R. (ed.) 2004: *The Externalist Challenge* (Berlin: Walter de Gruyter).

Sosa, E. 1994: 'Philosophical Scepticism and Epistemic Circularity.' *Proceedings of the Aristotelian Society* suppl. vol. 68: 263–90.

____ 2004: 'Circularity and Epistemic Priority.' In R. Schantz (ed.), *The Externalist Challenge* (Berlin: Walter de Gruyter).

Stroud, B. 1994: 'Scepticism, ''Externalism,'' and the Goal of Epistemology.' *Proceedings of the Aristotelian Society* suppl. vol. 68: 291–307.

Sturgeon, S. 1997: 'Rational Mind and its Place in Nature.' In M. Sainsbury (ed.), *Thought and Ontology* (Milan: FrancoAngeli).

Vahid, H. 2003: 'Content Externalism and the Internalism/Externalism Debate in Justification Theory.' *European Journal of Philosophy* 11: 89–107.

Williamson, T. 1998: 'The Broadness of the Mental: Some Logical Considerations.' *Philosophical Perspective* (*Noûs Supplement*) 12: 389–410.

Wright, C., Smith, B., and Macdonald, C. (eds.) 1998: *Knowing our Own Minds* (Oxford: Oxford University Press).

1

Externalism in Mind and Epistemology

Jessica Brown

1. INTRODUCTION

Prima facie, the debates between 'internalism' and 'externalism' in the philosophy of mind and epistemology are not obviously connected. In the philosophy of mind, the debate between those I will call 'semantic internalists' and those I will call 'semantic externalists' concerns thought content; in epistemology, the debate between those I will call 'epistemic internalists' and those I will call 'epistemic externalists' is best understood as concerning justification. However, despite the difference in their subject matters, it turns out that the two debates are connected. In particular, I argue that if, as is widely claimed, semantic externalism undermines the traditional idea that one has special access to one's thought contents, then it may undermine those forms of epistemic internalism which involve the idea that one has special access to one's justificatory status or the facts which determine justification. However, a version of epistemic internalism which has recently become prominent seems to escape the threat posed by semantic externalism. This version of epistemic internalism claims neither that one has special access to one's justificatory status nor that one has such access to the facts which determine justification. Instead, this version claims that justification is determined by facts which are 'internal' in the sense that they are mental, rather than in the sense that they are specially accessible. In the second part of the chapter, I examine whether there is reason to accept this alternative construal of epistemic internalism.

2. VARIETIES OF INTERNALISM IN EPISTEMOLOGY

Epistemic internalism is the view that a thinker's epistemic status depends wholly on matters which are 'internal' to that thinker, rather than at least partially on

Thanks to the participants at the Kentucky conference on externalism in mind and epistemology, and especially Sandy Goldberg, for helpful comments on the chapter.

matters which are 'external' to her, such as her relations to her environment. Let epistemic externalism be the denial of epistemic internalism. Given the truth condition for knowledge, it seems that internalism about knowledge is not a viable position. So, the internalism–externalism controversy in epistemology seems better understood as a debate about justification: does justification depend wholly on matters which are internal to the thinker, or also partly on matters which are external?

Different formulations of epistemic internalism result from different ways of filling out the notion of the 'internal'. One approach takes a thinker's mental states as the paradigm exemplars of the internal, and argues that whether a thinker is justified in believing p supervenes on that thinker's mental states. I will call this view, 'Supervenience on the mental' or 'SM.'

Supervenience on the mental (SM): whether a thinker is justified in believing p supervenes on that thinker's occurent and dispositional mental states.

In understanding SM, it is useful to distinguish two notions of justification, the notion of whether a thinker has a justification for her belief that p, and the notion of whether her belief that p is justified, or well grounded. Even if a thinker has reasons which provide justification for her belief that p, her belief might not be justified if she believes that p not for those good reasons but for some alternative bad reasons. Whether a belief is justified or well grounded depends on its causal history, how it was formed at some possibly distant past time. Prima facie, SM is most plausibly a view about the notion of having a justification for a belief, rather than the notion of a belief being justified or well grounded. For facts about the causal history of a belief needn't be reflected in the subject's current mental states (Greco 2005: 266). Indeed, this is the way that Conee and Feldman, two prominent recent defenders of SM, understand the view. Conee and Feldman defend what they call 'mentalism,' defined as the view that 'The justificatory status of a person's doxastic attitudes strongly supervenes on the person's occurent and dispositional mental states, events and conditions' (Conee and Feldman 2001: 234). Feldman (2005) makes it explicit that SM concerns the notion of having a justification for a belief, rather than the notion of its being well grounded (275). However, later in the chapter we will discuss a different version of SM defended by Wedgwood, a version which is concerned with the notion of a belief's being justified or well grounded. On Wedgwood's view, 'the rationality of belief supervenes purely on internal facts about the thinker's mental states,' where he defines 'internal fact' to apply to 'any fact that supervenes purely on the thinker's 'non-factive' mental states, and also to any fact about the explanatory relations in which such internal facts stand to each other' (Wedgwood 2002: 2).[1]

[1] See also Alston's 'perspectival internalism,' the view that to confer justification something must be 'within the subject's "perspective" or "viewpoint" on the world, in the sense of being something the subject knows, believes, or justifiably believes' (1986: 68–9). He contrasts this view with what he calls 'access internalism,' the view that one has direct access to justifiers and

Whereas SM understands the notion of the internal in terms of the notion of the mental, on a distinct approach, the notion of the internal is understood as that to which a thinker has 'special access.' On this latter approach, epistemic internalism is the view that whether a thinker is justified in believing p supervenes on those states to which she has such special access. I will call this view 'Supervenience on the accessible,' or SA for short.

Supervenience on the accessible (SA): whether a thinker is justified in believing p supervenes on those states to which she has special access.[2]

There are a range of possible views about what it means for a thinker to have special access to a state. Some epistemic internalists understand the relevant notion of access to be the notion of what's directly or non-inferentially accessible (e.g. Alston 1986: 92). A more common understanding takes the relevant notion to be the notion of what's accessible by reflection (e.g. Audi 1998: 233–4; Bernecker and Dretske 2000: 65–6).[3] However the notion of special access is filled out, it is implicitly relative to certain background conditions, e.g. that one is not drunk or asleep, and that one has the required conceptual repertoire. One cannot know directly (or indirectly) that one believes that p if one lacks the concept of belief. It seems that SA is most plausibly understood as applying to the notion of having a justification for a belief, rather than the notion of a belief's being justified or well grounded. For, it seems implausible that one has special access to the causal history of one's beliefs (cf. Greco 2005: 266).

On a variety of ways of filling out the notion of special access it seems that SM and SA are distinct, for there may be mental states which are not accessible in the relevant sense, and there may be non-mental states which are so accessible. Mental states which are unconscious or repressed may provide examples of mental states to which the subject can gain access only by inference from information about her behaviour or environment. In that case, SM does not entail SA. Further, on some views, there are features which are accessible in the relevant sense but not mental, or not part of one's current mental life. For instance, on a direct realist view, one might have direct access to the world through perception. On

their justificatory efficacy (92, 99). Sosa distinguishes 'Cartesian internalism' from 'Chisholmian internalism.' According to Cartesian internalism, 'justification requires only really proper thought on the part of the subject . . . where the appropriateness of the thought is a matter purely internal to the mind of the subject, and not dependent on what lies beyond,' whereas Chisholmian internalism holds that 'one can find out directly, by reflection, what one is justified in believing at any time' (1999: 147).

[2] Conee and Feldman call this 'accessibilism' (Conee and Feldman 2001: 233); Pryor calls it 'simple internalism', defined as the view that 'whether one is justified in believing that p supervenes on facts which one is in a position to know about by reflection alone' (2001: 104). Bernecker and Dretske initially define internalism in terms of access to the factors which determine justification, but later take it to involve access to one's justificational status (2000: 65–6).

[3] An alternative way of construing the notion of special access would be in terms of the strength of the access, rather than its route e.g. in terms of the idea that one's access is infallible. I omit this possible understanding in the main text since it's not standard among internalists. See also n. 6.

one view of our access to our past mental states, one's past thoughts are directly accessible through preservative memory, but not part of one's current mental life.[4] If justification supervenes on features which are specially accessible but not mental, then SA holds, but SM does not.

A last version of epistemic internalism agrees with SA in construing the notion of the internal in terms of special access, but holds that one has such access to one's justificatory status, rather than to the facts which determine that status. I will call this view 'Access internalism,' or AI for short.

Access internalism (AI): one has special access to one's justificatory status.[5]

As with SA, it seems more plausible to understand AI as concerned with the notion of whether one has justification for a belief, rather than the notion of whether it is justified or well grounded. For one does not have special access to the causal history of one's beliefs. In principle, one could attempt to understand the relevant notion of access as what's directly accessible. However, that would make the relevant internalism grossly implausible for it seems that knowing the justificational status of some beliefs requires complex reasoning.[6] It seems better, then, to construe the notion of special access in AI as the notion of what's reflectively accessible. As with SA, the notion of special access in AI is implicitly relative to certain background conditions, e.g. that one is not drunk or asleep, and that one has the required conceptual repertoire. It would be implausible to suppose that one has special access to one's justificational status if one lacks the concept of justification or if one's cognitive functioning is impaired due to drugs or alcohol.

AI is distinct from both SM and SA. Suppose that whether one is justified in believing p supervenes on one's mental states and assume that one has direct access by reflection to one's mental states so that both SM and SA hold. Still if having reflective access to whether one's justified requires simultaneously grasping very large sets of mental states, and complex relations between them, then it might be that one lacks reflective access to whether one's justified. For, grasping very large sets of mental states simultaneously, or the complex relations among those states, might be beyond one's ability (e.g. Sosa 1999: 148).[7]

[4] Bonjour makes this point, drawing on Burge's notion of preservative memory (1992: 133).

[5] This definition is based on Pryor's definition of 'access internalism' as the view that one always has special access to one's justificatory status (2001: 105). Compare Alston's 'access internalism,' which combines the views that justifiers and their justificatory efficacy are directly accessible (1986: 68, 99). Sosa calls the view 'Chisholmian internalism' and rejects it (1999). AI is defended in Bonjour 1985: chs. 1–2 and Chisholm 1977: ch. 6, S. 5.

[6] For similar reasons, it seems implausible to understand the notion of special access in AI as the notion of infallible access.

[7] There are a range of hybrid views which incorporate internalist and externalist elements. For instance, Alston 1986 embeds the internalist condition that one has direct access to the justifiers or grounds of one's beliefs within a theory which also contains externalist conditions.

3. INTERNALISM IN THE PHILOSOPHY OF MIND

Semantic internalism is a thesis in the philosophy of mind about thought content, i.e. what is believed, desired, hoped, and feared, etc. We can distinguish semantic internalism and its opposite, semantic externalism, by contrasting relational properties, such as the property of being taller than the Eiffel Tower, and intrinsic properties, such as the property of being made of iron. Semantic internalism treats the property of having a certain thought content as an intrinsic property, whereas semantic externalism treats it as a relational property.

Let Oscar and Twin Oscar share all their non-intentional intrinsic properties: they are molecule-for-molecule physical duplicates; they share all the same dispositions to behave, where behaviour is construed as bodily movement non-intentionally described; and they share all the same 'narrow' functional properties, i.e. functional properties defined in terms of causal relations within the body. Semantic externalism claims that Oscar and Twin Oscar may have different thought contents if they are in different environments. Semantic internalism claims that Oscar and Twin Oscar have the same thought contents. Thus we may understand the dispute between semantic externalism and internalism in terms of supervenience. Semantic internalism claims, while semantic externalism denies, that thought content supervenes on the non-intentional properties which duplicates like Oscar and Twin Oscar share.

At first sight, there may not be any obvious connections between the positions called 'internalism' and 'externalism' in philosophy of mind and epistemology. The internalist/externalist debates in semantics and epistemology have different subject matters, namely thought content and justification. Prima facie, one could combine semantic internalism and some form of epistemic externalism. For instance, one could hold that thought content supervenes on the non-intentional properties which duplicates share, but hold that the justification of a subject's thoughts is partly determined by their causal relations to the environment. Further, two of the varieties of epistemic internalism distinguished above, namely AI and SA, seem concerned with a notion of the internal distinct from that which figures in semantic internalism (cf. Pryor 2001: 103). These two varieties of epistemic internalism are concerned with the notion of the internal as that to which one has special access. SA holds that justification supervenes on features which are specially accessible. AI claims that one has special access to one's justificational status. By contrast, semantic internalism claims that thought content is internal in the sense that it supervenes on the (non-intentional) properties which duplicates share.[8] A property could be shared by duplicates, yet not specially accessible, such as a sub-personal property of one's nervous system.

[8] But see Farkas 2003 for the view that the debate between semantic internalism and semantic externalism is best understood in terms of whether two subjects who are in subjectively

In principle, a property could be specially accessible yet not shared by duplicates. For instance, as mentioned above, on a direct realist view of perception, certain features of the environment may be directly accessible.

The remaining version of epistemic internalism, namely SM, claims that justification supervenes on the internal, understood as the mental. This last version of epistemic internalism also uses a notion of the internal distinct from that used in semantic internalism. SM's characterization of the internal as mental cannot be used to characterize the debate in the philosophy of mind about whether thought content is internal or external. Rather, the latter debate focuses on the question whether thought content supervenes on the non-intentional properties shared by duplicates. Further, the distinctive claim made by SM, namely that justification supervenes on the mental, seems consistent with the claim that thought content is external in the sense used in semantic externalism, namely that it does not supervene on the non-intentional properties shared by duplicates. Indeed, one of the prime defenders of SM, Wedgwood, explicitly states that it is designed to be compatible with semantic externalism (2002: 2 n. 2).

Although the internalism–externalism controversies in mind and epistemology initially seem to be rather different, perhaps they may be connected by focusing on the consequences of these positions, rather than the positions themselves. It's been widely claimed that semantic externalism undermines the traditional idea that one has special access to one's thought contents. Plausibly, whether one has justification for a belief depends in part on the content of one's other beliefs. So, if semantic externalism undermines one's special access to one's thought contents, then it may undermine at least those versions of epistemic internalism which involve special access to one's justificational status and/or the factors which determine that status.[9] However, one version of epistemic internalism, SM, involves no claim that one has special access either to one's justificational status or to the factors which determine that status. SM merely states that justification is determined by, or supervenes on, a thinker's mental states, and makes no claim about the accessibility of those states. As a result, it is not threatened by semantic externalism even if the latter position threatens one's special access to one's thought contents.

In the next section, I attempt to fill out the thought that semantic externalism may undermine at least certain versions of epistemic internalism.[10] I distinguish a number of different claims about the kind of access one has to one's thought

indistinguishable states can have different thought contents (196). Since Farkas claims that 'first person authority extends only as far as things are subjectively indistinguishable' (203), it seems that she characterizes the debate as one about whether thought content supervenes on states to which one has authoritative first-person access.

[9] Cf. Bonjour 1992: 136.

[10] By contrast Chase 2001 denies that semantic externalism is inconsistent with epistemic internalism for, he suggests, epistemic internalists need not hold that one has special access to one's thought contents, whether that is construed in terms of a priori or direct knowledge. However, I

contents and examine how semantic externalism may undermine epistemic internalism by undermining one or more of these claims.

4. SEMANTIC EXTERNALISM AND ACCESS

It is sometimes suggested that semantic externalism undermines a subject's ability to know her thought contents introspectively, where to say that she knows them introspectively is to say that she knows them without inference from observations of her own behaviour or environment. In other words, it is alleged to undermine introspective knowledge of content, or IKC:

IKC: an individual can know the contents of her occurrent thoughts without inference from observation of her behaviour or environment.[11]

A key argument[12] for the claim that semantic externalism is incompatible with IKC uses a slow switch example in which a subject, S, is unwittingly switched between two environments in such a way that the semantic externalist would accept that S's thought contents are affected. For instance, we may suppose that S is unwittingly switched between Earth and a faraway planet, Twin Earth, which is exactly like Earth except that the liquid in lakes and rivers and which falls as rain is not water but a duplicate substance with a different chemical composition, twater. The case is so set up that semantic externalists would accept that, on Earth and before the switch, S thinks that water is wet, but that on Twin Earth as a result of the switch, S thinks that twater is wet. Incompatibilists argue that S lacks introspective knowledge of her thought contents. Take a time at which S thinks that water is wet. The incompatibilist argues that S cannot know introspectively that she thinks that water is wet, for she cannot introspectively distinguish the actual situation in which she thinks that water is wet and the twin situation in which she instead thinks that twater is wet. S is switched in such a way that she is unaware of the switch in her environment. Further, it seems that she is unaware of the switch in her thoughts. She would fail to notice when she starts to have thoughts involving the concept twater. If, after the switch, she is asked whether the concept she expresses by 'water' has changed, she would deny that it has. The incompatibilist argues that in order for S to know that she thinks that water

focus on one standard interpretation of epistemic internalism as the view that either the determinants of justification or one's justificational status are 'specially accessible.'

[11] The terms IKC and IKCC are taken from Falvey and Owens 1994, although their definition of IKCC does not use the qualification that the two thoughts be simultaneous. See also Goldberg's similar distinction between knowledge of content and discriminative knowledge of content (Goldberg 1999). In the context of this chapter, the temporal qualification is important since, plausibly, only a thought held at the same time as a second affects whether one has a justification for the latter.

[12] Other arguments include the memory argument and McKinsey-style reductio arguments (see Boghossian 1989; McKinsey 1991; Brown 1995; Boghossian 1997).

is wet, she needs to use empirical information about the kind of environment she is in.

Semantic externalists typically reject the claim that semantic externalism is incompatible with IKC. They point out that even the slow switch subject cannot be mistaken in her judgements about her thought contents. The environment determines the concepts she uses in both her first-and second-order thoughts, so there cannot be a mismatch between the thoughts she ascribes to herself at second order and her first-order thoughts (e.g. Heil 1988; Burge 1986). In general, her second-order judgements about what she thinks are reliable. Further, as Burge (1986) points out, one's second-order thoughts of the form, I judge that I think that p, are self-verifying. Judging that one is thinking that p is itself a case of having some thought toward the content p.

This reply has not settled the controversy about whether semantic externalism is incompatible with IKC. Incompatibilists have argued that the self-verifying character of judgements of the relevant form is not sufficient to make those judgements knowledge (e.g. Vahid 2003; Brown 2004). There is not space to resolve this debate here. However, we can consider what the consequences would be for epistemic internalism were it to turn out that the slow switch subject is a counterexample to IKC. A first step in this assessment it to determine what slow switch cases show about normal humans. Some have suggested that ordinary humans do not undergo slow switches and so such scenarios do not undermine the ability of ordinary humans to have introspective knowledge of their thought contents. On pain of scepticism, not all possibilities undermine knowledge, but only those which are 'relevant' or 'nearby.' So, it may be suggested that the slow switch scenario is not a relevant possibility for ordinary humans and so does not undermine their introspective knowledge of their thought contents (Warfield 1997; Sawyer 1999; Brown 2004). By contrast, others have argued that ordinary humans do undergo slow switches, for instance when they move between two linguistic communities which use the same term in different ways; e.g. the terms 'chips,' 'pavement,' 'professor,' and 'football' are used in different ways in UK and US English. On this view, slow switches are a quotidian phenomenon (Falvey and Owens 1994; Ludlow 1995; Gibbons 1996). I will not attempt to adjudicate this dispute about whether ordinary humans undergo slow switches. Rather, I will outline the consequences for epistemic internalism under the worse case scenario in which slow switches undermine IKC and ordinary humans frequently undergo such switches.

If slow switches undermine IKC and ordinary humans frequently undergo such switches, then ordinary humans lack introspective knowledge of at least some of their thought contents. On this worse case scenario, semantic externalism undermines epistemic internalism. If one can know some of one's thought contents only by inference from observation of one's behaviour or environment, then one lacks special access to one's thoughts in either of the ways epistemic internalists understand that notion (as direct knowledge or reflective knowledge).

If one lacks special access to at least some of one's thought contents,[13] one lacks special access to some of the factors which determine whether some of one's beliefs are justified and SA is false. If one lacks special access to some of the factors which determine whether some of one's beliefs are justified, it also seems that one lacks special access to the justificational status of some of one's beliefs, and so AI is false.

Although there is disagreement about whether semantic externalism is incompatible with IKC, there is widespread agreement that semantic externalism is incompatible with a different claim about one's access to one's thoughts, which we may call Introspective Knowledge of Comparative Content, or IKCC for short:

IKCC: with respect to any two occurrent thoughts, or thought constituents, which an individual has at a time t, at t she can know whether they have the same content or not without inference from observation of her behaviour or environment.

Notice that even if semantic externalism is incompatible with IKCC, it does not follow that it is incompatible with IKC. As we have seen, the standard defence of the compatibility of semantic externalism and IKC stresses the self-verifying nature of second-order judgements of the form, I judge that I think that p. As semantic externalists have pointed out, the slow switch subject's judgements of the form, I judge that I think that p, are self-verifying even though she cannot distinguish introspectively between water and twater thoughts. Indeed, semantic externalists who defend the compatibility of semantic externalism and IKC standardly accept the incompatibility of semantic externalism and IKCC (e.g. Falvey and Owens 1994).

If semantic externalism undermines IKCC then it may also undermine AI. Suppose that, as a consequence of semantic externalism, a thinker has two thoughts but cannot know introspectively whether they have the same or different content. Whether two of one's thoughts have the same or different content affects the logical and rational relations between these thoughts. In turn, those relations affect the justification one has for those thoughts. So if, as a consequence of semantic externalism, a subject cannot tell introspectively whether two of her thoughts have the same or different content, this undermines her ability to know by reflection whether some of her beliefs are justified.

It may be helpful to consider some standard examples of a failure of IKCC and examine how they undermine reflective access to justificational status. IKCC may fail in two different ways. A subject may have two thoughts, or thought constituents, with the same content, and yet be unable to know introspectively that they have the same content. Alternatively, a subject may have two thoughts, or thought

[13] Notice that on the minority view that the notion of special access in SA amounts to the claim that one has infallible access, semantic externalism does not undermine SA. For, as we have seen, semantic externalism is compatible with the claim that one's judgements of the form, I judge that I think that p, are self-verifying. However, this construal of special access is not plausible for AI (see n. 6).

constituents, with different contents, and yet be unable to know introspectively that they have different contents. I consider an example of each kind of case.

Suppose that Rudolph[14] partially understands the terms 'coriander' and 'cilantro' and defers to experts for their correct explication. He knows that each term names a herb, although he is ignorant that they name the same herb. Rudolph knows cilantro as the fresh herb used in Mexican cooking and is familiar with its appearance and flavour. By contrast, he knows coriander as a dried herb and is again familiar with its appearance and flavour. In virtue of these facts, it seems that a semantic externalist would accept that Rudolph has the concepts expressed by 'cilantro' and 'coriander.' Further, since 'cilantro' and 'coriander' have the same referent, many[15] semantic externalists would accept that Rudolph expresses the same concept by these terms. However, since Rudolph is ignorant of the fact that the terms are co-referential, he is not in a position to discover that he expresses the same concept by these two terms without using empirical information. It seems, then, that Rudolph is a counterexample to IKCC. That Rudolph cannot know introspectively that he expresses the same concept by 'cilantro' and 'coriander' may undermine his ability to know by reflection the justificational status of some of his beliefs.

For instance, suppose that Rudolph has the belief he would express by saying 'Coriander was known to the ancient Romans.' Further, suppose that Rudolph also has the beliefs he would express by saying 'Cilantro was brought to Europe after Columbus' expedition to the Americas' and 'Columbus' expedition occurred after the fall of ancient Rome.' On the semantic externalist account under consideration, Rudolph expresses the same concept by his use of 'cilantro' and 'coriander.' So, Rudolph's further beliefs undermine his first belief. However, given that Rudolph needs to use empirical information to know that he expresses the same concept by 'cilantro' and 'coriander', it seems that he cannot know by reflection that these further beliefs undermine his first belief. So, Rudolph's inability to know introspectively that he expresses the same concept by 'cilantro' and 'coriander' undermines his ability to know by reflection the justificational status of the belief he expresses by 'Coriander was known to the ancient Romans.'

Now consider a failure of IKCC in which a subject has two thought constituents with different contents yet cannot know that they have different contents introspectively. Slow switch cases are standardly used to provide examples of this sort. It is useful to look at Goldberg's variant of a slow switch case which overcomes some of the problems raised for the original slow switch examples

[14] This example is from Falvey and Owens 1994.

[15] Though not those that combine semantic externalism with the notion of Fregean sense. However, the notion of Fregean sense cannot be similarly used to challenge examples in which the subject has two thoughts constituents with different contents but cannot tell that they have different contents. (See Brown 2004: ch. 5.)

(Goldberg 1999).[16] Goldberg's variant case focuses on a subject, Sally, who is a committed semantic externalist who also has reason to think that she's been a victim of a slow switch. In particular, Sally thinks that some of the beliefs she would express with the term 'pragmatist' were acquired in one linguistic community in which 'pragmatist' has one meaning, and some were acquired in a distinct linguistic community in which that term has a different meaning. Goldberg assumes the two-concept interpretation of slow switch cases in which, after the switch, the subject retains the relevant pre-switch concept at the same time as acquiring a new post-switch concept.[17] So Sally has reason to think that some of the beliefs she would express with 'pragmatist' involve one concept and others involve a different concept. As a result, it seems that Sally may be unable to tell introspectively whether some of her thoughts have the same or different contents. For instance, suppose that Sally is not sure when she formed the beliefs she would express by (1) 'Pragmatists have no principles' and (2) 'Jones is a pragmatist.' Given that Sally is unsure when she formed these two beliefs, it seems that she would be unsure whether they involve the same concept or not. In order to know whether they do, she needs to use empirical information about when these beliefs were acquired and the kind of environment she was then in. Further, Sally's inability to know introspectively whether the beliefs she expresses by (1) and (2) involve the same concept undermines her ability to know by reflection the justificational status of some of her beliefs. For instance, suppose that Sally has the belief she would express by (3) 'Jones has no principles.' Whether Sally has justification for her belief depends on whether it is supported by her other beliefs. For instance, if the beliefs she would express by (1) and (2) involve the same concept, then they would provide reason for the belief she would express by (3). But, since Sally needs empirical information to know whether they involve the same concept, she cannot know by reflection whether these beliefs provide a justification for her belief (3).[18]

As we have seen in the case of Rudolph and Sally, if a subject has two thought constituents but cannot know introspectively whether they have the same or different content, this undermines her ability to know by reflection the justificational status of some of her beliefs. It seems that many of us may find

[16] The original slow switch examples are open to Burge's objection that a subject making an inference would intend to use the same concept throughout so that there cannot be equivocation between the steps of that inference. Since, in Goldberg's variant, the subject suspects that she has been switched, it would not be rational for her to form such an intention (Goldberg 1999).

[17] Not all semantic externalists accept the two-concept view, although many do. For a case in which the subject has two thoughts with different contents and cannot know introspectively that they have different contents but which does not require the two- concept view, see Frances 1999.

[18] A further way in which semantic externalism may undermine AI is by undermining explicative knowledge of content. On Burge's view, a subject may have certain thoughts without fully understanding them. As a result, certain of her beliefs may provide reason for others, even though she cannot know this a priori (Brown 2000).

ourselves in situations analogous to those of Rudolph and Sally. Like Rudolph, we may acquire two distinct terms for a single natural kind and understand them well enough to be credited with the relevant concepts, yet fail to realize that they are co-referential. Like Sally, we may suspect that we have been unwittingly switched between two linguistic communities which use a certain term in different ways, but be unsure which of our beliefs was acquired in one and which in the other. Such scenarios are not far-fetched but seem ordinary and everyday. If semantic externalism is true and many of us regularly find ourselves in situations akin to those of Rudolph and Sally, then we are regularly in situations in which we cannot know by reflection the justificational status of some of our beliefs and AI is false.

Although the cases of Rudolph and Sally can be used to show that semantic externalism may undermine AI, it is less clear that they provide reason to suppose that semantic externalism undermines a different version of epistemic internalism, SA. According to SA, justification is determined by, or supervenes on, things to which a thinker has special access, whether that is understood as direct access, or reflective access. Plausibly, the rational relations between one's mental states supervene on their contents, and the attitudes one has to those contents. As a result, even if semantic externalism undermines IKCC, it needn't undermine SA as long as it doesn't undermine special access to one's thought contents or the attitudes one holds to those contents. [19]

Let us summarize the conclusions arrived at in this section. We have seen that semantic externalism potentially threatens those versions of epistemic internalism that build in a requirement of special access. However, which of these versions of epistemic internalism are threatened depends on what type of knowledge of content is undermined by semantic externalism and the prevalence of failures of that type of knowledge. Compatibilists and incompatibilists agree that cases like those of Rudolph and Sally show that semantic externalism undermines IKCC. As a result, all should agree that if many of us regularly find ourselves in situations akin to those of Rudolph and Sally, then semantic externalism undermines AI. Compatibilists argue that, although semantic externalism undermines IKCC, it does not undermine IKC. If semantic externalism does not undermine

[19] Semantic externalism has also been alleged to undermine one's introspective knowledge of the attitudes one bears to one's thought content, or IKA for short:

IKA: an individual can know the attitude component of her occurrent attitudes without inference from observation of her behaviour or environment.

Whether a belief is justified depends not only on the content of one's other mental states, but also one's attitudes to those contents. For instance, the belief that X was in Paris at the time of the murder in London may epistemically justify the belief that X is innocent, but the desire that X was in Paris at the relevant time does not do so. As a result, if semantic externalism undermines introspective knowledge of the attitudes one bears to one's thought contents, it seems to undermine both AI (one has special access to the justificational status of one's beliefs) and SA (one has special access to the facts which determine justification). Similarly, it may undermine Alston's internalist condition, that one has special access to justifiers.

introspective knowledge of thought content or of attitudes to thought content then, although semantic externalism undermines AI, it does not undermine SA. By contrast, incompatibilists argue that slow switch cases show that semantic externalism does undermine IKC. If they are right, and slow switch cases are common for normal humans, then semantic externalism undermines not only AI but also SA.

Although semantic externalism threatens those versions of epistemic internalism that build in a requirement of special access, it does not threaten SM, a version of epistemic internalism which does not build in such a requirement. Given that SM is not potentially threatened by semantic externalism, it is useful to examine whether there are any reasons to prefer SM, rather than SA and AI, as our understanding of epistemic internalism. In the next section, I examine two recent attempts to argue for epistemic internalism construed as SM.

5. SM

The main recent proponents of SM are Conee and Feldman, and Wedgwood. Conee and Feldman's positive[20] defence of SM consists in a set of six examples, each of which concerns a pair of subjects who differ in the level of justification they intuitively have for a given belief. Conee and Feldman argue that 'these contrasts are best explained by supposing that internal differences make the epistemic difference' (236). They suggest that we can draw a general conclusion from these cases:

> It is reasonable to generalise from these examples to the conclusion that every variety of change that brings about or enhances justification either internalises an external fact, or makes a purely internal difference. It appears that there is no need to appeal to anything extramental to explain any justificatory difference. (238)

However, as Conee and Feldman acknowledge the cases hardly constitute a knockdown argument for epistemic internalism understood as SM. They define SM as the thesis that whether a thinker is justified in believing p supervenes on that thinker's occurrent and dispositional mental states; externalism is the denial of this claim. So, externalists can accept that sometimes internal differences make for a difference in justification, so long as they deny that justification wholly supervenes on internal facts. Thus, even if Conee and Feldman are correct in thinking that the difference in justification in their cases results from an internal difference, that is compatible with externalism. What's crucial is not what is the correct analysis of the six cases, but rather whether one can generalize from those cases to the claim that all differences in justification result from internal

[20] They also argue negatively that SM can meet a diverse range of objections levelled at internalism.

differences. But Conee and Feldman provide no reason for thinking that the six cases exhaust all possible cases, admitting that they 'have no proof that there is no exception to the pattern exhibited by the examples' (238). In the light of these problems, let us examine the rather different defence of a version of SM, that offered by Wedgwood.

Wedgwood explicitly focuses on the notion of the 'rationality' of belief, according to which 'the rationality of belief supervenes purely on "internal facts" about the thinker's mental states' (2002: 2). He makes it clear that he considers this account of rational belief as part of the wider debate between internalism and externalism about justification, on the assumption that 'a belief is "justified" if and only if it is rational' (1 n. 1). At first sight, Wedgwood's view may seem similar to that of Conee and Feldman. However, in fact, it is rather different since he defines 'internal fact' to apply to 'any fact that supervenes purely on the thinker's "non-factive" mental states, and also to any fact about the explanatory relations in which such internal facts stand to each other' (14). So, for Wedgwood, but not for Conee and Feldman, the supervenience base for rationality/justification includes facts about causal explanatory relations. Relatedly, Wedgwood's primary interest is in rational belief revision, rather than whether one has a justification for a belief.

Wedgwood takes it that internalists must defend their view by providing an explanation of it. In particular, he thinks that internalists must explain 'which facts about a thinker count as these "internal facts" upon which the rationality of a belief or decision supervenes . . ., and, also explain why rationality should supervene on internal facts in this way' (3). Wedgwood argues against the traditional understanding of internalism as SA and for his own version of internalism by arguing that his view better fulfils the explanatory task.[21] To see if Wedgwood's view succeeds in doing so, we need to better understand his conception of this explanatory task. We can do so by examining his rejection of SA (the view that the justification (or rationality) of belief supervenes on facts to which one has special access).

One common way of explaining the traditional understanding of internalism appeals to the ideas that 'to say that a belief is rational is just to say that in holding the belief, the thinker is proceeding in a cognitively blameless fashion,' and that 'one cannot fairly be blamed for not responding to a fact that one was not in a position to know' (4). Wedgwood agrees with other commentators that this common defence fails. First, even if the notion of rational belief is identical with the notion of cognitively blameless belief, it is not clear that the argument motivates the claim that one has special access to the factors which determine the rationality of belief. At best, the argument seems to show that

[21] 'I shall propose an alternative conception of rationality; and in SS. 4–5, I shall argue that this alternative conception provides a better explanation of what exactly these "internal facts" are, and of why it is that the rationality of beliefs and decisions supervenes on them' (3).

one knows, or can know, the factors which determine the rationality of belief (4–5; see also Goldman 1999: 271–93). Second, the notion of rational belief is not identical with the notion of cognitively blameless belief (see also Pryor 2001; Plantinga 1993). Consider a subject who uses a flawed principle to arrive at a belief, although she is cognitively blameless in doing so. Perhaps she is using the principle on the authority of an acknowledged expert whom she has no reason to distrust, and she lacks the cognitive capacity to understand herself that the principle is incorrect. Even though we are supposing that she is cognitively blameless in using the principle, since it is in fact flawed her belief is not rational.[22]

It seems that Wedgwood criticizes the standard defence of internalism on two distinct grounds: he complains that it makes an implausible proposal about what it is to say that a belief is rational (namely that the thinker is epistemically blameless in holding that belief) and that this proposal would not result in the relevant version of internalism. From his discussion, we can draw two criteria for the adequacy of any explanation of internalism. First, the explanation should make a plausible proposal about what it is to say that a belief is rational and, second, it should show how this proposal results in internalism. These two conditions do seem to be reasonable criteria for any explanation of internalism of the relevant form. Wedgwood attempts to meet the first criterion by proposing that rational belief revision consists in following certain kinds of rules. He attempts to meet the second by arguing that his proposal about the nature of rational belief when combined with certain facts about the nature of psychological explanation results in internalism. I will consider each part of his view in turn.

In more detail, Wedgwood claims that rational belief revision is a matter of following or being guided by what he calls 'basic' rules which are of the form (Ø if C) where it makes sense for one to conform to those rules in order to pursue the aim of truth (S3). Central to this proposal is the notion of a basic rule, which we can elucidate by considering the notion of following one rule by following another. One may follow the rule 'Stop when the light is red' by following the different rule 'Stop when one believes the light is red.' Wedgwood suggests that one may follow the truth rule 'Believe that p if and only if p is true' by following the rule 'Believe p whenever one has an experience as of p's being the case (and no special reason to distrust one's experience in the circumstances).' By contrast, if a rule is basic one can follow it 'directly;' one's following it 'cannot be analysed at the folk-psychological level of explanation into a series of sub-processes that include one's following any other rule' (12).

It is crucial for Wedgwood's internalism that he defines rational belief revision in terms of following basic rules. Suppose, by contrast, that rational belief revision were defined in terms of following non-basic rules, such as the truth rule (believe that p if and only if p is true). In that case, what is rational for a normal person

[22] Wedgwood also argues that SA is undermined by Williamson's anti-luminosity arguments (Williamson 2000: ch. 4).

and her mental duplicate who is unfortunately an evil demon victim might be very different. For, their beliefs have different truth values. For instance, the normal person's belief that she has hands is true, but the demon victim's belief that she has hands is false. If a belief were rational if and only if true, then it would be rational for the normal person, but not her demon twin, to believe that she has hands.

Let us consider, then, how plausible it is to elaborate the concept of rational belief revision in terms of following basic rules. Now Wedgwood claims only to be making a proposal about rational belief which he argues explains internalism, rather than arguing for this proposal.[23] Nevertheless, as we saw in Wedgwood's own rejection of the traditional understanding of internalism as SA, a proposal provides a good explanation of internalism only if that proposal is itself plausible. He criticized the traditional view on the grounds that it implausibly equated the notion of belief being rational and the notion of a thinker's being blameless in holding it.

Why, then, should we hold that rational belief revision is a matter of revising one's beliefs by following basic rules? According to Wedgwood, we follow some rules by following others, but basic rules are ones which we follow 'directly.' These points about how we follow rules seem to be points about the psychological explanation of our following rules. But, it's not clear why a point about psychological explanation should establish any substantive claim about the nature of a normative notion, such as the notion of being rational. Consider an analogy. On one popular view, assertion is governed by the rule, assert p only if one knows that p (e.g. Williamson 2000: ch. 11). Someone who is trying to follow this rule may do so by asserting p only if she believes that she knows that p. Wedgwood would presumably put this in terms of the subject's following one rule, namely the knowledge rule for assertion, by following another, say the rule, assert p only if one believes that one knows that p. However, even if one follows the former rule for assertion by following the latter, this psychological fact about how one follows the rule for assertion does not establish anything about the normative notion of the correctness of an assertion. In particular, even if one follows the knowledge rule by following the rule, assert p only if one believes that one knows that p, that does not show that one's assertion is correct if one believes that one knows that p. Rather, according to the knowledge rule for assertion, one's assertion that p is correct only if one knows that p. Similarly, even if one follows the truth rule, believe p if and only if p is true, by following the distinct rule, believe p whenever one has an experience as of p's being the case and no special reason to distrust one's experience in the circumstances, it does not follow that a belief is rational if it accords with the latter rule.

Perhaps Wedgwood might try to argue for the claim that rational belief revision is a matter of following basic rules as follows. Consider again the person

[23] See the quote at n. 21.

who follows the knowledge rule for assertion by following the different rule, assert p only if one believes that one knows that p. Consider a case in which she asserts p since she reasonably but falsely believes that she knows that p. According to the knowledge rule for assertion, she lacks warrant for her assertion. However, despite this, her assertion is understandable. Further, we wouldn't blame her for her assertion; after all, she reasonably believed that she knew that p and so it was reasonable for her to assert that p (Williamson 2000: 256–7). Similarly, consider someone who follows the truth rule, believe p if and only if p is true, by following the distinct rule, believe p whenever one has an experience as of p's being the case and no special reason to distrust one's experience in the circumstances. In a case in which she believes p on the basis that she has an experience as of p's being the case and no special reason to distrust her experience in the circumstances, but p happens to be false, it seems that she is not blameworthy for believing p. She was doing the best she could under the circumstances. It seems, then, that the most obvious way for Wedgwood to defend the idea that rational belief is a matter of following basic rules is by arguing that (1) a belief revision is rational if and only if the agent is epistemically blameless in revising her belief in that way, and (2) it is by reference to basic rules that we judge whether a subject is so blameless.

However, there is an obvious objection to the suggested method of justifying the idea that rational belief is a matter of following basic rules. As Wedgwood himself points out in his rejection of the standard defence of internalism, the notion of justified or rational belief is not equivalent to the notion of epistemically blameless belief. As a result, Wedgwood would not want to use the proposed defence of the elaboration of rational belief in terms of following basic rules. But, it is not obvious how else one could defend that elaboration of the notion of rational belief. It seems, then, that Wedgwood's proposed explanation of internalism faces a problem similar to that which faced the traditional explanation, namely that it is not clear why one should accept the proposed elaboration of the notion of rationality. The traditional explanation explicitly suggested that the notion of a belief being rational is equivalent to the notion of the thinker being cognitively blameless in holding it. Wedgwood explicitly suggests that the notion of being rational is equivalent to the notion of following certain basic rules. It is not at all clear what would motivate this latter equivalence other than an implicit appeal to the traditional claim that the notion of being rational is equivalent to the notion of being cognitively blameless.

Even if Wedgwood could find some defence for his proposed elaboration of the notion of rational belief, it is a further question whether this proposed elaboration results in internalism. Wedgwood argues that this conception of rational belief yields internalism when combined with certain facts about psychological explanation. Although we needn't examine all the details of his argument, it is useful to have the main points at hand. At the heart of Wedgwood's argument is his claim that 'whenever a thinker revises her beliefs through following a rule, a fully-articulated folk psychological explanation of that belief revision will

identify the proximate explanation of that revision with an internal fact about one's mental states,' where he defines the expression 'internal fact' to apply to 'any fact that supervenes purely on the thinker's "non-factive" mental states, and also to any fact about the explanatory relations in which such internal facts stand to each other' (14). Wedgwood uses this claim to argue that each of the following is an internal fact: that it makes sense for one to conform to the rule, (Ø if C) (S5); that one is following it (S4); and, condition C (S4). Since, on his conception, the rationality of belief is just a matter of forming beliefs by following basic rules of the form (Ø if C) which it makes sense for one to conform to, he thinks that this establishes that the rationality of belief supervenes on facts which are internal in his sense.

Some may object to Wedgwood's claim that his proposed elaboration of rational belief results in internalism by objecting to one of the key claims he uses in that argument, namely the claim that whenever a thinker revises her beliefs through following a rule, a fully articulated folk psychological explanation of that belief revision will identify the proximate explanation of that revision with an internal fact about one's mental states. Some may argue that the proximate explanation of a belief revision need not be an internal fact, and might instead be a factive mental state.[24] Here, I leave such worries aside and consider whether, even if we grant Wedgwood's claim about psychological explanation, his proposal about rational belief results in internalism.

Prima facie, it seems implausible that an internalist view could result from understanding rational belief revision in terms of whether the subject's belief revision results from her following rules. The notion of whether a belief revision results from following rules seems to involve the notion of the causal explanation of that revision. But, at first glance, the fact that a belief revision can be causally explained in a certain way does not obviously count as an intuitively internal fact. Greco (2005) suggests that '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' (266). Although Greco's comment seems to use the notion of the internal as the accessible, his point extends to the notion of the internal as the mental. Even if a belief revision was caused by another mental state it is not obvious that the fact that the belief revision was caused in a certain way is itself a mental fact. On standard analyses of the notion of causation, the truth of a singular causal statement depends on the truth of certain generalizations and counterfactuals. But, it is far from obvious that the fact that certain generalizations and counterfactuals obtain is a mental

[24] For instance, Williamson. Bird forthcoming objects to Wedgwood on the different ground that it is difficult to characterize the set of mental states on which, according to Wedgwood, justification supervenes. In particular, it is not sufficient to specify the set as those mental states which are non-factive since some non-factive mental states still seem external in a problematic way, e.g. mental states which entail that a certain proposition be false. The state of refuting p may be an example of such a state.

fact. Further, that a belief revision was caused in a certain way, namely by one of the subject's past mental states, does not supervene on the subject's current mental states.

Interestingly, Wedgwood makes no substantial case for the claim that if one mental state is caused by another, the fact that it is so caused is an internal fact. He simply stipulates that he is using the expression 'internal fact' to apply to 'any fact that supervenes purely on the thinker's "non-factive" mental states, and also to any fact about the explanatory relations in which such internal facts stand to each other' (14). The absence of any justification for this move is curious in the light of the significant role it plays in his argument for the claim that his elaboration of rational belief in terms of following rules results in internalism. For instance, his stipulative definition of the notion of the internal is crucial to his argument that when one revises a belief by following a basic rule, that one is following the rule is an internal fact. In more detail, he argues that when one forms the belief that p by following a basic rule of the form (Ø if C), that one is following that rule is a fact about an explanatory relation that holds between one's coming to believe p and C, where C itself is an internal fact. By his definition, the notion of an 'internal fact' applies to any fact that supervenes purely on the thinker's "non-factive" mental states, and also to any fact about the explanatory relations in which such internal facts stand to each other. So, it is a simple consequence of his account of following a basic rule and his definition of the notion of the internal that the fact that one is following a basic rule in revising one's belief counts as an internal fact.[25]

The role of this stipulative definition undercuts Wedgwood's claim to have explained internalism. To be worthy of the name internalism, a view must show that rational belief supervenes on states which are in some intuitive sense internal. For instance, a view would not count as internalist if it held that rational belief supervenes on 'internal facts,' where the latter are stipulated to include relations between a subject and her environment. As Greco's challenge highlights, it is far from clear that causal explanatory relations between non-factive mental states are themselves internal facts. It doesn't answer this challenge merely to stipulate that one is using the expression 'internal facts' in such a way that it applies to such causal explanatory relations. Rather, one needs to provide an account of why it is reasonable to regard such causal explanatory relations as internal, by appeal to the established use of that term. Otherwise, regardless of one's stipulation about one's use of the term 'internal,' it seems that one's view is not reasonably

[25] 'Suppose that one revises one's beliefs by directly following this basic rule: one comes to believe p in response to the internal fact that one has an experience as of p's being the case, and no reason to distrust one's experience in the circumstances. Then the fact that one is directly following this rule is itself a fact about a certain explanatory relation that holds between this internal fact and one's coming to believe p. That is, the fact that one is directly following this rule is itself a fact about the explanatory relations in which one internal fact stands to another. So, given my definition of the term, the fact that one is directly following the rule is itself an "internal fact"' (23).

regarded as internalist and one has merely conceded the externalist's point by another name.

6. CONCLUSION

Although the internalism/externalism controversies in philosophy of mind and epistemology concern different subject matters, namely thought content and justification, it turns out that there is a deep connection between them. If, as is widely claimed, semantic externalism undermines the traditional notion that one has special access to one's thought contents, then it may undermine those versions of epistemic internalism which build in the claim that one has special access to the facts which determine justification or one's justificatory status. However, which versions of epistemic internalism are threatened depends on the type of knowledge of content undermined by semantic externalism. In more detail, compatibilists and incompatibilists agree that one can construct scenarios which show that semantic externalism is incompatible with IKCC. If ordinary humans frequently find themselves in such scenarios, then semantic externalism undermines AI, the claim that one has special access to one's justificatory status. By contrast, it is controversial whether semantic externalism is incompatible with IKC. If slow switch cases show that semantic externalism is incompatible with IKC, and slow switches are a regular occurrence for ordinary humans, then semantic externalism undermines not only AI but also SA, the claim that one has special access to the facts which determine justification. One form of epistemic internalism is immune to the challenge posed by semantic externalism, namely, SM which involves neither the claim that one has special access to the facts which determine justification nor the claim that one has special access to one's justificatory status. However, I have argued that recent defences of this version of epistemic internalism fail. In lieu of any other defence, it seems that there is no defensible form of epistemic internalism which is not potentially threatened by semantic externalism.

REFERENCES

Alston, W. P. 1986: 'Internalism and Externalism in Epistemology.' *Philosophical Topics* 14: 179–221. Reprinted in Kornblith 2001.

Audi, R. 1998: *Epistemology: A Contemporary Introduction to the Theory of Knowledge* (New York: Routledge).

Bernecker, S. and Dretske, F. 2000: *Knowledge: Readings in Contemporary Epistemology*. (Oxford: Oxford University Press).

Bird, A. forthcoming. 'Justified Judging.' *Philosophy and Phenomenological Research.*

Boghossian, P. 1989: 'Content and Self-Knowledge.' *Philosophical Topics* 17: 5–26.

——1992: 'Externalism and Inference.' In E. Villanueva (ed.), *Philosophical Issues* 2: 11–28.

——— 1994: 'The Transparency of Mental Content.' *Philosophical Perspectives* 8: 33–50.
——— 1997: 'What the Externalist Can Know A Priori.' *Proceedings of the Aristotelian Society* 97: 161–75.
Bonjour, L. 1980: 'Externalist Theories of Empirical Knowledge.' *Midwest Studies in Philosophy* 5: 53–74. Reprinted in Kornblith 2001.
——— 1985: *The Structure of Empirical Knowledge* (Cambridge, Mass.: Harvard University Press).
——— 1992: 'Externalism/Internalism.' In J. Dancy and E. Sosa (eds.), *Blackwell Companion to Epistemology* (Oxford: Blackwell).
Brewer, B. 1999: *Perception and Reason* (Oxford: Oxford University Press).
Brown, J. 1995: 'The Incompatibility of Anti-individualism and Privileged Access.' *Analysis* 55: 149–56.
——— 2000: 'Critical Reasoning, Understanding, and Self-Knowledge.' *Philosophy and Phenomenological Research* 61: 659–77.
——— 2004: *Anti-individualism and Knowledge* (Cambridge, Mass.: MIT Press).
Brueckner, A. 2002:'The Consistency of Content Externalism and Justification Internalism.' *Australasian Journal of Philosophy* 80: 512–15.
Burge, T. 1979: 'Individualism and the Mental.' *Midwest Studies in Philosophy* 4: 73–121.
——— 1988: 'Individualism and Self-Knowledge.' *Journal of Philosophy* 85: 649–63.
Chase, J. 2001: 'Is Externalism about Content Inconsistant with Internalism about Justification.' *Australasian Journal of Philosophy* 79: 227–46.
Chisholm, R. 1977: *The Theory of Knowledge* (Eaglewood Cliffs, NJ: Prentice Hall).
Conee, E. and Feldman, R. 2001: 'Internalism Defended.' In Kornblith 2001.
Falvey, K. and Owens, J. 1994: 'Externalism, Self-Knowledge, and Skepticism.' *Philosophical Review* 103: 107–37.
Farkas, K. 2003: 'What is Externalism?' *Philosophical Studies* 112: 187–208.
Feldman, R. 2005: 'Justification is Internal.' In M. Steup and E. Sosa (eds.), *Contemporary Debates in Epistemology* (Oxford: Blackwell).
Frances, B. 1999: 'Contradictory Belief and Epistemic Closure.' *Mind and Language* 14: 203–26.
Gallois, A. and O'Leary-Hawthorne, J. 1996: 'Externalism and Scepticism.' *Philosophical Studies* 81: 1–26.
Gibbons, J. 1996: 'Externalism and Knowledge of Content.' *Philosophical Review* 105: 287–310.
Goldberg, S. 1999: 'The Relevance of Discriminatory Knowledge of Content.' *Pacific Philosophical Quarterly* 80: 136–56.
Goldman, A. 1999: 'Internalism Exposed.' *Journal of Philosophy* 96: 271–93.
Greco, J. 2005: 'Justification is Not Internal.' In M. Steup and E. Sosa (eds.), *Contemporary Debates in Epistemology* (Oxford: Blackwell).
Heil, J. 1988: 'Privileged Access.' *Mind* 97: 238–51.
Kornblith, H. (ed.) 2001: *Epistemology: Internalism and Externalism* (Oxford: Blackwell).
Ludlow, P. 1995: 'Externalism, Self-Knowledge and the Prevalence of Slow Switching.' *Analysis* 55: 45–9.
McKinsey, M. 1991: 'Anti-Individualism and Privileged Access.' *Analysis* 51: 9–16.
Plantinga, A. 1993: *Warrant: The Current Debate* (New York: Oxford University Press).

Pryor, J. 2001: 'Highlights of Recent Epistemology.' *British Journal for the Philosophy of Science* 52: 95–124.

Sawyer, S. 1999: 'An Externalist Account of Introspective Knowledge.' *Pacific Philosophical Quarterly* 80: 358–78.

Sosa, E. 1999: 'Scepticism and the Internal/External Divide.' In J. Greco and E. Sosa (eds.), *The Blackwell Guide to Epistemology* (Oxford: Blackwell).

Vahid, H. 2003: 'Externalism, Slow Swithing and Privileged Self-knowledge.' *Philosophy and Phenomenological Research* 66: 370–88.

Warfield, T. 1997: 'Externalism, Self-Knowledge, and the Irrelevance of Slow Switching.' *Analysis* 57: 282–4.

Wedgwood, R. 2002: 'Internalism Explained.' *Philosophy and Phenomenological Research* 66: 349–69. Page references to the version on his website.

Williamson, T. 2000: *Knowledge and its Limits* (Oxford: Oxford University Press).

2

What and About What is Internalism?

Richard Fumerton

Internalism/externalism controversies rage on in both epistemology and the philosophy of mind. Obviously, it is critical to any satisfactory discussion of these issues that we get clear about just exactly what constitutes an internal state and what constitute the targets of one's internalist/externalist analyses. So, for example, Conee and Feldman (2004) often identify their epistemic internalism with the thesis that all evidence is identical with the mental states of a believer—they call this thesis mentalism. But in an afterword to 'Internalism Defended' (81) they acknowledge that the existence of the internalism/externalism debate in the philosophy of mind might require a more subtle characterization of their epistemological internalism. In general, it seems to me that if an epistemic internalist understands mental states as the psychological externalist suggests, then that epistemic 'internalist' identifying evidence with mental states is really an externalist 'in disguise.'

Even after internalists get clear about just exactly what makes a state 'internal' they must obviously be careful in identifying the target of their internalist analysis. As Luper-Foy (1985) pointed out, for example, few self-proclaimed epistemic internalists will want to extend their internalism to the analysis of knowledge. If one endorses a truth condition for knowledge and one allows that one can have knowledge of truths about the external world, then clearly there will be external conditions necessary for knowledge of those truths. Similarly, while the metaphysical internalist in the philosophy of mind may want to identify *some* intentional states with internal states, such internalists should be very careful before endorsing an internalist thesis concerning the truth makers of so-called *de re* ascriptions of belief.

As someone interested in defending both epistemic and psychological internalism, I will try in this chapter to identify both a plausible and interesting concept of a state's being internal, and a plausible and interesting account of precisely about what an internalist should be an internalist.

1. EPISTEMIC INTERNALISM

While the terminology itself suggests that the epistemic internalist is advancing a thesis about the connection between exemplifying certain epistemic properties and being in certain internal states, it is far from clear that this lies at the heart of all epistemological internalisms. I have argued elsewhere (1996) that it is important to distinguish a number of quite different theses associated with epistemic internalism. The 'internal state' epistemic internalist does indeed identify a person's exemplifying certain epistemic properties with that person's being in certain internal states. But epistemic internalism has come to be associated just as closely with a variety of theses concerning the need for actual or potential *access* to the epistemic properties of one's beliefs.[1] So, for example, the *actual* access epistemic internalist argues that there is a necessary connection between a belief's being justified for S and S's having introspective noninferential reason to believe that the belief is justified. The *potential* access epistemic internalist argues that there is a necessary connection between a belief's being justified for S and S's being *able* to access that fact, where usually the potential access is also construed as a priori or introspective. While self-proclaimed internalists often stress these access requirements, the view is still probably not unrelated to the more natural interpretation of internalism as a thesis about the connection between epistemic states and internal states. After all, it was long assumed by a great many philosophers in the empiricist tradition that it is a mark of 'internal' mental states that one has a kind of unproblematic introspective access to such states. But however common it was to assume a strong connection between being in internal states and having at least the capacity to access introspectively those states, it is more than a bit dangerous to claim that it is an *essential* feature of all internal states that we have the capacity to introspectively access them. Although I won't argue the point here, it does seem to me that such a view threatens various forms of regress, most of them vicious. As we shall see later, rejecting access or potential access as a defining characteristic of the mental does not preclude appealing to such access as *a* mark of the mental.

In any event, let's return to the idea suggested by the very label for the view, the idea that the internalist wants to identify someone's exemplifying certain epistemic properties with that person's being in certain internal states. Clearly, a precise understanding of such a view would require a precise understanding of what is meant by an internal state. The first decision we might make concerns the interpretation of state. Is 'state' in this context shorthand for state of affairs—so the contrast is between internal states of affairs that involve S and external states of affairs that involve S? Or should we construe the two kinds of states as two

[1] Jessica Brown makes a similar point in her chapter in this volume.

kinds of properties? In the end it might not make much difference. If a state of affairs is understood, roughly, in terms of an object's exemplifying certain properties, then the difference between internal states of affairs in which S is a constituent and external states of affairs in which S is a constituent is probably going to be understood in terms of the critical properties of which the state of affairs is partially constituted.

As a first pass, we might suppose that the internal properties of S are *nonrelational* properties of S. S's internal states are constituted by S's exemplifying nonrelational properties. External states of S are constituted by S's exemplifying certain relational properties where at least one of the relata is an entity other than S. Physicalists might already be uneasy, however, for they are most likely to understand the internal states that are the most plausible candidates for the states upon which epistemic properties supervene to be brain states. And it is more than a bit controversial to suppose that someone S is simply identical with his brain. The exemplification of nonrelational properties by S's brain is not, therefore, unproblematically identical with the exemplification of nonrelational properties *by S*. While physicalism is, I believe, false, we don't want to beg that sort of question here. In order to avoid doing so, we should probably allow that the internal states of S include not only those states identical with S's exemplifying nonrelational properties, but also those states identical with some part of S exemplifying nonrelational properties. There aren't all that many hard-core dualists left, but those sympathetic to substance dualism will probably be happy to construe the internal states in question as nonrelational properties of the self or mind that constitutes the essence of S.

The difficulty with the above characterization of internal states is that it will inevitably exclude from the internalist camp a great many philosophers who have traditionally been thought of as internalists. So, for example, Conee and Feldman clearly want to identify S's sensations as internal states of S. While they don't, to my knowledge, take a position on the ontological analysis of sensation, I suspect they won't consider their internalism refuted should it turn out that the sense-datum theorist has the correct analysis of sensation. On a sense-datum theory, a sensation of S, whether veridical or not, always involves S's standing in a *relation* to an object, a sense datum, that exemplifies phenomenal properties (colors and shapes, for example, in the case of visual sense data). S's having a sensation is not to be understood in terms of S's exemplifying a nonrelational property. To be sure, there is the adverbial theorist's controversial alternative to the sense-datum theory. On the adverbial theory, the grammatical structure of the language philosophers have employed in trying to capture the occurrence of sensation is ontologically misleading. Just as Jones's dancing the waltz should probably not be understood in terms of some action, dancing, performed upon some object, the waltz, so also, Jones's feeling pain should not be construed as Jones standing in some relation of feeling to pain. Pain just is a *way* of feeling. And sensing phenomenal blue is just a way of sensing. Put more abstractly

still, the adverbial theorist is likely to construe sensing as the exemplification of a nonrelational property. And if the adverbial theorist extends that analysis to all experiential states, including intentional states like belief, desire, fear, and thought, and that adverbial theorist goes on to defend epistemic internalism, then that adverbial theorist might be content to identify the internal states upon which epistemic properties supervene as the exemplification of nonrelational properties. But again, it is a bit odd to suppose that choosing sides on the controversy between adverbial theorists and sense-datum theorists will necessarily determine whether or not one is an epistemic internalist.

If we want a broader understanding of internal state, we might allow that S's internal states can include the exemplification by S (or constituents of S—hereafter I'll for convenience omit this qualification) of relational properties, but only if the relata of these relations include nothing other than 'mind-dependent' objects. So some classic sense-datum theorists held that sense data, while genuine objects of awareness, exist only so long as one is aware of them. Paraphrasing Berkeley, their existence is their being the objects of awareness. Still others might allow that sense data of which we are unaware can exist but their existence is nevertheless tied to the existence of the person to whom they 'belong.' On this revised understanding of internal state, then, we can still construe the occurrence of sense data 'in' S and S's awareness of such sense data as internal states of S. Of course, sense-datum theorists like Moore also thought long and hard about whether one should countenance the possibility of sense data existing outside of mind. Perhaps at least some sense data should be identified with (logical) parts of physical objects.[2] And it is again a bit odd to suppose that if Moore, for example, woke up one morning and decided that the sense data that are constitutive of some sensations are not perceiver dependent, he would necessarily move into the externalist camp when he also thought that the epistemic status of S's beliefs depend in part on the sensations that S has.

There is a related worry. On at least some classic foundationalist views, our justification for believing necessary truths consists in a direct awareness of the truth makers for such truths. While some radical empiricists often thought that necessary truths were made true by relations between ideas (Hume, for example), Russell and Frege sought to identify the relevant truth makers with Platonic sorts of entities and relations that hold between such entities—relations between universals or Fregean senses. Neither Russell's universals nor Frege's senses are in any obvious sense mind dependent, so if the justification we have for believing necessary truths consists in whole or in part of direct awareness of such entities, our justification will not be internal according to the above criteria.

[2] Ayer, for example, would talk of physical objects as being constructed out of actual and possible sense data. He probably didn't mean it, however. It would be a horrible mistake to construe the existence of an actual sense datum as a necessary condition for the existence of a physical object.

Nothing much hinges on the terminology we adopt. We could just identify Russell and Frege as among the first epistemic externalists. On the other hand, they seem to me to have much more in common with paradigmatic epistemic internalists than contemporary epistemic externalists. And I think the reason is that Russell and Frege would both identify the mind-independent objects which are partially constitutive of epistemic justification as objects of direct awareness or direct acquaintance. Furthermore, Russell, at least, would allow that we have introspective access to the fact that we are in such states. And that suggests yet another way in which one might broaden the category of internal state if one wants to include as internalists the likes of Russell and Frege. Specifically, one might identify S's internal states with S's exemplifying nonrelational properties, S's exemplifying relational properties where the relata are themselves mind dependent, or lastly, S's exemplifying the relational property of being directly aware of some object(s) where S has introspective access to the fact that S has such awareness.

Now the concept of direct awareness as understood by classical foundationalists is, of course, anathema to most contemporary philosophers. In ordinary discourse there is a nontechnical use of 'is aware that' that is more or less synonymous with 'knows that.' The question 'Are you aware that the store closes at 9:00?' usually just means 'Do you know that the store closes at 9:00?' To separate genuine from spurious epistemic internalists, then, it is crucial that we emphasize the philosophically technical sense of awareness in our attempt to expand our understanding of internal states. Williamson, no internalist, will be content to identify evidence with objects of awareness if objects of awareness get understood in terms of truths that are known.[3] So if we are getting at any interesting concept for use by internalists with all this talk of direct awareness, we will have to understand direct awareness the way Russell did, as a sui generis relation that holds between a person and some object, property, or state of affairs. Unlike knowledge, direct awareness does not need to take as its object a bearer of truth value.[4] When I am directly acquainted with pain, for example, there is nothing in this relational state of affairs that can plausibly be identified with a bearer of truth value. Indeed, it is this feature of the traditional concept of the 'given' that has led many to question the plausibility of direct awareness playing an *epistemic* role. That is, however, a controversy I have addressed elsewhere.[5] Here, I am primarily concerned with establishing a framework within which to define interesting and important epistemological and metaphysical controversies.

[3] Williamson 2000 claims that S's evidence is simply identical with what S knows.

[4] Though it can—one can be acquainted with one's thoughts and one's thoughts can be true or false.

[5] See Fumerton 1996, 2001.

2. FINDING THE RIGHT PROPERTY ABOUT WHICH TO BE AN EPISTEMIC INTERNALIST

I have already pointed out in my introductory remarks that an internal state epistemic internalist will obviously need to be careful to restrict appropriately internalist theses about epistemic properties. If we understand internal states in the way discussed above, then it will be more than a bit difficult to identify S's *knowing* some proposition P with S's being in some internal state. If we can know truths about an external world and knowledge involves a nonredundant truth condition for what is known, then it is difficult to see how S's internal states can exhaust the conditions required for knowledge. At least that seems to follow on the supposition that at least some of what is known is constituted by facts with which one is not directly acquainted.

It is perhaps less commonly recognized, however, that one should be equally cautious before identifying S's believing P justifiably with S's being in some internal state. Most epistemologists accept a distinction between there being justification for S to believe P and S's believing P justifiably. It seems intuitively plausible to suppose that S may possess perfectly good epistemic reasons to believe P without believing P. Moreover, even if S believes P, that belief might not be appropriately *based* on the justification S possesses. When S's belief is justified, S not only believes P when S has justification for believing P, but S *bases* his belief that P on the relevant justification. Not everyone agrees that there is a basing relation crucial for justification,[6] and it is an understatement to suggest that among those who insist on a basing requirement, there is no consensus on just how to understand basing. But it is at least tempting to suppose that when S appropriately bases his belief that P on available justification, that justification must play a causal role in producing or sustaining the belief.[7] In any event, if the *causes* of a belief are relevant to its epistemic status, the plausibility of identifying the epistemic status of a belief with internal states of the believer depends critically on one's account of causation. If a regularity theory of causation were true, for example, then the existence of a causal connection between S's justification and S's belief would be hostage to past, present, and future correlations between certain properties—a complex fact indeed, and certainly a fact that involves more than the internal states of an individual believer. And even if one adopts some version of a necessitarian account of causation, the relata of the necessary

[6] See Lehrer 1971 and Foley 1987: 174–5.

[7] And even that may not be enough. As Nozick 1981 and others have pointed out, S's belief might be causally overdetermined. There may be both good epistemic reasons sustaining the belief and bad epistemic reasons that also causally support the belief. The epistemic status of S's belief might then depend on which reasons causally dominate, a matter that is probably most plausibly defined in terms of counterfactuals.

connection and the necessary connection itself are unlikely to be internal states of a believer (in the sense defined above).

The moral to draw, then, is that internalists might be on safer grounds if they restrict their internalist thesis to the property of having justification to believe, rather than the property of having a justified belief.[8] This suggestion should be doubly appealing to those internalists who stress the importance of actual or potential introspective access to the conditions that constitute justification. However one understands causation, it seems to me patently absurd to suppose that one can discover through introspection what is causing one to believe what one believes. To be sure, I often am subjectively confident that my philosophical views are based on nothing but my careful consideration of sound arguments with plausible premisses, but I hope I am not so naive as to rule out the possibility that I am caused by epistemically irrelevant facts to find attractive some of the views I hold. I am certainly not so naive as to rule out the possibility that other philosophers I know are caused by epistemically irrelevant facts to find attractive *their* views.[9]

If we do restrict our internalism to a thesis about the existence of justification, there is at least hope that we can identify the conditions constituting justification with internal states in the broad understanding of 'internal' discussed above. In the case of *inferential* justification, the key to defending such an internalism would be a defense of a Keynesian (1921) concept of epistemic probability as an internal (a different sense of the term) relation[10] holding between evidence and the conclusion that is inferentially justified. If the evidence is constituted by internal states (sensations, apparent memories, belief states, and the like) and Keynes is right, then the existence of those internal states may be metaphysically sufficient for a probability relation holding between those states and propositions made probable by them. Indeed, I would suggest that if the kind of internal state evidentialism defended by Conee and Feldman, for example, is correct, then either radical skepticism is true or there are relations of making probable as understood by Keynes.

On my own (rather unpopular) view, the mere existence of such probability connections between the justification one possesses and other propositions does not, in fact, constitute justification, at least ideal justification, for believing those other propositions. Rather, the existence of inferential justification requires *awareness* of the connections. But again, on our liberal interpretation of internalism, Keynesian probability relations, like relations between universals, are just the sorts of things with which one can become directly acquainted. That direct acquaintance will yield a priori justification in the case of necessary truths

[8] Again, this is a point Brown makes in her chapter.

[9] The fact a statistically high number of students find plausible the views of their teachers is surely not a remarkable coincidence.

[10] An internal relation is one that necessarily obtains when its relata exist.

while it yields only a posteriori justification in the case of contingent truths. But the source of noninferential a priori and a posteriori justification—direct acquaintance—is the same.

3. PSYCHOLOGICAL INTERNALISM

If the epistemic internalist must be careful to identify the right target of an internalist analysis, the psychological internalist walks a veritable minefield. Putnam (1975) distinguished intentional states with narrow content (so-called 'bracketed' beliefs, or, following Dennett, 'notional' world views) from those with broad content, and once seemed amenable to the suggestion that we might identify intentional states having narrow content with internal states. Burge (1979) and Davidson (1987) are clearly suspicious of the idea that there are genuinely intentional states that have so-called narrow content, and Putnam (1996: p. xxi) seems to concede that he shouldn't have introduced the idea. The dispute centers around the issue of whether anything genuinely and exclusively internal could constitute an *intentional* state. But even if nothing genuinely internal constitutes an intentional state, it is not clear that one should restrict psychological or mental states to *intentional* states, and the internalist conceding radical externalism about *intentional* states might still find a plausible mental or psychological state on which to stake an internalist claim.

As noted in the introduction, even the more ambitious internalist who wants to identify at least *some* intentional states with internal states must nevertheless still respect *de re/de dicto* distinctions. The nature of such distinctions is itself a topic for a paper or a book, and here I will confine myself to making a few very general remarks. In trying to slog through *de re/de dicto* distinctions the first conclusion one must reach concerns whether or not the distinction involves an *ontological* distinction between two fundamentally different kinds of intentional state, or whether it involves merely a *linguistic* distinction between two quite different ways in which we *report* or *describe* intentional states. Those construing the distinction as ontological are apt to take the terminology seriously and construe *de re* beliefs as literally constituted, in part, by the objects that figure in their content. So to take a somewhat plausible example, one might suppose that when I believe of the pain that is occupying my attention right now that it is severe, the pain itself is literally a constituent of the belief state. When Russell (1912) held that belief is best construed as involving polyadic relations beween the believer and various entities, he in effect also took belief to be fundamentally *de re*—though on deep analysis the *res* in question were most often universals instead of particulars (and whatever it is, if anything, that quantifiers 'refer' to). Indeed, on this view of belief, it may not make sense to suppose that there are any *de dicto* beliefs (though there are beliefs that do not include as constituents particulars). Russell thought that the entities partially constitutive of belief states

were always restricted to entities with which we are directly acquainted. And for this reason he thought he had direct access (again through acquaintance) to the state of affairs that constituted his believing some proposition. Given our earlier account of internal states, belief states could be both *de re* and internal.

Contemporary externalists about content, by contrast, bring into belief states objects as distant in space and time as Neptune and Julius Caesar. Though their views vary significantly, it is nomological connections that secure the content of intentional states and the relevant causal chains can extend indefinitely into distant times and places. Obviously, no traditional internalists can accept such accounts of content and construe the complex state of affairs that is, for example, one's believing that Caesar crossed the Rubicon as an internal state. No one now is directly acquainted, at least in Russell's sense, with the long dead Caesar, nor any state of affairs that includes him as a constituent.[11] Some externalists vehemently deny it, but it does seem to me that their ontology of belief states makes impossible introspective access to the state of affairs that constitutes a belief.[12] And some internalists will take that to be a *reductio* of externalism, and a reinforcement of their internalist intuitions that belief is an internal state.

In claiming that psychological internalism is inconsistent with introspective access to truths about our intentional states, I should emphasize that I am presupposing a particular version of psychological externalism.[13] Specifically, I am presupposing that the psychological externalist wants to argue that descriptions of intentional states are analytically equivalent to propositions describing complex nomological connections between external factors and internal states—alternatively, that the fact that is S's being in an intentional state is identical with the fact that S stands in complex nomological relations to external factors, external factors that are partially constitutive of the state. On such a view, it is misleading to suggest that an internal state *becomes* a representational state in virtue of its causal history. To be in an internal state is *never,* by itself, to be in an intentional state. Being in an intentional state is an extraordinarily complex property—the property of being in an internal state having a certain

[11] An anonymous reader points out that it might be controversial to claim that to be directly acquainted with Caesar's having crossed the Rubicon one would have to be directly acquainted with the constituents of that fact. The reader points out, for example, that one might be directly acquainted with a table without being acquainted with molecular constituents of the table. I am presupposing that facts are nothing over and above their constituents related and that one's acquaintance with a fact is completely parasitic upon one's acquaintance with the constituents of the fact. Though it is a long story, I would argue that the fact that there is a table before me is not in any way *constituted* by sub-atomic particles (though such entities might take the value of variables involved in the description of such facts).

[12] For a collection of papers defending a wide range of views on this controversy see Ludlow and Martin 1998.

[13] That I was doing so, and that my presuppositions are controversial, was made clear to me by conversations I had with some of the participants in the conference at which the papers in this collection were presented. I benefited particularly from discussions with Sarah Sawyer, Sandy Goldberg, David Henderson, and Jim Pryor.

causal history. There are at least some externalists who at least sometimes seem to talk as if the internal state itself is the intentional state but only when it has the right causal history. It is as if the internal state is 'magically' transformed into an intentional state replete with its representational capacities by virtue of its having the history it does. This sort of view doesn't even strike me as a version of externalism, properly speaking. If a genuinely internal state has an intrinsic character that makes it representational, then even if that character is in some sense determined by a *different* property of the state (its causal history), we still have a view according to which intentional states are internal states.[14]

The distinction between the two versions of externalism can be illustrated nicely with an example that Davidson uses in an attempt to reconcile externalism with introspective access to intentional states. Davidson points out that there is, of course, a difference between being sunburned and having a rash, and there is that difference even if the condition of the skin looks exactly the same. A condition of the skin counts as a sunburn only if it is caused in the right way—by exposure to the sun. Still, Davidson seems to suggest, one can obviously see the sunburn without seeing the sun that caused it. Or consider a quarter.[15] A given piece of metal has the property of being a quarter only if it originated in the US Mint, an institution whose identity criteria themselves involve a host of other social facts. But surely one can hold a quarter in one's hand, even if one can't hold the US Mint in one's hand. These examples, however, ultimately argue against the position Davidson had in mind. The property of being sunburned just is the property of having a skin condition of a certain sort caused by the sun. The fact that one is sunburned just is the fact that one has that skin condition caused by the sun. The proposition that one is sunburned is a proposition whose meaning involves reference to the sun and its effects on the skin. Similarly, the property of being a quarter is an extraordinarily complex relational property. The fact that I hold a quarter in my hand has that same complexity. The proposition that I hold

[14] Consider as an analogy Moore's view about the nonnatural property of being good. Moore admitted that goodness supervenes on natural properties, but insisted that that fact wouldn't make goodness a natural property. The fact that an intentional state supervenes (in Moore's sense) on external factors wouldn't make the state external. An anonymous reader for this volume suggests that perhaps the Davidsonian idea is better construed as the idea that the property of having an external cause of a certain sort is an essential property of a belief state, where the essential property is not even partially constitutive of that which has it. So Kripke argues, for example, that having the parents I do is one of my essential properties, even if having those parents is not, presumably, partially constitutive of me. I must confess that such views strike me as utterly mysterious, but in any event still strike me as versions of internalism. Just as I can be directly aware of myself without having any knowledge of my parents, so I can be directly aware of my belief states without having any knowledge of their causes. Certainly, it seems to me that classic externalist accounts of content, like those of Dretske, really are intended to provide analytically necessary and sufficient conditions for a belief's having a certain content. Since I have no idea how to evaluate claims about essential properties (other than such trivial properties as being self-identical, or being red or not-red), I'll continue to presuppose that the paradigm of an externalist view of content is one that builds external features into the very account of what an intentional state *is*.

[15] The example was given to me by David Henderson.

a quarter in my hand is equivalent in meaning to a very complex proposition describing, in part, social institutions. To know that someone is sunburned is to know a proposition describing the causal origins of a skin condition and to know that one holds a quarter in one's hand is to know a proposition describing complex social facts. Whatever access one has to the skin or the piece of metal is not access to the condition of being sunburned or to the state of having a quarter, and that's because on any plausible view, the fact of being sunburned just is a fact that includes as a constituent the sun; the fact that that's a quarter just is a fact that includes as a constitutent complex social institutions. In any event I'll continue to presuppose that the externalist is committed to the view that intentional states are in some important sense *de re* and include external factors as constitutents and there is an important sense in which *de re* beliefs, for example, include as constituents objects remote in time and place.

The above discussion presupposes that the *de re/de dicto* distinction is an ontological distinction concerning the character of belief states. Radical internalists, by contrast, can quite consistently hold that all intentional states are internal states. There is, to be sure, a distinction between *de re* and *de dicto ascriptions* of belief (and other intentional states), but while a *de re* ascription of belief often involves a commitment to the existence of an object external to mind, that by itself doesn't imply that belief states can contain an object external to mind. When I say of Henry Hudson that he believed of the bay in which he died that it was a passage to the Orient, I certainly seem to be committed to the existence of a bay in which he died. So the truth maker for my claim about Hudson's belief includes factors that lie outside of Hudson's internal states. But those who want to construe belief as an internal state have all sorts of devices to accommodate this claim. Some employ scope distinctions. S believes of the F that it is G, when there is one and only one F such that S believes that it is G. As many have pointed out, however, it is not clear what the content of S's belief is on such an analysis. The pronoun 'it' refers back to the F and if we want to be internalists about belief it is not clear that we have avoided the appearance of the F—the thing itself—creeping back into the belief state. A more plausible, but more complex approach involves treating the *de re* ascription of belief as one that involves an existential claim whose variables range over propositions. When we say of S that he believes of the F that it is G, we assert that one and only one thing is F, and we further claim that there is *some* proposition whose subject concept picks out the F such that S believes that proposition. At least that's the rough idea. The devil is in the details and the view requires modification to deal with counterexamples.[16]

It is important to realize that *de re* ascriptions of belief pervade language. There is no linguistic cue that tells us whether or not we should interpret an ascription of belief as being *de re*. Sometimes context, coupled with a principle of

[16] I sketch such a view in Fumerton 1986.

charity, makes it clear. Our description of Hudson's belief, for example, doesn't make much sense unless it is interpreted as a *de re* ascription. But there are other contexts in which we might need to ask in order to determine whether or not the speaker really meant to identify the object of the belief described as the proposition picked out by the noun clause completing the intentional verb.

So if one can argue that there are no *de re* intentional states at all, or that there are no *de re* intentional states that have as constituents objects with which one is not acquainted, one may be well positioned to advance an internalist thesis concerning all intentional states. But suppose that one is convinced by thought experiments involving twin earths, swamp men, or arthritis that the externalists about content are on to something and that for something going on in me to become a representation it must causally interact in an appropriate way with the thing represented. Put more accurately, and more carefully, suppose that one becomes convinced that internal states are never, qua internal states, representations, but that the state of affairs that is S's representing X is always a complex state of affairs that is some internal state of S standing in complex nomological relations to X (or things of the same kind as X, or kinds of things out of which X is in some sense constituted). At that point one must abandon an internalist account of intentional or representational states. But the somewhat loaded description of this externalist position obviously still allows that internal states are critical *constituents* of representational states. And how could it be otherwise? If one offers a causal account of representation, one needs to find the right sort of relata to constitute the first and last links of the relevant causal chains. And surely no one in his right mind thinks that in thought a person can end up representing some object or state of affairs if nothing happens in that person's brain (or mind). To be sure, the externalist shouldn't allow that the occurrence of the critical internal state is identical with the occurrence of an intentional state. But it or some other internal state must occur if there is to be representation.

Surely all this, however, is small comfort to the internalist trying to find a target for an internalist analysis from within the framework of an externalist account of intentionality. The internalist's thesis is in danger of reducing to the claim that one should adopt internalism with respect to those states of a conscious being that are internal states! And *that's* not a very interesting thesis. It becomes more interesting, however, if the internalist can successfully argue that at least some internal states are psychological or mental states. But what would be the difference between internal states that are psychological or mental and the indefinitely many internal states (like heart valves opening and closing) that are not? Well, there are two famous criteria proposed as a mark of the mental. The first, of course, is intentionality. But we are now presupposing the (what I think is, in fact, false) thesis that the externalist has a correct account of intentional states and that at least some representational states include external constitutents (though

not, of course, necessarily the very object represented).[17] There is, however, another historically prominent and equally plausible proposal concerning criteria for identifying the mental, and that is introspective accessibility. Indeed, some philosophers argue that it is an *essential* feature of a mental state that we be conscious of that state, or at least have the capacity to become conscious of that state. While I'm not sure *that* thesis is plausible, it is surely tempting to suppose that it is a *sufficient* condition for a state's being mental that it is possible to become introspectively conscious of the state.

There are a host of questions that must be asked, however, before the proposal is even clear enough to evaluate. Accessibility is a modal notion and philosophers must always take care to identify the relevant modality. We also need an account of introspective accessibility. And here the deep connections between epistemic and psychological internalism again become evident. Contemporary externalists are probably inclined to analyze introspection in terms of belief about a psychological state that is caused without the mediation of cognitive states. So a reliabilist, for example, might understand one's introspective knowledge of pain as belief that one is in pain produced by a highly reliable belief-independent process where the input is the pain itself. Such accounts of introspection are, to my way of thinking, hopelessly implausible. They fail to take account of the way in which the pain itself—not just a belief about the pain—is directly and immediately before consciousness. Suppose, for example, that one becomes convinced that there can be pains of which one is not conscious and that one is induced by a hypnotist to believe that one is in such pain. That belief will not constitute being conscious of one's pain. Having introspective knowledge of one's pain involves standing in the relation of direct acquaintance to the pain itself, a relation that cannot be assimilated to any intentional state.

Can we rely on this concept of introspective access to identify a plausible target of psychological internalism? At this point most internalists will rely on a kind of argument familiar from centuries of philosophical controversy over the nature of perception. Those rejecting direct realist accounts of our perceptual access to the physical world try to introduce a common denominator of both veridical and nonveridical experience. Whether we veridically see a table or hallucinate a table, there is surely something, the argument goes, that we are, or can become, aware of—something that is common to the two states of affairs. In precisely the same way, the internalist can ask us to compare two hypothetical situations—one in

[17] Internalists and externalists alike must acknowledge that one can think of (fear, hope for, search for, etc.) things that don't exist. So externalists must be very careful in identifying the way in which external factors determine content. At most, the careful externalist would insist that for S to represent X, S must have come into contact directly or indirectly with the kind of thing represented. And even that thesis is bound to be too strong. The careful externalist will surely borrow a leaf from the page of the radical empiricist who distinguished between simple ideas and complex ideas. Only simple ideas, that empiricist claimed, are copies of prior impressions. The externalist should similarly restrict claims about the need for connection to external reality to simple ideas of external objects.

which the externalist's conditions for being in an intentional state are met, and one in which one scrapes away all, but only, the external components. That there would be internal components left seems almost trivially true. That some of these internal components would be mental is more controversial, even on our second proposal for identifying the mental. But what reason could anyone have for denying that there are internal states that fail to satisfy the externalist's requirements for intentionality, but that can still become the phenomenal objects of introspection? The complex causal chains that create representation according to the externalist do, after all, have internal links. And there seems to me no good reason to deny that we can become aware of those internal links. Consider again perception. One can embrace all one wants Moore's (1903) and Harman's (1990) suggestion that perceptual states are in a sense 'diaphanous'—that our attention is always, or nearly always, directed outward in perception. But Moore himself would have been the first to allow that there can be nonveridical counterparts of veridical perception, and that these nonveridical counterparts have a character of which we can become aware, and even which we can describe (though we may have to 'borrow' a language that is created primarily for the description of external reality). Brains in a vat, swamp men, solitary linguists, and the like may not be capable of representing what we are capable of representing, but their very description makes sense only on the supposition that there is something common to them and us, and there is no incoherence whatsoever in the supposition that they are, or can become, conscious of the character of this state. If the externalist can't give sense to introspective access to the character of such states, then so much the worse for externalism.

All of this might be just fine with the externalist. I suspect that if we confine our internalist theses to states that fall short of robust intentionality the externalist might be willing to hand the internalist a few philosophical crumbs. Dialectically, however, concessions of this sort are dangerous for the externalist and promising for the internalist. Consider an example that is, to be sure, favorable to the internalist. In their zeal to construe all interesting mental states as intentional, some externalists will construe even pain as a representational state. Being in pain is just being in a state that represents something like damage to the body. Our conciliatory externalist grants us that there may be an internal state of which we can become directly aware but insists that it is only *pain* insofar as it represents damage to the body. Such representation, the externalist insists, is a complex matter that includes facts about the causal history of this, or this kind of, state. We internalists know that we just mean by pain a state of this phenomenal kind—the kind with which we are directly acquainted. If the externalist won't give us the word 'pain' to describe the state, we can just asterisk the word 'pain' and claim 'pain*' for our own. We won't let the externalist preclude us from inventing a language to talk about those states the existence of which even the externalist concedes. Of course, we'll be quite confident that the pain we have always been talking about is, in fact, pain*, but there is no point in quarreling over the use

of a word. In similar fashion, I suspect that we can reclaim *all* that was near and dear to our internalist hearts and leave to the philosophical externalists and cognitive psychologists the more complex, and certainly interesting, nomological states of affairs that they want to study empirically.

Consider again other obvious candidates for representational states. Even if we accept in broad outline the details of an externalist account of representation, there still remains the question of just what we are introspectively aware of in nonveridical representation—in perceptual hallucination or fear of ghosts, for example. The representationalists are notoriously difficult to read on this subject. Their usual suggestion is always that introspection has precisely the same direction (outward) as the intentional state introspected. Such a suggestion faces enormous difficulties with the obvious need to account for our ability to introspect the difference between different intentional states (fear and desire, for example) with precisely the same content. But setting that problem aside, there is no existent *object* in hallucination that can serve as the shared 'object' represented by both experience and introspection. The contemporary representationalists' naturalism will surely make them uncomfortable with a world of Meinongian non-existent objects to secure content for misrepresentation, and in their search for something to identify as the 'objects' represented, they may turn to something like unexemplified properties.[18] You never know what a naturalist takes to be consistent with naturalism—it usually depends primarily on whether or not they think they need the category of thing under discussion. But if unexemplified properties play a critical role in both their account of nonveridical representational states and introspection of those states, we have found just the sort of state of affairs about which we can advance an internalist thesis. And we have also found a pretty good candidate for the intentional* state that we always thought we were talking about when we used the language of our folk psychology.

REFERENCES

Burge, T. 1979: 'Individualism and the Mental.' *Midwest Studies in Philosophy* 4: 73–121.

Conee, E. and Feldman, R. 2004: *Evidentialism* (Oxford: Oxford University Press).

Davidson, D. 1987: 'Knowing one's Own Mind.' *Proceedings and Addresses of the American Philosophical Association* 60: 441–58.

Foley, R. 1987: *The Theory of Epistemic Rationality* (Cambridge, Mass.: Harvard University Press).

Fumerton, R. 1986: 'Essential Properties and De Re Necessity.' *Midwest Studies in Philosophy* II: 281–94.

——1996: *Metaepistemology and Skepticism* (Savage, Md.: Rowman and Littlefield).

[18] That it is awareness of properties involved in nonveridical perception and the introspection of such perception is the gist of Mark Johnston's suggestion 2002 that in hallucination we are aware of uninstantiated 'profiles' of universals.

Fumerton, R. 2001: 'Classical Foundationalism.' In M.DePaul (ed.), *Resurrecting Old-Fashioned Foundationalism*, (Savage, Md.: Rowman and Littlefield).
Harman, G. 1990: 'The Intrinsic Quality of Experience.' *Philosophical Perspectives* 4: 31–52.
Johnston, M. 2002: 'The Obscure Object of Hallucination.' *Philosophical Studies* 120: 113–83.
Keynes, J. 1921: *A Treatise on Probability* (London: Macmillan).
Lehrer, K. 1971: 'How Reasons Give Us Knowledge, Or the Case of the Gypsy Lawyer'. *Journal of Philosophy* 68. 10: 311–13.
Ludlow, P. and Martin, N. 1998: *Externalism and Self-Knowledge* (Stanford, Calif.: CSLI Publications).
Luper-Foy, S. 1985: 'The Reliabilist Theory of Rational Belief.' *The Monist* 68: 203–25.
Moore, G. E. 1903: 'The Refutation of Idealism.' *Mind* Oct.: 433–53.
Nozick, R. 1981: *Philosophical Explanations* (Cambridge, Mass.: Harvard University Press).
Putnam, H. 1975: 'The Meaning of "Meaning".' In *Mind, Language and Reality* (Cambridge: Cambridge University Press).
____ 1996: 'Introduction' to A. Pessin and S. Goldberg (eds.), *The Twin Earth Chronicles* (Armond, NY: M. E. Sharpe).
Russell, B. 1912: *The Problems of Philosophy* (Oxford: Oxford University Press).
Williamson, T. 2000: *Knowledge and its Limits* (Oxford: Oxford University Press).

3

Externally Enhanced Internalism

Earl Conee

1. INTRODUCTION

Refinements aside, epistemic internalism holds that justification is wholly determined inside of us. Refinements aside, content externalism holds that the contents of some mental states are partially determined outside of us. Content externalism can seem to threaten epistemic internalism. The justification of attitudes toward propositional contents is a main topic of internalism. According to content externalism, some attitude contents are differentiated externally. Those contents might appear to be too remote from internal factors for justification to be determined internally. In short, since content externalism seems to imply that we cannot always distinguish on a purely internal basis which proposition we have in mind, it can seem doubtful that the appropriate attitude toward a proposition is always determined internally.

In spite of this appearance, the fact is that content externalism aids epistemic internalism. At least this is true if internalism is what Richard Feldman and I have called 'mentalism.' The reason is simple. Mentalism is the thesis that for epistemic purposes the 'internal' is the mental. So according to mentalism, what is epistemically 'inside of us' is determined in whatever way mental content is determined. The basic internalist claim is that the justificatory status of any proposition for a person supervenes on what is internal to the person. Since content externalism expands the factors that fix the mental, content externalism expands the supervenience base for justification according to mentalism. Thus, content externalism makes more available for the internal determination of justification. That makes life easier for the epistemic internalist.

The claim of assistance is abstract and controversial. It is abstract in that it does not specify anything helpful that content externalism adds to the mentalist supervenience base for justification. The claim is controversial in that mentalism

I have benefited from comments by Rich Feldman and Sandy Goldberg on previous drafts.

is not universally regarded as the best version of epistemic internalism. Some illustration of the assistance from content externalism is in order, as is a defense of internalism as mentalism. These things will be attempted just below. Subsequently, three other prima facie conflicts between epistemic internalism and content externalism will be addressed and reconciled.

2. SOME EXTERNAL ASSISTANCE

Richard Feldman and I have discussed the following sort of seeming threat to epistemic internalism.[1] Given content externalism, an Earthling and a Twin Earthling who is apparently the Earthling's internal duplicate could be justified in believing different propositions. Supposing that the Earthling and her Twin are ordinarily well informed, the Earthling is justified in believing that water exists while the Twin is justified in believing that twater exists. But not vice versa in either case, since neither has any information about the aqueous substance in her counterpart's environment. So apparently we have the implication that justification does not supervene on internal states.

Rich and I have offered a modification of internalism to accommodate this apparent possibility. We defined a notion of a 'counterpart' proposition. A counterpart to some proposition that is being considered is a proposition that differs from the considered one in an environmentally determined way, if at all. Using this counterpart idea, we held that internalism can be understood as the thesis that the justificatory status of counterpart propositions cannot differ between internal duplicates. Thus understood, internalism implies only that the water proposition must be as well justified for the Earthling as some counterpart of that proposition is for the internal duplicate Twin Earthling. Twins are no threat to this thesis.

Given a modest further assumption, though, no modification of internalism is necessary. The further assumption builds on such highly credible observations as that no one can be justified in believing that water exists while lacking the concept of water. More generally, a proposition has a justificatory status for someone only if the proposition is graspable by the person. Concepts help us to grasp propositions. To bear any epistemic relation to a proposition, we need concepts than connect us psychologically to it. Let's call this relation 'having concepts enabling one to engage the proposition.' The crucial further assumption is that a proposition has some justificatory status for someone only if she has concepts enabling her to engage the proposition. Let's call this assumption 'the conceptual requirement.'

The conceptual requirement is quite plausible. To cite one humble instance of its explanatory assets, it accounts for the fact that we become justified in believing

[1] The discussion occurs in the afterword to Conee and Feldman 2004.

new propositions from a technical discipline only when we acquire the relevant technical concepts.

The conceptual requirement helps with the twins example. It supports the claim that the Earthling must have the concept of water, while her twin must have the concept of twater, in order for them to be justified in believing the respective propositions about the existence of the respective substances.

Content externalism readily accommodates the conceptual requirement. It is available to a content externalist to hold that appropriate differences in external conditions determine the difference in the concepts that the twins possess. This environmental determination of conceptual content is just as intuitive and defensible as the standard content externalist view that the twins' propositional contents are differentiated environmentally.

Relying on the conceptual requirement, the mentalist version of internalism needs no modification to accommodate the Twin Earth examples.[2] Mentalism takes 'internal' duplicates to be mental duplicates. Mental duplicates must be justificatory duplicates. Given the conceptual requirement, the twins are not mentally alike. They have differing concepts that enable them to grapple with the differing contents of their beliefs. So mentalism does not imply that the same propositions are justified for each of them. This is one way in which content externalism aids internalism, when internalism is understood as mentalism.

The conceptual requirement is of equal service to mentalism on the topic of empty names. We are justified in believing that Socrates was a philosopher. The mentalist says that any who are mentally the same are justificationally the same. What does that claim imply about worlds in which we are as much as possible mentally the same as we actually are, but for the fact that our use of 'Socrates' fails to refer because Socrates never existed?

In considering this question, it is harmless to employ the most drastic content externalist view of empty names. This is the view that, if a name has no referent, then no proposition is expressed by any sentence using the name.[3] Thus, in the possible worlds in question, our maximal duplicates' use of the sentence 'Socrates was a philosopher' formulates no proposition. Does mentalism imply that the proposition that Socrates was a philosopher is justified for our maximal duplicates, given that it is justified for us?

[2] Assuming that in the twins example the differing aqueous propositions are respectively believed, the twins differ in their beliefs. This is an uncontroversial mental difference. So it can be inferred, without using the conceptual requirement, that the twins are not mental duplicates. But in some possible cases the propositions are not believed, although one of them is justified for the one twin while the other is justified for the other. Something is needed to entail a mental difference across the board, including these latter cases. The conceptual requirement does the job.

[3] The points to be made about the no-proposition view will clearly apply to other externalist metaphysical views about contents, such as the view that where a name does not refer, the sentence expresses a proposition that has a gap in it where the name's referent would be. (See e.g. Braun 1993.) Such views imply a difference in content rather than an absence of content. Either way, there is a mental difference.

Internalism is about both doxastic and propositional justification. Doxastic justification is a matter of actually believing a proposition for which one has justification and basing the belief on some undefeated justification for it. For the question about doxastic justification, the conceptual requirement need not be invoked. It is not controversial that our maximal duplicates have to share our beliefs, in order to be mentally the same as we are. Thus, in any world where our maximal duplicates' use of 'Socrates was a philosopher' does not formulate any belief, they are not mentally the same as we are. And this lack of a belief is what our content externalism implies about worlds in which the name 'Socrates' does not refer. So we have no mental duplicates in such worlds. Consequently, mentalism does not imply anything about their inhabitants sharing our doxastically justified beliefs.

Now let's take up propositional justification, i.e. the sort of epistemic justification that a proposition can have for someone, whether or not the person actually has any doxastic attitude toward it. In worlds where 'Socrates' does not refer because there was no Socrates, it seems that no attitude toward that proposition is justified for our maximal duplicates. The content externalist view that we are employing tells us that the proposition does not even exist in those worlds, and there is nothing that our maximal duplicates there could do about that. Assuming all this, how could any attitude toward that proposition be called for on epistemic grounds?

Mentalists need not say that any attitude toward the proposition is epistemically justified under such circumstances. It is open to the internalist, and plausible, to hold that without Socrates in a world, no one could there have the concept of Socrates by which we understand the name 'Socrates.' The concept is partly externally determined, just as is the proposition.

It is reasonable to employ the conceptual requirement here. We can claim that some concept of Socrates that derives in part from Socrates himself is needed, in order for the proposition that Socrates was a philosopher to have any justificatory status. No one has any such concept where Socrates does not exist. Concepts are mental items. So in a world where Socrates does not exist, our maximal duplicates are not conceptually equipped for that proposition to have any justificatory status for them. Thus, the mentalist can exploit external aspects of the mental content and reasonably contend that this Socratic proposition is not propositionally justified for us in any world in which Socrates does not exist.[4]

[4] Clearly the same point can be made about a need for a different concept if the word had had a *different* referent rather than none, owing to a difference at the origin of the reference-determining causal chain. This need for a concept of the subject, in order to have any justified attitude toward a proposition formulated with a proper name, might be thought to conflict with the view that proper names lack all 'descriptive' or 'qualitative' content. It does not conflict with that view, given a suitably broad understanding of concepts. The idea implemented here is that only someone who understands some name for the individual can grasp any proposition that can be formulated with a proper name designating the individual. The requirement of understanding is being expressed as

3. AN ADVANTAGE OF MENTALISM

These helpful contributions of content externalism to epistemic internalism rely on taking internalism to be mentalism. A main thesis of mentalism is this strong supervenience claim:

S. The justificatory status of a person's doxastic attitudes strongly supervenes[5] on the person's mental states, events, and conditions.[6]

The mentalist account of epistemic internalism is controversial. It is time to say something on its behalf. I shall argue that it compares favorably to rival interpretations of internalism. The principal advantage of mentalism is that it more clearly includes some intuitively internalist positions.

The main rival version of internalism identifies what is internal for someone with what is available to be related to the person by some sort of access. The access is frequently understood as an outcome of a mental process, such as introspection. For instance, here is Jim Pryor's basic account of internalism, which he calls 'Simple Internalism':

SI. Whether one is justified in believing P supervenes on facts which one is in a position to know by reflection alone (2001: 104).

Pryor explains that by the term 'reflection' he means 'a priori reasoning, introspective awareness of one's own mental states, and one's memory of knowledge acquired in these ways' (ibid.).

S and SI are not far apart. Both accounts classify as internalists a wide range of those philosophers who would count themselves as internalists. Both accounts exclude brazenly externalist views such as reliabilism and proper functionalism.

Nonetheless, mentalism has advantages. For one thing, it is nicely noncommittal. In particular, mentalism leaves open what states, events, processes, or mechanisms can yield any sort of positive epistemic status, and it implies nothing about how much positive epistemic status is available. In contrast, SI has substantial psychological and epistemic commitments. SI implies that if there are any

the need to have a concept of the named individual. This notion of a concept implies nothing about having any descriptive or qualitative content.

[5] It is worth noting that strong supervenience is important to the intuitive merits of the classification. William Alston formulates a version of internalism that holds only that 'justifiers' must be '(fairly readily) accessible' (2001: 101). Subsequent discussion by Alston makes it clear that this thesis is to be understood as a weak supervenience thesis. Alston argues that a reliabilist could agree with the thesis on the ground that the thesis allows that the status of something as a justifier could depend on whether this something happened to be part of a reliable belief-forming process (102). Given that the thesis fails to exclude this reliabilist view, it seems too weak to be a good version of internalism. In contrast, since S requires that justification strongly supervenes on any justifying condition, the condition cannot justify contingently as the reliabilist would have it.

[6] Conee and Feldman 2004: 56. Thesis S is not the whole of mentalism, because it is silent about propositional justification. That omission will not affect the issues discussed below.

justified beliefs, then a priori reasoning and introspection are real psychological processes, and they can give knowledge to any who have justified beliefs.

These are not bold assumptions. But they seem inessential to the internalist perspective. The fundamental internalist idea is just that justification is internal to us. It seems extra to imply that our innards have any particular psychology. It also seems extra to imply that knowledge of our innards is definitely available to us. It seems that an internalist could coherently deny these things.

For example, a philosopher might regard a priori reasoning and introspection as discredited categories of folk psychology. She might nonetheless affirm, for instance, that one's epistemic justification strongly supervenes on one's sensory states. A 'Sensationalist' of this sort ought to qualify as an internalist.

SI does not out-and-out imply otherwise. It may be that whatever supervenes on our sensory states also supervenes on things that we are in a position to know by reflection alone. If so, then SI agrees that our Sensationalist is an internalist. But SI still displays a relative weakness here. For one thing, the Sensationalist could conceivably be right about a priori reasoning and introspection. It is conceivable that mature science will reveal that there are no such things. If they are indeed unavailable to us, then SI unfortunately implies that nothing differs in justificatory status for us from anything else. And second, even if there are such processes, the Sensationalist could not coherently count herself as an internalist by the test of SI, since she denies that those processes exist while SI explicitly requires them to yield knowledge of a supervenience base for justified beliefs. It is a liability of SI that such an intuitively clear internalist as the Sensationalist could not use it to recognize her own internalism.

A second disadvantage of SI concerns its epistemic commitments. An internalist might accept the existence of the psychological kinds that are definitive of 'reflection,' but skeptically deny that they can yield knowledge. An internalist who is skeptical about the fruits of reflection could coherently affirm that the conditions that furnish our justification for any proposition are determined by facts that can be more or less justifiably believed by reflection. The position of the reflection-skeptic just implies that such facts do not thereby become known.

Again, the problem for SI is not that it definitely misclassifies someone who is an internalist. SI does count our reflection-skeptic as an internalist, unless the skeptic turns out to be correct about the incapacity of reflection to yield knowledge. But again, for one infelicitous thing, the skeptic could conceivably be correct about the lack of knowledge, and it is additionally unfortunate that the skeptic cannot coherently use SI to regard herself as an internalist.

In contrast, mentalism has no problematic psychological or epistemic commitments. Mentalism requires only that justification is determined by the mental, leaving open what mental processes exist and what knowledge they can supply.

Might a mere commitment to the mental be excess baggage for an account of internalism? In other words, could there be a coherent version of internalism that denies the existence of the mental? That is doubtful. The intelligibility of

any such internalism depends on its explaining how a mindless being might have epistemic justification, and also explaining what would be 'internal' about the justification. Even if this could be done, the prospects for finding a seriously credible position of this sort are dim. The ontological commitments of mentalism are correspondingly light.

Other versions of the accessibility approach to internalism add epistemic implications beyond those of SI. These other versions have internalism imply that some epistemic facts are always somehow 'accessible' to us, e.g. knowable by reflection. The accessible epistemic facts might include the justificatory status of our beliefs, or the identity of the justifying evidence for any justified belief, or general epistemic principles about when there is justification or knowledge.

There are respectable internalist positions that affirm such things. Perhaps some such position is correct and some of those epistemic facts are indeed always available in the required way. But claiming that we have these epistemic capacities is not mandatory just in order to be an internalist. Our Sensationalist is an internalist, no matter what she says about the availability of epistemic facts. She is an internalist simply because she holds that justification derives entirely from sensory states, and sensory states are as internal as it gets. To deny that this is enough to qualify as an internalist and to insist that only those who impose epistemic access conditions are true internalists is analogous to holding that only thoroughgoing Cartesians are true foundationalists. Defining a category by the views of its extremists is a mistake.

4. THE SUBSTANCE OF INTERNALISM RETAINED

Does mentalism, abetted by content externalism, gain defensibility and inclusiveness at the cost of rendering internalism unmotivated?[7] More specifically, are there distinctive reasons for thinking that justification is internal that are obliterated by counting any mentalist theory as internalist?

It should be acknowledged that mentalism is too broad to imply on its own the premisses of the classic internalist arguments. For instance, mentalism does not imply that epistemic justification depends on conditions that provide for blameless believing, nor does mentalism imply that justification depends on the evidence that one currently possesses. Seemingly worse, one central and powerful intuitive point favoring internalism is that victims of pervasively deceptive demons might have the same justification for their beliefs as is had by their closest counterparts among people in ordinary environments. The present version of mentalism allows in the supervenience base for justification mental states that are individuated by content externalist conditions. Given many content externalist

[7] I am grateful to John Greco for raising this sort of question at the 2006 Pacific APA meetings in the discussion of a previous draft of this paper.

views, victims of many drastic deceptions do not have the same mental states as their closest counterparts among ordinary people.[8] So internalism as mentalism does not imply that such victims are internally the same as their closest ordinary counterparts. In consequence, it cannot be validly inferred from internalism that such deceptions result in the same justifications for the victims as for the ordinary counterparts. This might be thought to show that something important to internalism is lost when it is taken as mentalism.

Regarding internalism as mentalism does not cost any internalist any argument. At most, the source of premisses must be more carefully specified. For instance, given the ways in which external factors can affect mental content, an internalist who is a responsibility theorist could not plausibly say that all mental conditions provide for blameless believing. A remote causal factor might affect the content of a belief in a way to which the person might not be positioned responsibly to respond. But in order to be an internalist the responsibility theorist does not have to say that justification depends on internal conditions in general. The internalist responsibility theorist can say, with no loss of credibility, that justification depends on those more specific mental conditions that she takes to be the basis for blameless believing. The mentalists' candidate for the property of being internal plays no central role in the articulation or defense of the responsibility theory. But that does not detract from the view or give any reason to deny that internalism is mentalism.

Likewise, various versions of mentalism are in a position to benefit from the powerful intuition that victims of drastic deceptions have the same justification as do their ordinary counterparts. This is compatible with mentalists acknowledging differences in some of the most extreme deceptions between some cognitive contents of the victims and those of their ordinary counterparts, such as in cases where external differences give them different concepts. The claim of shared justification is borne out by a variety of more specific internalist bases for justification. For instance, if the fundamental source of justification is experiential evidence that is shared by the counterpart pairs, then they have the same justification. Similarly, if the fundamental source of justification is a basis for responsible belief that the counterparts share, then they have the same justification.

In general, the reasons that internalists can give for their views are unaffected by taking internalism to be mentalism. All that is demoted is the role in arguments of the property of being internal. This change is a benign consequence of interpreting internalism inclusively enough to accommodate all positions about justification that it makes sense to classify as internalist.

[8] To separate issues, we should restrict our attention to deceptions that produce mental differences, according to standard versions of content externalism, but allow the contents of the justified attitudes to be the same. Recently envatted brains, for instance, can consider the same propositional contents as people in ordinary environments, even if they differ mentally in virtue of differences, say, in what is actually demonstrated in attempts at demonstratively based believing.

Mentalism includes a diversity of theories, from views that demand for justification direct acquaintance with a truth-making fact to views according to which all beliefs are justified unless their contents obviously conflict. In consequence, mentalism is less incisive than some access accounts of internalism. Internalism as mentalism is a general theoretical perspective, like empiricism, rather than a war cry, like logical positivism.

Mentalism is inclusive, but it is not noncommittal. Its strong supervenience thesis does exclude contingent or purely environmental variations from affecting justification. In particular, reliability requirements remain beyond the pale, as do claims of variations in justification that derive from purely communal standards. Mentalism remains a substantial thesis that is tied to the etymological roots of internalism. Justification does have to be 'internal' to something, namely, the mind.

5. PARTIAL UNDERSTANDING AND JUSTIFICATION

Let's turn now to other issues where content externalism and epistemic internalism interact, to the apparent detriment of internalism. One apparent threat arises from cases of doxastic attitudes based on partial understanding. Content externalists hold that some aspects of the contents of some such attitudes are socially determined. Suppose, for instance, that Smith partially understands the word 'vegan.' Smith intends to use it in the ordinary way. He realizes that the word applies to people on the basis of some special policy that they have adopted pertaining to vegetables. But Smith does not have any evidence or opinion about what particular policy makes one a vegan. Suppose further that Smith knows that Jones has a strict policy of eating vegetables and not eating any animal products. Still, Smith's uncertainty about what policy makes one a vegan leads Smith to withhold judgement on the thought that he considers by use of the sentence, 'Jones is a vegan.' Smith realizes that Jones's dietary intentions are just one of many policies about vegetables that people might have. From Smith's point of view, the chances are not high that Jones's is the policy that makes one a vegan.

Epistemic internalists seem required to count this withholding of judgement by Smith as justified. After all, Smith's internal condition seems to give Smith no good reason to apply to Jones the term 'vegan,' given the limited extent to which Smith understands it.

According to various versions of content externalism, though, the proposition on which Smith is withholding judgement is the very one that the words 'Jones is a vegan' formulate in ordinary English. Yet Smith does know that Jones has the dietary policy that is as a matter of fact definitive of being a vegan. In some views, the proposition that Jones has that definitive policy is identical to the proposition that Jones is a vegan. Since these views seem to imply the most severe challenge to the epistemic internalism, let us assume that they are correct. Given such a

view, it appears that Smith's justified attitude toward the vegan proposition is belief rather than withholding judgement. Smith out-and-out knows that Jones has precisely the dietary policy that is definitive of veganism.[9]

Yet as we have seen, epistemic internalism appears to imply that withholding is Smith's justified attitude. Again, as far as matters internal to Smith seem to be, he has no good reason to believe it as he considers the proposition using the word 'vegan.' So apparently in some cases of partial understanding something beyond internal conditions can justify belief.

In spite of this appearance, internal conditions are sufficient for us to account for justificatory status in such examples. We should note that there are changes in what is intuitively justified for Smith as his thinking proceeds. Smith is considering a proposition by use of English sentences. His partial understanding of the word 'vegan' gives him too little information about the proposition that any sentence with that word expresses to give him any good reason to believe that that it stands or falls with the proposition that his 'strict diet' sentence expresses. When he thinks about the proposition that Jones is a vegan using the 'vegan' sentence, Smith has undefeated reasons to withhold judgement on it. From Smith's perspective, the chances are not high that Jones's policy concerning vegetables is a vegan-making policy. So on those occasions intuitively Smith is justified in withholding judgement. When Smith thinks about the proposition that Jones is a vegan using the 'strict diet' sentence, he has undefeated reasons to affirm it—he knows that Jones has that dietary policy. So intuitively Smith is justified in affirming the proposition on those occasions. This change in what attitude is justified for Smith is quite understandable. He lacks information about the concept of a vegan that would enable him to recognize it as the concept that he deploys when he uses the 'strict diet' expression. Internal differences underlie these justificatory changes. It is congenial to internalism to attribute this pattern of justified attitudes.

Objection: This internalist account attributes to Smith reasons to believe the proposition while he is allegedly justified in withholding judgement. All along Smith has whatever supporting reasons give him knowledge of Jones's dietary policy. So it must be that the appearance that withholding is justified is an illusion. Smith has the knowledge-giving justifying reasons all along.

Reply: The clear facts are facts about the attitudes that are justified for Smith as he brings the proposition to mind by use of one or another sentence. Until Smith better understands the word 'vegan,' the attitude that he is justified in having toward the proposition depends on the sentence he is using to think of the proposition. The reasons that give him his stored knowledge that Jones has the

[9] Even in views according to which the propositions differ, they are exquisitely similar. In any case Smith's justification for the dietary policy proposition seems to justify the vegan proposition, as is attested by the fact that one who has full understanding of the vegan proposition would be justified in believing it by whatever gives him knowledge of Jones's dietary policy. So it seems that belief, rather than withholding judgement, is the justified attitude for Smith to take toward the proposition that Jones is a vegan.

dietary policy do not affect the justificatory status of Smith's thought when he considers a proposition using the 'vegan' sentence. Smith's partial understanding of the word 'vegan' gives him crucially limited information about the concept that he deploys when using the word. He has at that time no reason to connect the proposition that he is considering with his evidence about Jones's dietary policy. So under such circumstances he does not have as evidence for the thought the reasons that give him stored knowledge of the proposition. Withholding judgement is thus justified when Smith uses the 'vegan' sentence. This allows that affirming is justified, when Smith uses the 'strict diet' sentence.

Epistemic internalists have more than one credible alternative about the justification of Smith's attitudes toward the proposition. Another internalist approach to this example is to concur with the objector that belief is the attitude that is justified for Smith. This second approach claims that Smith does have throughout good reasons to believe the proposition, in spite of his limited appreciation of the bearing of those reasons. The alternative view implies that those reasons do affect the justificatory status of the proposition for him, whatever sentence he uses to think of it.

This alternative has the initial liability that it fails to account for the reasonableness of Smith's withholding judgement when he uses the 'vegan' sentence to consider the proposition that Jones is a vegan. But the alternative may be defensible, perhaps by accounting for the apparent reasonableness of Smith's withholding judgement as Smith's failing to appreciate the actual bearing of some evidence that he nonetheless does have. In any event, both of these approaches are available to an internalist.

Objection: Content externalism seems to be making trouble here rather than helping epistemic internalism. Dropping the content externalist view, we could say that Smith is unable even to think that Jones is a vegan when Smith uses the 'vegan' sentence. We could say that he does not understand the sentence well enough to use it to consider the proposition that is expressed by it in English. If internalists say this, then they do not have to worry about what attitude Smith is justified in having toward the proposition. Belief is unproblematically the justified attitude, since he knows that Jones has the definitive diet.

Reply: Content externalism is defended by intuitive judgements about the contents of some thoughts in some hypothetical situations. The intuitive data support the conclusion that people sometimes do adopt propositional attitudes by use of partially understood sentences. Epistemic internalists must have some response to these data. In particular, they must have something to say about the epistemic status of propositions that are apparently considered using partially understood sentences. Content externalism helps by making available mental equipment for the epistemic internalist to use in order to explain this epistemic status. The favored internalist proposal has it that withholding judgement is sometimes Smith's justified attitude toward the proposition that Jones is a vegan. In order for this attitude to be justified, Smith has to be able to grapple

intellectually with that proposition while withholding judgement is justified. Content externalism provides a crucial intellectual device to attribute to Smith in order for him to do this. It allows us to attribute to Smith the concept of a vegan. We can thereby account for Smith's being in a position to think about the proposition. Smith's partial understanding of the word 'vegan' gives him limited grasp of the concept using the expression, and this helps to explain why withholding judgement is his justified attitude when he uses that word to bring the concept to mind.

6. JUSTIFICATION FROM KNOWLEDGE?

Points defended in Timothy Williamson's book *Knowledge and its Limits* (2000) pose two apparent challenges to mentalism that are bolstered by content externalism in one way or another.

The first challenge arises from a main thesis of the early chapters of Williamson's book. The thesis is that knowledge is a mental state (2000: 21 and ff.). Let's formulate this as follows.

KM. Any state that consists in someone knowing a fact is a mental state of the knower.

The principal way that content externalism supports KM is to block an otherwise plausible objection to KM. The objection consists in the following argument.

Some known facts have external constituents, such as the fact that Braintree is in Massachusetts, which has Braintree as a constituent. A known fact is a constituent of any state of knowing it, and the constitution relation is transitive. So something external is a constituent of some states of factual knowledge. Yet no mental state has any external constituents. Hence, at least some states of factual knowledge are not mental states.

Content externalism facilitates replying to this argument as follows. Content externalism standardly attributes to us beliefs with external constituents. All beliefs are mental states. So the premiss in this argument claiming that no mental state has external constituents is objectionable, if standard content externalism is correct.

Let's take KM for granted and see how it affects mentalism. Here is an apparent threat.

Assuming KM and mentalism, states consisting in a person knowing a fact are eligible to be justifiers for the person. Yet consider a case in which Smith knows, in some ordinary way, that Nutley is in New Jersey. Let's call this known truth 'the Nutley proposition.' It is not plausible that any part of Smith's justification for believing the Nutley proposition is his state of knowing the Nutley proposition. For one thing, since knowledge entails truth, Smith's having justification that includes the state of knowing would give Smith entailing justification for the Nutley proposition. Yet intuitively our justification for such external world beliefs is not that strong.

A further problem is this. It looks as though it follows from KM and mentalism that Smith could also know, via an entailing justification, that he knows the Nutley proposition. After all, the mental state of his knowing that proposition entails itself. Yet knowledge that one possesses some particular item of external world knowledge seems not to be that epistemically secure. This sort of meta-knowledge seems at least as insecure as the content of the knowledge, and as we just noted, that content seems not to have entailing justification.

Thus, if KM is correct, then mentalism seems to be in trouble. It seems to imply that we have stronger external world justification than we actually have.

Internalism is a classification of theories of justification by the location of a theory's supervenience base for justificatory status. Internalism does not imply that everything internal actually justifies. The combination of KM and the mentalist version of internalism merely makes states of knowledge eligible to justify within a view that is internalist. The combination implies nothing about whether any given mental states justify, or what propositions the justifying ones justify. Those specifics are the job of specific internalist theories.

It is intuitive that Smith's knowledge of the Nutley proposition does not justify his belief in that proposition. This intuition is best understood as the intuition that the knowledge does not give him a particular sort of justification for the belief, namely, the justification that helps to constitute his knowledge of it. We can see this as follows. One intuitive specification of the epistemic justification that is required for knowledge is that the justification is 'how the person knows.' When we consider how Smith knows the Nutley proposition, the intuitive answer is that this is accomplished on the basis of other mental states and events. In typical cases of continuing knowledge, these would prominently include Smith's memories about Nutley's location. In typical cases where the belief is currently becoming justified, the knowledge-giving justification would be his perception of a report of the Nutley proposition from some source that Smith has sufficient reason to trust. It is this knowledge-giving sort of justification that is plainly not entailing.

A mentalist who affirms KM is free to say that these knowledge-constituting states contribute justification, and that no part of Smith's knowledge-constituting justification for the belief consists in his knowing it to be true. Clearly the same goes for any knowledge of that knowledge. If Smith does have this meta-knowledge, he knows it by having some justification for it other than itself. If we ask Smith, 'What justification do you have for your belief that you know that Nutley is in New Jersey?,' he cannot accurately answer this question by replying, 'My justification is simply that I do know that Nutley is in NJ.' Intuitively, that sheer fact that it is knowledge is not how he knows that it is. He needs something else that serves as good grounds for him to regard it as knowledge, These good grounds might, for instance, be products of Smith's reflection on the strength of his knowledge-constituting evidence for his belief in the Nutley proposition.

Given that Smith has the separate justification that helps to constitute his knowing the Nutley proposition, he may also have his knowing of the Nutley proposition as justification for some things. It may even be that one justification that Smith has for the Nutley proposition itself is that he knows that proposition. Mentalists can say this. But again, that knowledge is not the justification by which he knows the Nutley proposition, and mentalists who affirm KM have no reason to say that it is.[10]

Thus, while mentalists can attribute a justifying role to states of knowledge, they need not. The implications of KM leave mentalists free to not count states of knowledge as justifiers of anything, including themselves and their entailments. This versatility is a strength of mentalism.

7. EXTERNAL PARTS OF KNOWLEDGE

A known fact is usually external to the knower, as are at least some of the causal connections that make the fact rightly related to the person to be known. It might be thought that an internalist must say that knowledge is otherwise entirely internal. Internalists hold that any epistemic justification that is required for knowledge is internal.[11] And of course believing is internal. So it might be thought that an internalist would be committed to this much about the internal contribution to knowledge:

PK. If S has some possibly true justified belief, B, then there is some possible situation in which an internal duplicate of S knows B.

Intuitively put, the possible situations that confirm PK are ones where the external world cooperates with the internally justified belief so as to yield its truth and its proper connection to the believer.

In *Knowledge and its Limits* Williamson has in effect argued against PK, on the assumption that content externalism is true (2000: 58). He argues by use of the following example. Suppose that Smith has the experience of the sound of a barking dog in what appears to Smith to be a situation where a barking dog is not unexpected. Smith uses the sentence, 'That dog is barking,' in an attempt to

[10] It is helpful to compare this case with another where internalists clearly have the same sort of latitude. Wanting to be happy is an internal state. Although this is clearly so, internalists can unproblematically claim that having this desire does not always contribute to justifying a person's belief that she has the desire. Internalists can reasonably affirm that, at least sometimes, someone's belief that she wants to be happy is justified only by inference from other mental states, if at all. These other mental states might be things like her memories of the role that has been played by opportunities to become happy in guiding her practical deliberations.

[11] As Sandy Goldberg has pointed out, internalists need not hold that the sort of justification about which they are internalists is a requirement for knowledge. But it is generally agreed that the justification about which internalists theorize is a requirement for knowledge, if any justification is required.

believe a proposition expressed by the sentence. From this, Smith infers that some dog is barking. The latter belief seems to be epistemically justified, at least by internalist lights. So it might seem that Smith's belief that some dog is barking would be something that he knows, as long as the external world is cooperating by including a barking dog. But the final fact in Williamson's example is that Smith is having an auditory hallucination rather than hearing the sound of a barking dog. Given this final fact, intuitively Smith does not know that some dog is barking.

Now consider the version of content externalism according to which, owing to Smith's failed attempt to demonstrate a dog, Smith's token of 'that dog is barking' expresses no proposition. In any possible circumstance in which Smith is mentally the same, Smith has the same beliefs. So given this version of content externalism, in all such situations Smith also fails to believe a proposition by the corresponding use of the sentence 'that dog is barking.' Smith's inference to the conclusion that some dog is barking is always based on an empty mental affirmation, one that has no propositional content. Consequently, the inference is always defective.[12] Whether or not it is a fact that some dog is barking, and however that fact relates externally to Smith, this defective inference would prevent Smith from thereby knowing that some dog is barking under all such possible circumstances. Thus, if the internalist is right about the justification of the belief, then PK is not true.

This example refutes the claim that any possibly true belief which is epistemically justified by internalists' lights can be known with the addition of a truth-maker for the proposition believed that is appropriately externally connected to the person. It might be thought that PK remains plausible. That is, it might be thought that any genuinely justified possibly true belief can be known by having appropriate external conditions in place. If PK is correct for this reason, then the internalist view that the belief in our example is justified is a mistake. Perhaps there is some external element in justification that is missing in the present case, because of the externally defective basis on which the true belief is inferred. So now internalism has been brought into question.

There is no genuine trouble for internalism here. First, when internalism is understood as mentalism, an internalist can reasonably deny that Smith's belief that some dog is barking is even justified. This is another place where content externalism aids mentalism. Smith's inference from the attempted singular belief to the existence of some barking dog can be argued not to justify the latter. It can be contended that this inference has a mental justificatory defect. It can be contended that a justifying inference here requires Smith to have concepts

[12] The externalist view according to which the sentence fails to state any proposition makes it clearest that the inference is defective. But an externalist view according to which there is a proposition believed, but it has a gap in it and could not be true, also supports the defectiveness claim.

enabling him to grasp the thought that is its premiss. Employing content externalism, a mentalist is in a position to deny that Smith has any concept formulated by 'that dog' under the circumstances. Smith has no such concept because an external condition needed for its existence is missing, namely, the presence of a demonstrated dog. The mentalist can credibly claim that, in the absence of this concept, the failed attempt to form a particular belief does not result in a mental state capable of justifying anything by inference from its apparent content.

For the sake of concurring that Williamson's example refutes PK, it is optional for an internalist to deny that Smith has a justified belief that some dog is barking. An internalist can hold instead that an inference from an empty premiss does justify an inferred belief under certain conditions, such as the following: the one who does the inferring is suitably proficient and has suitable background information, the inference presents no appearance of being defective, and the person lacks any other reason to question its success. Smith's belief that some dog is barking might meet some such conditions. Counting the belief as then justified does not leave a mentalist without resources to explain the lack of knowledge in Williamson's example. If Smith's belief that some dog is barking is a justified true belief, then it is best to regard the example as a Gettier case. The barking dogs that make the belief true never include one about which a particular belief is formed by Smith, since his attempt to form a particular belief about a barking dog is based on a hallucination. So the fact that some dog is barking is made true in a way that is accidental in relation to Smith's basis for his belief. This sort of accident is the hallmark of a Gettier case. Internalism presents no obstacle to denying that Gettier cases are knowledge.[13]

8. CONCLUSION

Content externalism makes no trouble for epistemic internalism understood as mentalism. Where the impact of content externalism is not positively beneficial, it is positively benign.

[13] It is of interest that, given content externalism and mentalism, the same knowledge-excluding accidental connection to the fact exists in all possible cases that are internally just like this one. But such examples arguably do not require implications of content externalism. Suppose that some form of inference is subtly fallacious, and Smith is bearing in mind a reliable assurance that it is a valid form. This may be enough for the form of inference to justify its conclusions for Smith. Still, if Smith employs an actually defective inference, then Smith does not gain knowledge of the conclusion drawn. So in any case that is internally just like the actual one, Smith fails to know any conclusion drawn by this form of inference. In this sort of example, content has no role in explaining the defect. So it seems that the examples where content externalism does play a role do not raise any special epistemic issue.

Ram Neta has pointed out that there are other sorts of apparent counterexamples to PK. For instance, concerning the proposition that no one knows anything, B1, it seems possible that B1 is justified for someone. B1 is possibly true. But it is not possible for someone to know B1.

REFERENCES

Alston, W. 2001: 'Internalism and Externalism in Epistemology.' In H. Kornblith (ed.), *Epistemology: Internalism and Externalism* (Oxford: Blackwell).

Braun, D. 1993: 'Empty Names.' *Noûs*, 449–69.

Conee, E. and Feldman, R. 2004: 'Internalism Defended.' In *Evidentialism: Essays in Epistemology* (Oxford: Oxford University Press).

Pryor, J. 2001: 'Highlights of Recent Epistemology.' *British Journal for the Philosophy of Science* 52: 95–124.

Williamson, T. 2000: *Knowledge and its Limits* (Oxford: Oxford University Press).

4

How to be a Neo-Moorean

Duncan Pritchard

> Moore's mistake lies in this—countering the assertion that one cannot know [. . .] by saying 'I do know it'.
>
> (Wittgenstein 1969: S. 521)

1. MOOREANISM

My aim here is a somewhat ambitious one: to sketch the contours of a research project. In particular, what I'm interested in is whether we can add some flesh to a certain style of response to the problem of scepticism, one that has historically been the source of much derision but which, as we will see, derives support from a number of theses that one can find being defended in the current literature—so long, that is, as one integrates these theses into a coherent whole.

The response to scepticism that I have in mind mirrors in key respects the 'commonsense' proposal often ascribed to G. E. Moore. As the story goes, Moore would respond to radical scepticism as it is typically conceived of in the contemporary literature by offering the following sort of argument.[1] First, he would claim that he knew a paradigm 'everyday' proposition—i.e. a proposition

Earlier versions of the material in this chapter have been presented on a number of occasions, including the following conferences: 'Contextualism', Bled, Slovenia (June 2004); 'Contextualism', Amsterdam Free University (October 2004); 'Externalism and Internalism in Semantics and Epistemology', University of Kentucky (April 2005); and 'Disjunctivism', University of Glasgow (June 2005). Earlier versions of this material have also been presented at talks at Faculties of Philosophy at the Universities of York, Birmingham, Copenhagen, Aarhus (Denmark), Rijeka (Croatia), and Hull (UK) during 2004 and 2005. Special thanks to Jessica Brown, Tony Brueckner, Earl Conee, Gary Ebbs, Sandy Goldberg, Adrian Haddock, Marie McGinn, Alan Millar, Ram Neta, Jim Pryor, Finn Spicer, and two anonymous referees. Finally, I am grateful to the Arts and Humanities Research Council for the award of a research leave grant which enabled me to conduct work in this area.

[1] I say 'as the story goes', since my interest here is not with what Moore actually said, or meant to say, but rather with what he is typically supposed to have said. Indeed, I think that a close examination of the key articles in this regard (Moore 1925, 1939) indicates that there is good reason to think that he might not respond to BIV scepticism in the manner outlined here. For an overview of some of the issues, see Baldwin 1993.

that all of us would think we knew, in those circumstances—such as, famously, that he has two hands. A claim of this sort is surely intuitive. Second, he would note that since having two hands is inconsistent with the relevant sceptical hypothesis, such as the hypothesis that he is a (handless) brain-in-a-vat (BIV), it follows that if he knows that he has two hands then he also knows that he's not a BIV. Again, this claim is also intuitive. Finally, he would conclude that he knows that he's not a BIV. Here, then, is a run-down of the argument:

Mooreanism
M1. I know that I have two hands.
M2. If I know that I have two hands, then I know that I'm not a BIV.
MC. I know that I'm not a BIV.

An anti-sceptical conclusion is thus validly derived from two intuitive premisses. In what follows, we will treat this argument as representative of the Moorean response to scepticism.

Note that the whole point of Moorean anti-scepticism is that there is nothing more to the stance than the presentation of an argument of this sort. The sceptic has called our knowledge into question, via the presentation of the sceptical hypothesis, and the Moorean, via his opposing argument, has rebutted the sceptic's claims. Thus, there is no case to answer, and hence nothing more that needs to be said. In this sense, then, the Moorean stance is a *pre-theoretical* proposal, in that it attempts to deal with the sceptical challenge in an entirely commonsense way which avoids the need for a theoretical response to the problem.

I think that many who engage with scepticism for the first time find a response of this sort compelling; indeed obvious. Interestingly, though, this conviction usually doesn't last very long. Instead, many initial proponents of Moorean-style anti-sceptical arguments quickly abandon their claims once they realize that, on closer inspection, such an argument is not as intuitive as it first appears. With Mooreanism abandoned, the scene is set for various other anti-sceptical theses to come to the fore which are all far more complex and theoretical than the Moorean response—such as arguments for non-closure, contextualism, and contrastivism, to name but three of the big anti-sceptical theories that have come to prominence in recent years.

Nevertheless, I think it is true to say that many epistemologists secretly look back on the abandoned Moorean response to scepticism with a rueful sigh, and still view it as being the, alas unrealizable, apogee of anti-sceptical endeavour. Roughly speaking, the (unspoken) thought here is this: *if only we could have salvaged the Moorean response, then we wouldn't have had to engage with all these myriad anti-sceptical theories, all highly theoretical and all beset by their own fundamental problems.*

It is not my aim here to show how one might salvage Mooreanism as such, since I share the conventional wisdom that this view is unsalvageable. That said, however, I do think that there is a sister view available—what I have termed 'neo-Mooreanism'—which shares many of the key features of Mooreanism but

which *can* be made to work.[2] As we will see, such a view is certainly more complex and theoretical than its Moorean counterpart, but it is still—or so I claim—less revisionary than other competing anti-sceptical views, and thus it is in this sense more Moorean than anti-Moorean.

2. THE FAULTS OF MOOREANISM

Before we can resurrect the Moorean proposal, we first need to be clear just what is wrong with it. While most epistemologists agree that Mooreanism is a non-starter, there is very little in the way of consensus about exactly why the view is problematic. I want to suggest that there are six key problems with the view, many of them interrelated, which thus identifies six problems that a neo-Moorean position has to overcome.

Perhaps the most common complaint levelled at the Moorean argument is that there is something question-begging about responding to the sceptical problem in this way, in that it simply takes as an unquestioned premiss in its argument the denial of the very claim that the sceptic will want to motivate as a conclusion of her argument. Call this the *dialectal impropriety* objection.[3]

We can get a grip on what the problem is here by considering the sceptical argument that the Moorean stance is supposed to be a response to. Keeping to the first-person, as Moore does, and sticking with the examples used above of having two hands and not being a BIV, we can formulate the opposing sceptical argument as follows:

Scepticism
S1. I don't know that I'm not a BIV.
S2. If I know that I have two hands, then I know that I'm not a BIV.
SC. I don't know that I have two hands.

Both of these premisses are intuitive. The first is intuitive because it seems that whether or not we are BIVs is just something that we could never know because sceptical scenarios are defined such that there is nothing in our experiences that could offer us any definitive indication one way or another as to whether we are the victim of such a scenario. The second premiss is exactly the same as the second premiss in the Moorean argument, which we saw was intuitive above. This premiss can be further motivated in terms of the closure principle:

Closure
For all S, φ, ψ, if S knows φ, and knows that φ entails ψ, then S also knows ψ.[4]

[2] I first coined the phrase 'neo-Mooreanism' in Pritchard 2002d.
[3] See, for example, Wright 2002 for a complaint against Mooreanism of this sort.
[4] One might want to modify this principle in a number of ways in order to deal with potential counterexamples of a trivial sort (such as possible cases where the agent doesn't even believe the entailed proposition), but this unembellished version of closure should suffice for our purposes here.

Closure certainly seems plausible, since it is hard to see how this principle could fail. How could one know one proposition, know that it entailed a second proposition, and yet fail to know the second proposition? Crucially, however, with closure in play—and given that one knows the relevant entailment, as presumably one does—it follows that if one knows that one has two hands then one also knows that one is not a BIV. We thus get the second premiss, motivated in terms of the highly intuitive closure principle.

With these two premisses in hand, however, the sceptical conclusion immediately follows. And since this argument can be repeated with any number of everyday propositions (one would just have to vary the sceptical hypothesis to suit), so the full intellectually devastating radical sceptical conclusion—that we are unable to know most of the empirical propositions which we typically think we know—is in the offing.[5]

With the sceptical argument and the Moorean argument set side by side, one can see that the debate here encapsulates that old philosophical chestnut that one philosopher's *modus ponens* is another philosopher's *modus tollens*. Whereas the Moorean takes his everyday knowledge as secure and argues on this basis that he also has the required anti-sceptical knowledge; the sceptic begins by highlighting the implausibility of anti-sceptical knowledge and argues on this basis that we also lack everyday knowledge. Indeed, notice that the second premiss of the Moorean argument can be regarded as resting on closure just as much as the second premiss of the sceptical argument. It is clearly not the closure principle, then, that is at issue in this debate.

With the debate so construed, however, one can see why the Moorean strategy can seem so dialectically inappropriate. The sceptic has given us an apparently compelling argument for thinking that we lack everyday knowledge. In response, the Moorean simply helps himself to the denial of the contested conclusion and reasons on this basis to the negation of the premiss of the sceptical argument. Given that the Moorean argument begins and ends with this strategy, it is little wonder that few find it persuasive. Imagine, for example, an atheist responding to an apparently compelling proof for the existence of God—let us suppose that such a proof could exist—by citing as a premiss God's non-existence! Clearly, we wouldn't think that such a person had engaged with his opponent, and the same seems true of the Moorean.

A second, and related, difficulty with the Moorean response is that it seems to offer us, at most, a *draw* with the sceptic, rather than a resolution of the sceptical problem. After all, given that the sceptical argument is just the *modus tollens* to Moore's *modus ponens*, and since both arguments have intuitive premisses, it appears that the dialectical situation is that we are faced with two opposing arguments of equal force. If this is right, then even despite the Moorean argument we still have just as much reason to be sceptics as to be Mooreans. Put another

[5] For more on the contemporary discussion of scepticism, see Pritchard 2002c.

way, it is still the case on the Moorean view that we have no good reason *not* to be sceptics. This is a kind of second-order scepticism which, while not obviously reducible to its first-order cousin (which would hold that we have reason to be sceptics), is still enough to make Moorean anti-scepticism not nearly as intellectually satisfying as it might at first appear. Call this the *impasse* objection.[6]

The third problem with Mooreanism is that the Moorean argument seems to consist of a series of assertions which strike one as conversationally inappropriate, if not just plain absurd or contentless. As a number of commentators have noted—most trenchantly Wittgenstein (1969) in his final notebooks—the assertions in question in the Moorean argument seem to offend against our usual usage of the term 'know'. In particular, we very rarely use the phrase 'I know that *p*' in order to convey the fact that we have knowledge of *p*—instead, we just assert '*p*'. This phrase thus plays a very special role in our practices of knowledge self-ascription, but, crucially, not one that seems applicable to the kind of anti-sceptical assertions that the Moorean makes. We never normally say we know the denials of sceptical hypotheses, and neither do we usually say we know everyday propositions which are just plain obvious to everyone in that conversational context. Call this the *conversational impropriety* objection.

The fourth problem with the Moorean stance is that it isn't backed up by a plausible epistemological theory about how it is that we might come to know that we are not BIVs, still less does it engage, for obvious reasons, with the contemporary literature on this topic which features, as we will see, a number of different proposals in this regard. The problem here is that, as we saw when we looked at the opposing sceptical argument, it is very intuitive to suppose that this is the sort of proposition which one could never come to know. Since it is an empirical proposition it isn't obviously in the market to be known *a priori*, but neither does it seem to be the kind of proposition that one could know empirically either since, *ex hypothesi*, there is no empirical investigation that one could undertake which would indicate to you that this hypothesis did not obtain. Accordingly, to be told on pre-theoretical grounds that one can know such a proposition is unlikely to offer any intellectual satisfaction, since one would also want to be told just how this could be so. Call this the *no supporting epistemology* objection.

This complaint about a lack of a supporting epistemology connects with a fifth worry about the Moorean proposal, which is that it does not support the anti-sceptical line with a *diagnostic* story which explains why we ever found scepticism to be so plausible in the first place. By the lights of this approach, it seems just plain *mysterious* why we were ever taken in by the sceptical problem. After all, if the proper way to deal with scepticism is this straightforward, then why wasn't it recognized all along? Why did we expend so much effort engaging

[6] Wright 1991 discusses this 'impasse' objection to Mooreanism, and offers an argument which shows that the second-order scepticism which results collapses into first-order scepticism.

with the problem in a myriad of complex ways when we could have dealt with it simply by offering the Moorean argument? Call this the *no diagnosis* objection.

There is also a sixth and final difficulty facing the Moorean response to scepticism that I want to mention, but which requires a little more explaining. In short, this problem is that Mooreanism doesn't really engage at all with a related form of scepticism which targets the *evidential basis* of our knowledge. Call this the *evidential scepticism* objection.

Consider the following argument:

Evidential Scepticism

ES1. My evidence for E does not favour E over the known to be incompatible BIV hypothesis.

ES2. If I know E, then my evidence for E favours E over the known to be incompatible BIV hypothesis.

ESC. I don't know E.

Whereas the sceptical argument we considered above, like the Moorean counterargument, traded on the closure principle for knowledge, this argument rests on the logically weaker 'underdetermination' principle, which we can roughly express as follows:

Underdetermination

For all S, φ, ψ, if S knows φ, and S knows that φ entails ψ, then S's evidence for believing φ favours φ over $\neg\psi$.[7]

In essence, this principle—a version of which can plausibly be found in ancient Pyrrhonian sceptical writings—demands that one's knowledge be evidentially supported, where evidential support here means support which *favours* what is believed over known to be incompatible alternatives (i.e. which provides *more* support for what is believed than it does for the known to be incompatible alternatives). For example, if you know that you are at present in the town's Odeon cinema, and you know that if you are in the Odeon then you are not in the town's other cinema, the Grand, then the evidence which supports your belief that you are at present in the Odeon must prefer this belief over the known to be incompatible alternative that you are at present in the Grand.

So construed, the principle seems entirely uncontentious, since it is hard to see how one's evidence could be genuinely supporting evidence if it did not perform this 'favouring' function. The trouble is, however, that once one feeds sceptical hypotheses into this principle then one immediately generates the sceptical problem. This is because if you know that, for example, you have two hands, and you also know that having hands is inconsistent with being a BIV, then it follows, via underdetermination, that your evidence for believing that you have hands must favour your belief that you have hands over the alternative

[7] For a thorough discussion of the underdetermination principle, including some of the different ways in which it can be formulated, see Pritchard 2005d.

hypothesis that you are a BIV. *Ex hypothesi*, however, this is impossible, and thus you are unable to know that you have two hands (and much else besides).[8]

The main point to note about the underdetermination principle is that it is more fundamental to the sceptical problem than the closure principle in at least two ways. To begin with, notice that the underdetermination principle is logically weaker than closure. I have not the space to elucidate this point at length here—see Pritchard (2005d) for the full argument—but one can get a sense of why this is so simply by looking at the kind of sceptical inferences that the two principles license. Taking the knowledge of the relevant entailment as given, the closure principle licenses an inference from knowledge of an everyday proposition, such as that one has two hands, to knowledge of an anti-sceptical proposition, such as that one is not a BIV. In contrast, the underdetermination principle only licenses an inference from knowledge of an everyday proposition, let's say two hands again, to the claim that one's evidence favours one's belief that one has two hands over sceptical alternatives, such as the BIV hypothesis. Since the antecedent is the same in both cases, we just need to focus on the consequent. Intuitively, the claim that one knows that one is not a BIV is much stronger than the claim that one's evidence favours having two hands over being a BIV. Indeed, it seems that all the latter immediately entails is that one does *not* know that one *is* a BIV, and that is certainly weaker than the claim that one *knows* that one is *not* a BIV. *Prima facie*, then, the underdetermination principle is logically weaker than closure, and this means that merely denying the closure principle will not suffice by itself to block the sceptical challenge.

There is also a second—and, I think, more important—sense in which the underdetermination principle is more fundamental to the sceptical problem than the closure principle. This is that the sceptical argument is best thought of as primarily attacking the evidential basis of our knowledge. Think, for example, of how the sceptic motivates the first premiss of her argument, (S1). Clearly, the claim here is that one is unable to know that one is not a BIV because one, perforce, lacks evidence for a belief in this proposition. An evidential claim is thus right at the heart of the standard sceptical argument, even though there is no explicit mention of the agent's evidence in that argument. Making this evidential aspect of the sceptical argument explicit is important because it highlights just how counterintuitive the Moorean conclusion, (MC), is. To say that we can know the denials of sceptical hypotheses is one thing; to say (as seemingly the

[8] Now one might want to qualify this principle in order to accommodate the intuition that not *every* belief needs to be evidentially grounded in order to be an instance of knowledge. Notice, though, that such an amendment to the principle wouldn't necessarily make any real difference to its ability to generate sceptical conclusions. After all, it is surely the case when it comes to *most* of our beliefs in empirical propositions that they are only evidentially grounded to a degree that would support knowledge provided that such evidence is able to play this 'favouring' role. And since the sceptic only needs to call the epistemic status of most of our beliefs in empirical matters into question in order to motivate her sceptical doubt, this weaker construal of the underdetermination principle would, it seems, serve the sceptic's purposes equally well.

Moorean must) that one has adequate evidence for one's belief in the denials of sceptical hypotheses is quite another, and far more problematic.

Evidential scepticism gets right to the heart of this issue by focusing on an evidential thesis and highlighting how this thesis can itself engender the sceptical problem. Moreover, note that attempting a Moorean response to the evidential sceptical challenge just does not seem possible. Suppose, for example, that the Moorean responded by 'inverting' the sceptic's reasoning in the way attempted above as regards closure-based scepticism. The evidential Moorean argument would then go something like as follows:

Evidential Mooreanism
ME1. I know E.
ME2. If I know E, then my evidence for E favours E over the known to be incompatible BIV hypothesis.
MEC. My evidence for E favours E over the known to be incompatible BIV hypothesis.

Clearly, unless this argument is supplemented with further philosophical support, it is extremely suspect. It is one thing to say, on 'commonsense' grounds, that since I know E it must also be the case that I know that the BIV hypothesis is false; quite another to specifically claim that I have evidence which favours E over the BIV hypothesis. The underlying problem here is that to respond to scepticism in this way is not, it seems, to offer a 'pre-theoretical' response to the sceptic at all since it directly issues in a very counterintuitive and highly theoretical claim about the evidential basis of our anti-sceptical knowledge. Given the oddity of the conclusion, the Moorean strategy of refusing to engage further with the sceptical problem once the relevant Moorean argument has been offered seems just plain dogmatic.

We have thus identified six problems with Moorean anti-scepticism, and this means that we have thereby set out what work needs to be done to get a neo-Moorean response on a sound footing. Essentially, what we are looking for is a neo-Moorean anti-sceptical account which shares as many of the central features of the Moorean response to scepticism as possible consistent with it no longer being subject to these complaints. This is what I will be attempting in what follows, with one further constraint. This is that we want the neo-Moorean response to scepticism to have dialectical advantages over its main competitor anti-sceptical theories, and so we need to evaluate the view in this light. As we will see, bringing in the competition in this way is useful in that neo-Mooreanism can exploit aspects of other anti-sceptical views to its own advantage.

3. SCEPTICISM AND CLOSURE

Before we get down to the nitty-gritty of the task before us, I want to make one simplifying assumption explicit. This is that in what follows I will take it

as given that closure holds (at least in some slightly modified form). I take it that most contemporary epistemologists no longer have the inclination to deny this principle outright. Indeed, I suspect that Fred Dretske (2005a; cf. Dretske 1970; Nozick 1981) is perhaps the only prominent epistemologist left who is still defending the outright denial of closure. Part of the reason for this is that epistemologists are now more aware of the alternatives to denying closure here—especially contextualism, which we will consider in a moment. But the main reason is still that rejecting such a principle seems to be itself a form of intellectual self-harm, and so unable to offer us any comfort in our dealings with scepticism. Moreover, recent work on closure has made explicit just how intellectually disastrous the rejection of this principle would be.[9] In any case, as we have already noted, denying closure is of little help when it comes to the sceptical problem unless one is also in a position to deny the underdetermination principle as well, and that fact in itself diminishes the importance of responses to scepticism that rest on the rejection of closure alone.

Accepting closure means accepting that one's response to the sceptic must be, as Jonathan Schaffer (2004: 22) has put it, *immodest*, in the sense that it must allow that we can know the denials of sceptical hypotheses. The primary task for the neo-Moorean, as with any proponent of an anti-sceptical strategy which is immodest in this way, is to account for such knowledge.

One final coda is in order here, which is that in moving directly from the claim that we should retain closure to the claim that we should therefore adopt an immodest response to the sceptic, one might object that I am ignoring the contrastivist response to scepticism, as developed, for example, by Schaffer (e.g. 2005b). On this view, knowledge is always to be understood as a contrastive notion, such that one never knows that *p*, but always knows that *p rather than* a set of contrasts, *Q*. Knowledge is thus to be understood as a three-place relation between a subject, a proposition, and a contrast class, rather than in the usual way as a two-place relation between a subject and a proposition.

It is indeed true that I am ignoring the contrastivist option here, and this is primarily because I am not convinced that contrastivists can retain closure, at least not in any form that we would recognize as closure. After all, contrastivists like Schaffer explicitly grant that there is no contrast relative to which we are able to know the denials of sceptical hypotheses, and thus the *prima facie* tension between closure and contrastivism is manifest. Schaffer (2005a) himself has gone to great lengths to show how one could 'contrastivize' the closure principle in order to reconcile this tension with closure intact, but it is hard to see why the astonishingly complex set of principles that results should be thought to model closure. That is, given the extent of the revision in play, it is hard to see why Schaffer isn't simply denying the principle rather than merely modifying it.

[9] See, for example, Hawthorne 2005, to which Dretske 2005b responds. I explicitly discuss anti-sceptical arguments that are predicated on the rejection of closure in Pritchard 2002a, 2002b.

In any case, there is an overriding reason why we can set both of these forms of immodest anti-scepticism to one side, which is that if we can make neo-Mooreanism a palatable anti-sceptical strategy, as I think we can, then there is no longer any obvious need to deny (or, if you prefer, radically modify) closure in order to deal with the sceptical problem, and thus the impetus to endorse anti-sceptical theories of this sort subsides accordingly. Moreover, as we will see, neo-Mooreanism as it will be developed here can accommodate a substantial part of the motivation for contrastivism anyway. With this in mind, we will now focus on those immodest anti-sceptical theories that remain under consideration.[10]

4. SCEPTICISM, CONTEXTUALISM, AND NEO-MOOREANISM

As I just noted, what all immodest anti-sceptical views have in common is that they claim that we can have knowledge of the denials of sceptical hypotheses. The chief anti-sceptical theory in this regard—and the neo-Moorean's main competition—is attributer contextualism, as defended, for example, by such figures as Keith DeRose (1995), David Lewis (1996), and Stewart Cohen (e.g. 2000). There are other forms of contextualism of course, such as the variety of subject contextualism—confusingly known as *subject-sensitive invariantism*—recently defended by John Hawthorne (2004) and Jason Stanley (2004, 2005). If, however, we can show that neo-Mooreanism has distinct advantages over attributer contextualism, then this will constitute a substantial part of the task of motivating the view over other contextualist theories as well, especially since, as we will see, the key advantage that neo-Mooreanism has over its main contextualist rival is that it demonstrates that we do not need to suppose that there is any substantial epistemic context-relativity in order to deal with the sceptical problem. Thus we will focus our attentions on attributer contextualism.

Attributer contextualism (henceforth just *contextualism*) holds that 'knows' is a context-sensitive term in the sense that the truth of ascriptions of knowledge will depend upon the context in which those ascriptions are made (and thus the context of the *attributer*). In particular, contextualists maintain that different contexts employ different epistemic standards, such that while an agent might meet the epistemic standards in operation in one context of ascription—so that relative to this context an ascription of knowledge to this agent would express a truth—this is consistent with that agent failing to meet the more demanding set of epistemic standards in operation in another context of ascription—so that relative to this context an ascription of knowledge to this agent would express a falsehood. Given the broadly indexical nature of 'knows', however, there is no

[10] I consider the contrastivist response to scepticism in its own right in Pritchard forthcoming a.

conflict between these two claims, since the proposition that is being expressed in the one context is not the same proposition that is being expressed in the other, and thus the one can be true while the other is false without contradiction.

The basic contextualist idea is that while the epistemic standards are high enough in sceptical contexts to ensure that any ascription of knowledge of either everyday propositions or the denials of sceptical hypotheses would be false; nevertheless the epistemic standards are low enough in quotidian contexts to ensure that any ascription of knowledge of either everyday propositions or the denials of sceptical hypotheses would (normally) express a truth. Thus, relative to the standards in operation in quotidian contexts at any rate—which are presumably the contexts that are of most interest to us—we do come out as having the everyday knowledge that we typically take ourselves to have and also the anti-sceptical knowledge which we must have if closure is accepted and scepticism is false.

So construed, contextualism is primarily a linguistic thesis which must be backed up by an appropriate epistemological theory. We do not need to get into the details of the different formulations of the contextualist epistemological thesis in order to see that there is a Moorean aspect to this type of anti-scepticism (indeed, some contextualists have explicitly referred to the view as 'Moorean'). This aspect relates to the immodesty of the position in allowing that we can know the denials of sceptical hypotheses. Notice, however, that this Moorean commitment on the part of contextualism also points to why a neo-Moorean view ought not to be a contextualist thesis if this is at all possible. After all, the contextualist is committed to *two* revisionary theses here, when it seems that just one of them would suffice. That is, the contextualist is denying not only the intuition that we cannot know the denials of sceptical hypotheses but also the invariantist (i.e. non-contextualist) intuition that 'knows' is not a substantively context-sensitive term (i.e. not context-sensitive in the kind of radical way that it would need to be if it is to enable us to resolve key epistemological problems). Interestingly, however, the first claim alone would suffice by itself to block the closure-based sceptical argument, since this is the denial of one of the two key premisses in that argument, (S1). Furthermore, notice that going contextualist without also denying this premiss is of little appeal, since it would mean that one would have all the disadvantages of being a contextualist along with all the disadvantages of rejecting closure.[11] Given that one should always minimize one's revisionism where possible, and bearing in mind also the general nature of the Moorean response which tries to evade the need for complex theorizing, it should be clear that the natural way for the neo-Moorean to go is to aim to simply deny the first premiss of the sceptical argument without also advancing a form of contextualism.

[11] For more on this point, see my discussion of Heller 1999 in Pritchard 2000.

There is also a further reason to prefer a form of neo-Mooreanism run along these lines over a contextualist account, and this concerns the inherently concessive nature of the contextualist thesis. After all, it is part of the contextualist response to scepticism to allow that there are some contexts of ascription in which it would be true to say that we lack everyday knowledge. Given that sceptical contexts are by their nature contexts in which more demanding epistemic standards are being employed than is usual, the problem with this concession is that it prompts the natural thought that, strictly speaking, we do not have the knowledge which we typically ascribe to ourselves. That is, that while, speaking loosely, it is in some sense correct to ascribe knowledge to ourselves in quotidian contexts where the epistemic standards are not very exacting, once one tightens up one's use of the language one realizes that those ascriptions are, strictly speaking, false. Put simply, the worry here is that contextualism seems to validate a train of infallibilist thinking which leads us right back into the sceptical problem. If we can avoid making such a concession, as neo-Mooreanism, as we are now understanding that view, would, then this is all to the good.

A related issue here is the unusual way in which the contextualist is understanding the relevant alternatives thesis. Relevant alternatives epistemology is a form of fallibilism, in that it demands that not all error-possibilities must be eliminated in order for an agent to have knowledge. Instead, just a relevant sub-set of the total class of error-possibilities will need to be eliminated. I take it that the core relevant alternatives intuition is that those error-possibilities which only obtain in far-off possible worlds are automatically irrelevant, in that modal closeness is a prerequisite for relevance. Crucially, however, it is the sceptic's claim that empirical knowledge (most of it at any rate) is impossible, and thus the sceptical argument should go through however we understand the actual world to be. With that in mind, we are entitled to suppose that the actual world is much as we take it to be, and thus that sceptical hypotheses are indeed modally far off, and then ask the question of whether we have any empirical knowledge. The reply given by those who subscribe to the core relevant alternatives intuition is that on this supposition we do have knowledge of much of what we take ourselves to know—because in this case sceptical hypotheses are irrelevant to our knowledge—and thus that the sceptic is wrong to say that it is impossible to possess the empirical knowledge that we typically take ourselves to possess.

There is a strong rationale for the core relevant alternatives intuition, and it arises out of the platitude that knowledge is, at root, non-lucky true belief. For so long as sceptical error-possibilities are indeed modally far off, then the mere fact that one would not be able to tell a sceptical and a counterpart non-sceptical scenario apart should not by itself suffice to indicate that one's true beliefs in everyday propositions—which entail the denials of sceptical hypotheses—are thereby only luckily true. Put another way, if sceptical error-possibilities are indeed far-fetched, then it is not the case that one's everyday beliefs—and

one's anti-sceptical beliefs for that matter—could have been in this regard very easily false.

It is an anti-luck intuition of this sort that, I take it, lies behind safety-based theories of knowledge—as defended, for example, by Ernest Sosa (1999)—where safety means something like 'could not have easily been false'. There are various ways of formulating this principle, not all of them plausible, but the basic formulation has it that for a true belief to be safe it must be the case that, across a wide range of nearby possible worlds, where the agent believes the target proposition (on the same basis), that belief continues to be true. Note that the anti-luck intuition encapsulated in safety is also what lies behind the contextualist's claim that we can know everyday propositions and the denials of sceptical hypotheses relative to quotidian contexts of ascription (indeed, DeRose 1995 explicitly appeals to a safety-type principle in this regard). What is important for our purposes, however, is that this intuition can be read as motivating the core relevant alternatives thesis that far-off error-possibilities are by default irrelevant to knowledge possession. The core relevant alternatives intuition and the anti-luck intuition thus go hand in hand.[12]

Contextualists are fallibilists, and thus relevant alternatives theorists, in the sense that they allow that an ascription of knowledge to an agent can be true even though that agent is unable to rule out all possibilities of error.[13] This will certainly be the case in quotidian contexts of ascription, where the epistemic standards are low. Notice, however, that contextualists do not subscribe to the core relevant alternatives intuition just noted, since they allow that even scenarios which are, *ex hypothesi*, modally far off can still be relevant to knowledge ascription if the epistemic standards in the context of ascription demand that one take them into consideration. In sceptical contexts, for example, where sceptical hypotheses are explicitly at issue, whether or not an agent is truly ascribed knowledge will be dependent upon whether that agent is in a position to rule out the sceptical hypothesis, even if, as a matter of fact, that hypothesis is indeed modally far off.

This points to a further juncture at which the neo-Moorean and the contextualist should part company, in that the neo-Moorean should, it seems, cling on to that core relevant alternatives intuition—and the associated anti-luck intuition—and insist that just so long as the actual world is roughly as we take it to be then we are able to have knowledge even despite our inability to eliminate modally far-off sceptical hypotheses. Of course, the neo-Moorean will have to explain how, given closure, one is able to know the denials of sceptical hypotheses given that this is the case, but that is a problem shared by the

[12] I discuss this point at length in Pritchard 2005b, *passim*. See also Pritchard 2004, 2005c, 2007.

[13] Notice that this formulation of fallibilism is consistent with Lewis' 1996 famous claim to be an infallibilist. This is because it is an *explicit* fallibilism that Lewis rejects, such that an ineliminable error-possibility is taken into consideration and yet the agent is ascribed knowledge nonetheless.

contextualist anyway, and so not relevant to the specific issue of theory choice when it is just these two theories under consideration.

There are thus strong *prima facie* grounds for preferring a neo-Moorean anti-sceptical strategy over a contextualist one. There is, however, one key feature of the contextualist position which speaks in its favour in this respect, and that is that contextualism seems able to meet some of the problems we saw facing the Moorean response above. To begin with, the contextualist can account for why the assertions made by the Moorean seem so improper, the simple explanation being that in the sceptical context in which these assertions are, perforce, made, what is asserted is false (because very demanding sceptical epistemic standards will apply). Furthermore, the contextualist can diagnose the attraction of scepticism, since on this view the problem emerges out of a failure to realize that 'knows' is a context-sensitive term. Finally, the contextualist has the supporting epistemology to back all of this up. It is thus essential to neo-Mooreanism that it is able to supply a similar response to the problems that beset the Moorean strategy, since without it the view loses much of its appeal relative to its competitor contextualist theory.

5. EVIDENTIAL CONTEXTUALISM AND EVIDENTIAL NEO-MOOREANISM

This discussion of contextualism—and a potential neo-Moorean rival theory—is all by-the-by, however, since, as we noted above, it is essential that any anti-sceptical theory be able to deal with the threat posed by evidential scepticism, and it is far from obvious how the views just set out (which make no mention of evidence) would do that. There is, however, a contextualist theory in the literature—due to Ram Neta (2002, 2003)—which also tries to deal with the evidential sceptical problem, so it is worthwhile turning to this view for inspiration in this regard.

Neta's proposal is that we should treat what counts as evidence as being context-sensitive, and thus offer a contextualist account of knowledge in virtue of the context-sensitivity of evidence. In particular, Neta's claim is that in different contexts of ascription what counts as an agent's evidence can change, so that the evidential support that the agent has for her belief relative to one context of ascription might be strong (strong enough to make an ascription of knowledge in that context true, say), even though that same agent has very weak evidential support for her belief relative to another context of ascription which employs different epistemic standards (and so may not have sufficient evidential support to make an ascription of knowledge in that context true). Thus, the truth of ascriptions of knowledge is still a variable matter as it is on the standard contextualist account; it is just that the variability is explicitly accounted for in terms of the shifting standards of what counts as evidential support.

In essence, Neta's idea is that an agent's evidence is determined by the error-possibilities at issue in the context of ascription in the sense that the agent's evidence is the evidence the agent would have were the relevant error-possibilities to be true. In a context in which sceptical hypotheses are at issue, therefore, an agent's evidence can only be the evidence she would have were she to actually be, say, a BIV, and that will of course severely limit the scope of the agent's evidence. As far as perceptual evidence goes, for example, her evidence for her perception-based belief that there is, say, a table before her can only consist of the way the world seems to her (that it looks to her as if there is a table in front of her). With this in mind, the underdetermination-based sceptical argument gets a grip, since relative to this conception of what counts as the agent's evidence it will not favour her belief in the everyday proposition (in this case that there is a table before her) over the known sceptical alternative, and thus she will be unable to have knowledge of this proposition.

In quotidian contexts, in contrast, the error-possibilities at issue will not include sceptical hypotheses, and thus what counts as an agent's evidence relative to this context will inevitably be much broader. In particular, Neta claims that in quotidian contexts in which sceptical hypotheses are not at issue an agent's perceptual evidence can include factive evidence, such as that the agent sees that such-and-such is the case, where *seeing that p* entails *p*. If this is right, then relative to quotidian contexts an agent's evidence can be such that it favours the agent's belief in an everyday proposition, such as that there is a table before her, over a sceptical alternative, since relative to these quotidian contexts the evidence for the former could logically exclude the latter alternative.

Notice that this view is most naturally construed along content externalist lines as being allied to a form of disjunctivism. This is because the factive nature of one's perceptual evidence in quotidian contexts intuitively reflects the factive nature of one's perceptual experiences in those contexts, such that the content of one's perceptual experience can be such that it is determined, at least in part, by facts in the agent's environment (for instance, in the case just described, the fact regarding *p*), facts that would not obtain in corresponding cases of illusion or delusion where as a result the content of one's experience, and thus the nature of one's evidence on this view, would be different.

This is an intriguing suggestion, but note that, when evaluated relative to neo-Mooreanism, it faces similar problems to those we noted above as regards the standard non-evidential version of contextualism. In particular, the natural question to ask at this juncture is that if we can indeed make sense of the idea not just of an agent knowing that she is not the victim of a sceptical hypothesis, but also that this knowledge is evidentially supported within the constraint imposed by the underdetermination principle, then why should we be attracted at all to an evidential contextualism? Why not instead simply deny, on content externalist grounds, the key premiss in the underdetermination-based sceptical argument that we are unable to have evidence which favours our beliefs

in everyday propositions over sceptical alternatives and leave the matter at that? Indeed, once the role of the disjunctivist thesis in the arguments for evidential contextualism is made explicit, then the idea that one lacks the factive perceptual evidence in sceptical contexts that one possessed in quotidian contexts starts to look rather suspect. After all, how can the content of one's perceptual experience vary in line with purely conversational matters? And if it doesn't vary, then why should we concede that what counts as one's perceptual evidence changes with the shift of context?

Moreover, adopting a neo-Moorean form of evidentialism ensures that nothing of substance is conceded to the sceptic, since, as with standard forms of contextualism, evidential contextualism leaves one with the nagging thought that, strictly speaking, one lacks evidential support for one's beliefs in everyday propositions, it is just that, loosely speaking in quotidian contexts, it is acceptable to regard such beliefs as appropriately evidentially supported. An evidential neo-Moorean view that eschewed contextualism would not face this problem.

A further motivation for the neo-Moorean stance in this regard comes from the core relevant alternatives intuition that we noted earlier. So long as sceptical scenarios are indeed far-fetched, then why should they be relevant to the determination of knowledge in *any* context? More specifically for our present purposes, why should they be relevant to the determination of the scope of an agent's evidence in any context? And since we can connect this thought to the more general, and widely held, intuition that knowledge is at root non-lucky true belief, the support this consideration provides for an evidential neo-Mooreanism is quite strong.

Of course, the major advantage that evidential contextualism has over evidential neo-Mooreanism is the same one that non-evidential contextualism has over non-evidential neo-Mooreanism, which is that it has a story to tell about why we get taken in by sceptical arguments, and why Moorean assertions can seem so plain odd. Again, then, we need to remember that neo-Mooreanism, evidentially construed or otherwise, is in serious trouble unless it can incorporate the necessary diagnostic story.

6. VARIETIES OF EVIDENTIAL NEO-MOOREANISM AND THE CLASSICAL EPISTEMIC EXTERNALISM/INTERNALISM DISTINCTION

So far we have characterized neo-Mooreanism mostly in negative terms; in terms of which theses the neo-Moorean rejects. In particular, we have noted that a neo-Moorean will argue that we can know the denials of sceptical hypotheses, but will not contextualize this thesis by claiming that such knowledge is only possessed relative to certain contexts of ascription. Moreover, we have seen that

such a view would need to be given an evidential spin such that one's evidence can favour one's beliefs in everyday propositions over sceptical alternatives, thereby ensuring that one's beliefs in the known to be entailed anti-sceptical propositions, such as that one is not a BIV, can also be appropriately evidentially grounded. Again, it is important to the view that this evidential claim is not contextualized. Finally—and this is the only positive aspect of the view so far—we have started to see one way of motivating the neo-Moorean position in terms of a certain construal of the core relative alternatives intuition which we have noted is closely associated with the anti-luck requirement common to most theories of knowledge, and which can be encapsulated in some formulation of the safety principle.

These theses alone clearly do not yet represent an anti-sceptical theory, nor do they suffice to motivate neo-Mooreanism over contextualism unless they are supplemented with further claims that can match the diagnostic appeal of contextualism. Before we can make any headway at developing this view, however, it is important to first factor in the classical externalism/internalism distinction in epistemology since, as we will see, this has important ramifications for how such a view will be developed.

By classical epistemic internalism here, I mean *access* internalism such that what makes an epistemic condition (i.e. a condition which, perhaps in conjunction with other epistemic conditions, can turn true belief into knowledge) an internal epistemic condition is that the agent concerned is able to know by reflection alone those facts which determine that this condition has been met. Meeting the justification condition, for example, at least as it is standardly conceived, involves the possession of grounds in support of the target belief, where these grounds—and the fact that they are supporting grounds—are reflectively accessible to the subject. Classical epistemic externalism denies this, and so holds that there are epistemic conditions which do not demand reflective access on the part of the subject of this sort. I will understand classical internalism about knowledge as being the view that meeting a substantive internal epistemic condition is *necessary* for knowledge possession, with externalism about knowledge as the denial of this thesis.

There are other ways of drawing the internalism/externalism distinction of course. One could put the point in terms of supervenience rather than access—see, for example, Earl Conee and Richard Feldman (2000)—or one might weaken the internalist requirement by saying that one only needs reflective access to the supporting grounds for one's belief and not also to the fact that they are supporting grounds, as William Alston (1988) suggests (though note that he doesn't regard this view as an internalist thesis as such). The account of the distinction offered here is fairly standard, however, and, I believe, it also gets to the heart of what is at issue in the standard debates about epistemic externalism and internalism. Think, for example, about the normal cases over which internalists and externalists diverge, such as the chicken sexer case. Here we have an agent

who is exhibiting a highly reliable cognitive ability (to distinguish between the sexes of chicks), and yet who has false beliefs about how she is doing what she is doing (she thinks she is using her senses of sight and touch, when actually it is her sense of smell) and who typically lacks good grounds for thinking that she is reliable in this respect. Clearly such an agent does not meet a substantive internal epistemic condition as we have just defined that notion. Nevertheless, that, for the externalist at least, there is still an issue about whether or not the agent has knowledge indicates that what is in question here is the *necessity* to knowledge of meeting such an internal epistemic condition, which is just as it should be given how I have just characterized the classical externalism/internalism contrast. We will consider a non-classical form of internalism below, but for now we will focus on the distinction as it has just been drawn.[14]

With this distinction in mind it ought to be clear that motivating a classical internalist version of neo-Mooreanism is not going to be at all easy. There are two reasons for this. The first is that the primary intuition behind the idea that we are unable to know the denials of sceptical hypotheses is an internalist intuition to the effect that since there is nothing in our present experiences that we can, as it were, introspectively point to in order to indicate that we are not the victim of a radical sceptical hypothesis, hence we are unable to have adequate reflectively accessible grounds to support our belief in this respect. Accordingly, on an internalist account, lack of knowledge of the denials of sceptical hypotheses seems to follow fairly quickly.

The second reason why a classical internalist version of neo-Mooreanism would be tricky to sustain concerns the evidential basis of our putative anti-sceptical knowledge on this view. By the lights of an internalist epistemology one would expect one's evidence to be individuated on internalist grounds in terms of facts which are accessible to the agent. If this is right, however, then one's evidence even in cases where one is not being radically deceived can be no better than one's evidence in counterpart cases in which one is being radically deceived. After all, if the deception is such that there is nothing that is reflectively accessible to the agent which could indicate which of these situations is the one that obtains, it follows that one's evidence reduces to the lowest common denominator—or,

[14] For further discussion of the chicken-sexer example, see Foley 1987: 168–9, Lewis 1996, Zagzebski 1996: S. 2.1 and S. 4.1, and Brandom 1998. For more on the classical epistemic externalism/internalism distinction in general, see Kornblith 2001. Note that the chief problem with supervenience internalism is that it seems it can only accommodate the traditional epistemic internalist intuitions provided one does not combine the view with content externalism. After all, most internalists take as a datum to be explained by the theory that if an agent has a justified belief then so too does his envatted counterpart. On most content externalist views, however, there will tend to be a difference in the content of the mental states of the two agents, and thus this claim will not obviously go through. It may be, of course, that all formulations of the epistemic externalism/internalism distinction are hostage to what conclusions are derived from the debate about content externalism/internalism, but it is odd that such a basic incompatibility between epistemic internalist intuitions and content externalism should be so immediately apparent on this way of drawing the epistemic externalism/internalism distinction.

if you prefer, the highest common factor—between the two cases. As a result, an evidential version of neo-Mooreanism would be unable to account for the evidential basis for our anti-sceptical knowledge and would also be immediately susceptible to the underdetermination-based sceptical argument.

It is unsurprising, then, that there are very few proto-neo-Moorean stances in the literature that are conceived along classically internalist lines. Indeed, perhaps the only such account is due to Crispin Wright (e.g. 2004). It is part of Wright's view to grant that no warrant can, as he puts it, be 'earned' for beliefs in anti-sceptical propositions. Nevertheless, he claims that it does not follow that such warrant is not possessed. In essence, Wright's idea is that what the sceptic highlights is that it is essential to our cognitive projects that one has a standing entitlement to certain anti-sceptical propositions, such that the warrant in question is 'unearned'. That is, the idea is that given that it is essential to enquiry that one believe certain anti-sceptical propositions—and since without engaging in enquiry one is unable even in principle to earn a warrant for one's beliefs, hence the moral of scepticism is that not all warrants are earned.

This view faces the immediate problem of accounting for how, by classical internalist lights, an unearned warrant is any warrant at all, but we do not need to get into this issue here.[15] This is because even if the classical internalist neo-Moorean can make sense of the idea of an unearned warrant—perhaps in terms of the requirements of an epistemic rationality—it is still not clear how such a notion would enable the proponent of a view of this sort to deal with the sceptical challenge posed by the underdetermination-based argument. How could it be that by internalist lights one has evidence which favours one's belief in everyday propositions over sceptical alternatives? Indeed, Wright's claim that the epistemic standing of our anti-sceptical beliefs is 'unearned' is surely meant to imply that they have no evidential support at all.[16]

[15] I expand upon this objection at length in Pritchard 2005e.

[16] Another position in the literature which might plausibly qualify as a classical internalist version of neo-Mooreanism is that set out by Pryor 2000. Roughly speaking, Pryor's idea is that we have a default warrant—and in this sense 'unearned' and, thus non-evidential, warrant—for our basic perceptual beliefs on account of how forming beliefs in this way is not epistemically blameworthy. The thought is then that given this default standing epistemic status of our basic perceptual beliefs, we are entitled to make inferences to (known to be entailed) anti-sceptical beliefs, where this default epistemic standing transmits to the inferred conclusion. Our beliefs in the denials of sceptical hypotheses thus have a default epistemic standing in virtue of the default epistemic standing of our basic perceptual beliefs and the associated (correct) inference. Moreover, Pryor wishes to make this anti-sceptical move from within a classical internalist epistemology, so that the warrant in question is to be understood as satisfying classical internalist criteria. There is certainly a tradition of understanding classical internalism deontologically in terms of obedience to epistemic norms (where it is essential, of course, that one is in a position to determine by reflection alone what these norms are, and what one needs to do in order to follow them). As Pryor 2001 himself elsewhere concedes, however, it is not plausible to suppose that the minimal epistemic standing that accrues on this picture simply in virtue of not having contravened any epistemic norm will itself suffice, with true belief, for knowledge, even if we further suppose that the belief in question also meets a further anti-Gettier external condition. Thus, even if we can assuage our worries about how a

In any case, what is perhaps more important for our purposes is that the classical externalist is clearly in a stronger position than the classical internalist when it comes to motivating a neo-Moorean position. After all, that one lacks adequate reflectively accessible grounds for believing that one is not a BIV will not on this view decide the issue of whether or not one can have knowledge of this proposition. Similarly, such a lack of reflectively accessible grounds need not mean that one's evidence is restricted to merely the evidence that one would have were one to be in the counterpart deceived case. This is especially the case if one allies the view to a form of disjunctivism which allows that the content of one's perceptual experience can be different in non-deceived and counterpart deceived cases, since there is surely a close connection between the content of one's perceptual experience and one's perceptual evidence. Thus, the path is cleared to allowing one's anti-sceptical knowledge to be evidentially grounded and thus to blocking the underdetermination-based sceptical argument.

7. KNOWING AND SAYING THAT ONE KNOWS

This aspect of the envisaged neo-Moorean view—namely, the combination of content externalism and epistemic externalism to meet underdetermination-style sceptical arguments—is, of course, found in recent work by Timothy Williamson (2000a, 2000b), though it is not advanced under this description. Significantly for our purposes, however, this view is not set within the kind of diagnostic account that we saw above was essential to any plausible rendering of the neo-Moorean proposal. It is thus vital that we do not end the anti-sceptical story at this point, but continue to consider the further issue of diagnosis. What is key here, I believe, is understanding just why the Moorean assertions, while (on this view) true, are nevertheless conversationally inappropriate.

To begin with, note that the classical externalist is again on stronger ground than the classical internalist in this regard. This is because of the connections between the conditions for knowledge possession and the conditions for appropriate assertions of explicit knowledge claims (i.e. explicit self-ascriptions of knowledge of the form 'I know that *p*'). On the classical internalist view, there will be a close connection between these two conditions. This is because one ordinarily needs good reflectively accessible grounds in order to properly make an assertion—especially assertions which involve explicit self-ascriptions of knowledge—and the possession of grounds of this sort will also be a prerequisite for

default warrant for propositions of this sort is to be maintained on the classical internalist account, and similarly overcome the concern we raised for Wright above regarding how such a default conception of warrant could respond to the specific challenge posed by evidential scepticism, it would still remain that this type of anti-scepticism is not directly relevant for our purposes, where it is scepticism about knowledge that is our focus.

knowledge on the classical internalist account. On the classical externalist account, in contrast, the connection between these two conditions will be relatively weak in that knowledge might be possessed even in the absence of good reflectively accessible grounds, and thus be possessed in cases where the agent would not be in a position to properly assert the proposition that she knows. Think, for example, of the chicken sexer case described above. Here we have an agent who has knowledge by classical externalist lights, but who is clearly not in a position to properly assert what she knows—much less assert that she knows it—since she lacks any good reflectively accessible grounds to back up that assertion.

With this in mind, it is going to be a lot easier for the classical externalist to account for the apparent impropriety of Moorean assertions in such a way as to retain the thought that what Moore is asserting is nevertheless true. It could be, for example, that one's anti-sceptical knowledge that, say, one is not a BIV is like the 'brute' knowledge possessed in the chicken sexer case, and if this is so then this would explain why it cannot be properly asserted. What makes this claim plausible is the fact that, as just noted, what does seem to be a clear consequence of the sceptical reasoning is that we are unable to possess good reflectively accessible grounds for believing that we are not the victims of sceptical hypotheses. On this view, however, the lack of such grounds will only affect the propriety of one's assertions in this respect, and need not undermine one's anti-sceptical knowledge.

Further reflection on the peculiar role of 'I know' in our linguistic practices also supports this sort of contention. Typically, one conveys one's knowledge of a proposition simply by asserting the proposition in question. Adding the further phrase 'I know' is rare, and standardly reflects not just emphasis but also an ability to resolve a particular challenge that has been raised. For example, one might initially convey one's knowledge of what the time is by simply asserting, say, 'It's 10.22 a.m.', but then be prompted into the further explicit claim to know this proposition by a challenge to one's original assertion. There are two main ways in which these challenges could be issued. The first concerns the presentation of an error-possibility which is held to be salient. Call this an *epistemic challenge*. The second concerns those occasions where it is pointed out that a lot hangs on the correctness of the assertion in question. Call this a *standards challenge*.[17]

In responding to either of these types of challenge with an explicit knowledge claim one is representing oneself as being in possession of stronger reflectively accessible grounds in support of one's assertion than would be implied simply by making the assertion itself. Notice, however, that the kinds of additional grounds required in each case can be very different. When it comes to standards challenges, for example, all that is normally required is stronger grounds *simpliciter*. In contrast, when it comes to epistemic challenges, the additional grounds have

[17] One might have challenges which are a mix of the two, of course, but we will bracket this possibility here in order to keep matters as simple as possible.

to speak specifically to the error-possibilities raised. In particular—and this is a point which, I think, has often been overlooked in this regard, despite its epistemic importance—the grounds one needs available to one in making a claim to know in response to this second sort of challenge must be such as to *discriminate* between what is asserted and the relevant error-possibility. This claim is important because, intuitively, the kind of evidential support one needs in order to have knowledge is weaker than this. All that is required here is the sort of 'favouring' evidence that we saw above in our discussion of the underdetermination principle (which is still a strong requirement on knowledge, as we also saw).[18]

In order to make this point clear, consider the famous 'zebra' case offered by Dretske (1970). Here we are asked to imagine someone who is at the zoo in normal circumstances and sees what looks like a zebra in the zebra enclosure. Clearly, such an agent would normally be attributed knowledge that the creature before him is a zebra, and we would be perfectly happy with any assertion he might make to the effect that there is a zebra before him—which would represent him as having knowledge of this proposition—since he has adequate reflectively accessible grounds to back up that assertion. Similarly, an explicit claim to know that he sees a zebra would also be deemed appropriate in this context, if the circumstances were right. For example, if the original claim that the creature before him is a zebra is challenged in some mundane fashion—perhaps by someone short-sighted who wonders out loud why, since they were expecting to be near the gorilla enclosure, there should be zebras here—then it would be unproblematic for our agent to respond to this challenge by saying that he knows that this creature is a zebra.

It is important to recognize why such an assertion would be entirely appropriate, given how we have described the situation. The reason for this is not just that the agent is in a position to offer very good reflectively accessible grounds in favour of what he asserts, nor even that he has good reflectively accessible grounds in favour of what he asserts which prefer what he asserts over the target error-possibility (that it is a gorilla rather than a zebra), but more specifically that he has good reflectively accessible grounds which discriminate between the target proposition and the target error-possibility—i.e. between creatures that are zebras and (non-zebra) creatures that are gorillas. That is, explicitly claiming knowledge in this context will generate the conversational implicature that one is able to offer grounds in support of the proposition claimed as known which would suffice to distinguish the scenario described by this proposition from the specified error-possibility. In this case, however, the agent does have the required grounds. After all, our agent is aware, presumably, that zebras and gorillas have very different shapes and gaits, and this will suffice to enable such discrimination

[18] I think a failure to make this distinction is what makes contrastivism such an initially plausible conception of knowledge. For more on this point, see Pritchard forthcoming a.

to take place. That such a discriminative ability is required in order to claim knowledge in this case should not, however, lead the neo-Moorean into thinking that it is thereby required for knowledge possession, since, intuitively, mere favouring evidence will suffice in this regard.

This last point is important because there are cases where one has the required favouring evidence but where one lacks the relevant discriminative capacity. Accordingly, if one fails to pay due attention to this point then one will be led into denying knowledge to the agent even though on the neo-Moorean view it is possessed. Imagine, for example, that the error-possibility that the zebra may in fact be a cleverly disguised mule is raised and taken seriously in that conversational context. Since the original assertion that the creature is a zebra has been challenged, it would ordinarily be appropriate for the agent, if he knows this proposition, to explicitly say so, just as he did in the case just described where the objector wonders why he isn't at present looking at a gorilla. If the agent now claims to know that there is a zebra before him in the light of this error-possibility being raised, however, then this will generate the implicature not just that the agent has reflectively accessible grounds which prefer the proposition claimed as known over the target error-possibility, but also that the agent has reflectively accessible grounds which could serve to discriminate between the proposition claimed as known and the target error possibility. That is, the agent is representing himself as having grounds which would suffice to enable him to tell the difference between these two creatures (i.e. a zebra and a cleverly disguised mule). Such grounds might be, for example, that he has examined the creature at close range and been able to determine that it is not painted. Typically, of course, the agent will not have grounds of this sort available to him, and so his claim to know will be inappropriate because it generates a false conversational implicature.

Notice, however, that this fact alone does not suffice to indicate that the agent lacks knowledge of the target proposition. After all, in the standard case at least, the agent will have evidence which favours the hypothesis that the animal before him is a zebra over the alternative hypothesis that it is a cleverly disguised mule. Think, for example, of the grounds he has regarding the implausibility of a zookeeper going to such lengths to deceive patrons, and the penalties that would be imposed were such a deception to come to light, as presumably it would eventually. Moreover, this evidence will typically be reflectively accessible to the subject, and so this point stands alone from any general considerations regarding the relevance of the classical internalism/externalism distinction in epistemology here. Thus, by both externalist and internalist lights, the agent has evidentially grounded knowledge that there is a zebra before him. The issue is solely whether or not he can properly claim to possess that knowledge in these circumstances; not whether it is possessed.

What is different about these cases and the sceptical case is that we can make sense of an agent having evidence—even reflectively accessible evidence—which

favours the believed hypothesis over the alternative hypothesis. Moreover, we can also make sense of there being agents who are in better epistemic positions relative to the believed proposition—such as, in the zebra example, zoologists—who are able to possess the reflectively accessible discriminating grounds required for an appropriate knowledge claim. In sceptical cases, in contrast, matters are very different. This is because we have difficulty comprehending evidence that can play the required favouring role (it is essential that the evidence not be understood along classical internalist lines if it is to play this role); and we can make no sense at all of the idea that one has adequate reflectively accessible grounds which could serve to discriminate the target hypothesis from the sceptical alternative. On the neo-Moorean account sketched here, then, we can only account for knowledge possession in these cases relative to a classical externalist epistemology, and we can make no sense at all of appropriate knowledge claims in sceptical contexts.

Consider again the Moorean anti-sceptical assertions. If we grant that knowledge possession requires evidence which favours one's everyday beliefs over their sceptical alternatives, but also that such evidence can be possessed by externalist lights, then we are in a position to maintain that the Moorean assertions could well be true. Nevertheless, they cannot be properly made, since they are entered in a context in which sceptical alternatives are explicitly at issue, and thus they will generate false conversational implicatures. In particular, the claim that, for example, one knows one has two hands will generate the false conversational implicature that one has reflectively accessible grounds which would suffice to distinguish between the scenario in which one has hands and the alternative sceptical scenario in which one is, say, a BIV who merely seems to have hands.[19] Similarly, the claim to know that one is not a BIV will generate the false conversational implicature that one has reflectively accessible grounds which would suffice to indicate that one can distinguish between the scenario in which one is not a BIV from the alternative scenario in which one is a BIV. It is little wonder then, on this view, why such knowledge can never be properly claimed even in cases when it is possessed. Once one factors in the further consideration of the dialectical impropriety of these assertions, it becomes manifest why a neo-Moorean stance must not try to deal with the sceptical problem head-on by making anti-sceptical assertions in this way.

It is interesting to note that Williamson's (1996) own account of assertion—in terms of the overarching rule that one should only assert what one knows—does not seem to offer any diagnosis of what is wrong with the Moorean assertions. Presumably, this rule demands in this case that one should know that one knows the relevant propositions (i.e. E, ¬ BIV) if one is to properly assert them.

[19] This consideration also explains, at least in part, why such claims to know what is (taken to be) plainly obvious *always* seem problematic. This is because it is hard to imagine a non-sceptical context in which such an assertion would be entered.

Accordingly, one might think, on broadly epistemically externalist grounds, that the problem with the Moorean assertions is that while one has the knowledge in question (such that what is asserted is true), one lacks the corresponding second-order knowledge and it is this fact that ensures that the rule of assertion has been broken and thus that the assertion is inappropriate.

It is far from clear, however, why the neo-Moorean view as it is described here should ally itself with a move of this sort. After all, if one does know the relevant everyday and anti-sceptical propositions, then intuitively one's beliefs that one knows these propositions could well be just as safe as one's beliefs in the propositions themselves. Thus, there seems no reason, on this view, for denying the second-order knowledge in this case.[20] The trouble is, of course, that the assertion would, intuitively, still be inappropriate even if the second-order knowledge were possessed. If this is right, then on the face of it the Williamsonian view contains an important lacuna, since it cannot offer the required diagnostic story that distinguishes the neo-Moorean view from its problematic Moorean ancestor.

In any case, with the neo-Moorean picture of the conversational propriety of explicit self-ascriptions of knowledge just set out in mind, it is hardly surprising that there will be a context-sensitivity in the propriety conditions for making such assertions. One type of context-sensitivity is that just considered regarding claims to know in response to epistemic challenges. In these cases, the evidential demands on appropriate assertion shift in response to features of the conversational context, such that the very same assertion can be appropriate in one context and yet inappropriate in another context even though all that has changed has been the introduction of an error-possibility to that conversational context. A similar sort of context-sensitivity is also in play when it comes to standards challenges, though, as noted above, the type of evidential demand made by a standards challenge is usually subtly different in that responding to such a challenge merely demands of the asserter a greater degree of reflectively accessible evidence, rather than reflectively accessible evidence which can specifically serve to discriminate between the asserted proposition and the target error-possibility. Repeating a claim to know in response to someone simply pointing out how important it is that one is right will typically involve merely representing oneself as having very strong reflectively accessible grounds in favour of the proposition claimed as known; it will not usually involve representing oneself as having specific discriminating grounds because there is no error-possibility at issue for which such grounds would be relevant in this case.

[20] This is especially so on the Williamsonian picture because of its commitment to the thesis that one's knowledge is identical to one's evidence (see Williamson 2000a). Accordingly, simply in virtue of having the first-order knowledge one thereby has excellent evidence to support any second-order belief that one might hold to the effect that one has this knowledge, and thus one is by default well on the way to having second-order knowledge as a result.

Contextualists often talk as if their focus is simply standards challenges, but it should be clear that it is epistemic challenges that are really what is at issue here. In any case, what is important is that this pragmatic account of the shifting propriety conditions for explicit knowledge claims removes much of the impetus for contextualism, in that one can accommodate the apparent context-sensitivity of our 'knows' talk without thereby treating 'knows' as a context-sensitive term. More needs to be done to complete the pragmatic story of course, since one needs to extend the view so that it deals with explicit knowledge ascriptions more generally, rather than just explicit self-ascriptions of knowledge, but the beginnings of such a view are clear to see.[21]

Furthermore, such a pragmatic story also goes a long to way towards offering the kind of diagnostic support for neo-Mooreanism that we noted above was so lacking, especially in light of the strong diagnostic support that contextualists can offer in favour of their view. In particular, one can now account for the impropriety of the Moorean assertions, and thereby explain the intuitive pull behind sceptical arguments, without conceding anything substantial to the sceptic. With this diagnostic story in place, the remaining theoretical advantage held by the contextualist over the neo-Moorean is finally undermined.

8. McDOWELLIAN NEO-MOOREANISM

Interestingly, there is a very different neo-Moorean position that has been proposed in the literature which deserves our consideration, one that does not comfortably fit into either of the classical internalist or classical externalist camps. This position is due to John McDowell (e.g. 1995), and what is distinctive about it (amongst other things) is that it incorporates two theses which, collectively, set the view apart from classical versions of epistemic externalism and internalism. These theses are: (i) a claim in the spirit of epistemic internalism which demands of a knower that she be in a position to know by reflection alone what the reasons which support her knowledge are; and (ii) a content externalist claim of the disjunctivist sort noted above which allows that one's reasons can be both empirical and factive—i.e. can be reasons for believing an empirical proposition and entail what it is that they are a reason for. This last claim entails content externalism since on McDowell's view one's experiences can function as one's reasons, and yet the content of one's experiences will clearly be sometimes—i.e. in those cases where the reasons are factive—determined by facts concerning one's environment (such that in corresponding cases of illusion or delusion where the fact does not obtain the content of the experience will be different, and hence the target reason will be absent). These two theses, when conjoined,

[21] I offer the extended account in Pritchard 2005a.

pose a problem for the classical way of understanding the internalism/externalism debate because it is standardly thought that what one has reflective access to cannot extend beyond the 'inner' to take in factive empirical reasons in this way. According to McDowell, however, this conventional wisdom of contemporary philosophy is false and leads to a philosophical picture which invites the sceptical challenge. Moreover, the source of the problem with this conventional wisdom, according to McDowell, is the failure to endorse the kind of content externalism that he has in mind.

For example, for McDowell one's reason for believing an empirical perception-based proposition—say, that John is in my office—could simply be the factive empirical reason that one *sees that* John is in my office. Moreover, since one's reasons are reflectively available to one (since otherwise they would not be one's reasons at all), it follows that one has reflective access to this factive empirical reason. That is, one's reason for believing that John is in one's office could be that one sees that he is in one's office, where one is able to know by reflection alone that one possesses this reason.

We saw above that there is an everyday conception of evidence according to which we can legitimately cite factive grounds in favour of our empirical beliefs in this way. According to the classical internalist, however, this everyday practice of offering factive grounds cannot be taken at face value. Instead, the classical internalist will claim that the evidential grounding of one's belief must be, strictly speaking, the non-factive counterpart of the factive claim, such as that *it seems to one as if* John is in one's office. Part of the reason for this restriction, presumably, is that it is only the non-factive counterpart that could be reflectively accessible to the subject. Any straightforward accommodation of this everyday conception of evidence is thus automatically placed in the classical externalist camp. It is this classical internalist orthodoxy that McDowell's view challenges.

On the face of it, the McDowellian line is susceptible to a straightforward problem, one that mirrors the famous 'McKinsey' puzzle that concerns the supposed compatibility of first-person authority and content externalism.[22] For it seems that one can use one's reflective access to one's factive empirical reasons, along with one's reflective knowledge of the relevant entailment (i.e. from the factive empirical reason to the empirical fact), to acquire reflective, and thus non-empirical, knowledge of the empirical proposition which the reason is a reason for (in this case that John is in one's office). Intuitively, this is just a *reductio* of the view, and this in part explains, I think, why very few commentators have taken McDowell seriously on this point and have stuck, instead, to the conventional wisdom on reflective access to reasons that is implicit in the classical way of drawing the externalism/internalism contrast in epistemology.[23]

[22] For more on this problem, see Nuccetelli 2003.

[23] For example, Greco 2004 takes the factivity of reasons claim seriously, but doesn't take the internalist reflective access claim seriously as a result, and therefore regards McDowell's view

It is not obvious, however, that McDowell's position is subject to a problem of this sort. The reason for this is that the difficulty only emerges if one *acquires* non-empirical knowledge by running through an inference of this type, and it is far from clear why, on the McDowellian view, this would be so. In particular, the McDowellian conception of perceptual knowledge in these cases is such that one could only be in a position to run through such an inference provided one already has the empirical knowledge in question, and thus the problem concerning the non-empirical acquisition of empirical knowledge does not arise.

In order to see this, one only needs to note that for McDowell a factive reason for *p* and a relevant belief that *p* which is based on that factive reason will suffice for knowledge that *p*. Note, however, that reflective accessibility is itself factive, in that if one is in a position to reflectively access that one has a factive reason for believing *p* then it must be the case that one has that factive reason for believing *p*. Furthermore, even though the possession of the factive reason is consistent with non-belief in the target proposition on the McDowellian view, it is obviously not going to be possible for someone to run through the reasoning described above without in the process acquiring the reason-based belief in the target proposition, thereby meeting all the conditions required on this view for empirical knowledge in this case. Thus, given the further trivial claim that if one has empirical knowledge of *p* then one cannot also have reflective (i.e. non-empirical) knowledge of *p*, it follows that there is no prospect of acquiring non-empirical knowledge in this case, and thus McDowell's McKinsey-style difficulty disappears.

On the face of it, then, there is a position available here, one that retains that aspect of the classical internalist thesis that insists on reflectively accessible grounds in favour of one's beliefs if one is to count as a knower, but which also allows that one's grounds, so construed, can be robust enough to meet the underdetermination-based sceptical argument. Accordingly, one could combine the contextualist pragmatic thesis outlined above with this line of argument to derive an internalist formulation of the view which does not concede that our knowledge of these anti-sceptical propositions is necessarily brute. Instead, one can have adequate reflectively accessible grounds in favour of such beliefs, such as one's factive perception-based reasons which entail the denials of sceptical hypotheses.

This view clearly merits further exploration.[24] Notice, however, that while it changes the shape of the dialectic here by opening up a new direction of research,

as simply being a version of classical epistemic externalism. Wright 2002, in contrast, takes the internalist reflective access claim seriously but as a consequence does not take the factivity claim seriously. On his reading, what one can have reflective access to is not the factive empirical reason but rather a 'disjunctive' reason—i.e. a reason for believing that either one is in the factive state or one is not in the factive state and deluded in some way.

[24] For more detailed discussion of the view, see Neta and Pritchard forthcoming and Pritchard forthcoming b.

it does not make any substantial difference to the key points raised earlier. For while it is now true on this view that one does have reflectively accessible grounds in favour of one's anti-sceptical beliefs, this does not mean that one is in a position to, for example, properly cite them in a sceptical conversational context in which we are taking sceptical error-possibilities seriously. Even if it is true that one possesses a factive empirical reason which entails the denial of, say, the BIV sceptical hypothesis, it can hardly be thought that such a reason—that I see that I have two hands, for example—would represent grounds which speak to this particular contrast, and yet this assertion in this context will certainly generate this conversational implicature. We have seen above just why this is so, since to explicitly claim knowledge in a sceptical context is to represent oneself as having reflectively accessible evidence which *discriminates* between the proposition asserted as known and sceptical alternatives. Crucially, however, one's reflectively accessible factive empirical reasons will not serve *this* role.[25]

I think this is part of the reason why McDowell—at least in his more careful remarks on this subject—only claims to be showing how one can legitimately *ignore* the sceptical argument, rather than claiming to have actually *responded* to it. The thought is that once one recognizes that one's reflectively accessible grounds can be factive then one should not feel the pressure to conversationally engage with the sceptical problem with anti-sceptical assertions any more, not that one thereby has a response which should silent even the sceptic (which is what the Moorean strategy seems to aspire to). (I take it that the classical externalist neo-Moorean would be inclined to agree with this claim, and thus on this score at least the two views are very similar.) Thus, the McDowellian line is very much in the spirit of neo-Mooreanism.

I here leave it open whether one should be a McDowellian neo-Moorean or simply a classical externalist neo-Moorean, since either will suffice to meet the sceptical problem. I do, however, want to make one small remark on this issue before I close, which points to how we should go about exploring this issue. This is that on the face of it the McDowellian view is plausibly in a better position to be counted as the true heir to the Moorean tradition on account of how it retains core internalist intuitions. After all, part of the desiderata of neo-Moorean positions is that they are able, where possible, to accommodate our pre-theoretical intuitions, and internalist intuitions are surely highly embedded within folk epistemology. Moreover, since both views appear to endorse some form of disjunctivism, it seems that one could similarly argue that what the McDowellian picture highlights is how disjunctivism, properly understood, enables one to evade the sceptical problematic *without* having to resort to the revisionism of classical epistemic externalism. Such fine-grained issues of which sort of neo-Mooreanism one should endorse once one has dealt with the sceptical problem can, however, be left for another occasion.

[25] For more on this point, see Pritchard 2003.

9. CONCLUDING REMARKS

We have thus seen that there is a viable neo-Moorean view available which can deal with the sceptical problem in a more satisfactory way than other competing anti-sceptical theories, and which avoids the problems facing Mooreanism. In particular, it is an anti-luck epistemology that retains closure while avoiding a contextualization of our understanding of 'knows', but which also accommodates contextualist intuitions by incorporating a view about the context-sensitivity of the propriety conditions for explicit knowledge self-ascriptions. Furthermore, this pragmatic account also serves to demarcate the neo-Moorean view from its Moorean ancestor, since it explains, in part, why the Moorean assertions were problematic (and thus why scepticism can seem so plausible) even though they were true. Finally, we have seen that one can resolve the evidential sceptical problem without resorting to an evidential form of contextualism by adopting a version of content externalism and either opting for a classical externalist epistemology or a McDowellian version of epistemic internalism.

There is one further element of the neo-Moorean account that I think is required, though I will not explore it at length here. This is that the neo-Moorean would be wise, I think, to concede something to the sceptic; to say that there is something right about the sceptical challenge. The key to this concession lies in the fact that the neo-Moorean anti-sceptical response is in a certain way necessarily *mute*, since it is part of the view that one cannot properly respond to the sceptical challenge by repeating one's everyday claims to know. I think there is a deep point here about the limits of our cognitive responsibility, a point that I have explored at length elsewhere.[26] Making such a concession to the sceptic does not undermine the view so long as one steadfastly retains the core claim that knowledge is nonetheless possessed in such cases, so that the epistemic lack at issue here, if that is the right way to characterize it, is not an epistemic lack that would undermine knowledge. Such a concession would thus only strengthen the view by accounting for our visceral attraction to sceptical arguments, even despite our strong anti-sceptical intuitions. The point would be that there is a deep truth in sceptical arguments, though not the deep truth that the sceptic advertises. Neo-Mooreanism, while obviously a particularly robust anti-sceptical theory, need not be a view that dismisses the sceptical problem entirely.

REFERENCES

Alston, W. 1988: 'An Internalist Externalism,' *Synthese* 74: 265–83.

Baldwin, T. 1993: 'G. E. Moore.' In J. Dancy and E. Sosa (eds.), *A Companion to Epistemology* (Oxford: Blackwell), 283–5.

[26] See, for example, Pritchard 2005b, *passim*; 2005c.

Brandom, R. 1998: 'Insights and Blindspots of Reliabilism.' *The Monist* 81: 371–92.
Cohen, S. 2000: 'Contextualism and Skepticism.' *Philosophical Issues* 10: 94–107.
Conee, E. and Feldman, R. 2000: 'Internalism Defended.' In H. Kornblith (ed.), *Epistemology: Internalism and Externalism* (Oxford: Blackwell), 231–60.
DeRose, K. 1995: 'Solving the Skeptical Problem.' *Philosophical Review* 104: 1–52.
Dretske, F. 1970: 'Epistemic Operators.' *Journal of Philosophy* 67: 1007–23.
____ 2005a: 'The Case against Closure.' In M. Steup and E. Sosa (eds.), *Contemporary Debates in Epistemology* (Oxford: Blackwell), 13–26.
____ 2005b: 'Reply to Hawthorne.' In M. Steup and E. Sosa (eds.), *Contemporary Debates in Epistemology* (Oxford: Blackwell), 43–6.
Foley, R. 1987: *A Theory of Epistemic Rationality* (Cambridge, Mass.: Harvard University Press).
Greco, J. 2004: 'Externalism and Skepticism.' In R. Shantz (ed.), *The Externalist Challenge: New Studies on Cognition and Intentionality* (New York, de Gruyter).
Hawthorne, J. 2004: *Knowledge and Lotteries* (Oxford: Oxford University Press).
____ 2005: 'The Case for Closure.' In M. Steup and E. Sosa (eds.), *Contemporary Debates in Epistemology* (Oxford: Blackwell), 26–43.
Heller, M. 1999: 'Relevant Alternatives and Closure.' *Australasian Journal of Philosophy* 77: 196–208.
Kornblith, H. (ed.) 2001: *Epistemology: Internalism and Externalism* (Oxford: Blackwell).
Lewis, D. 1996: 'Elusive Knowledge.' *Australasian Journal of Philosophy* 74: 549–67.
McDowell, J. 1995: 'Knowledge and the Internal.' *Philosophy and Phenomenological Research* 55: 877–93.
Moore, G. E. 1925: 'A Defence of Common Sense.' In J. H. Muirhead (ed.), *Contemporary British Philosophy* (2nd series) (London: Allen and Unwin).
____ 1939: 'Proof of an External World.' *Proceedings of the British Academy* 25: 273–300.
Neta, R. 2002: 'S Knows that P.' *Noûs* 36: 663–81.
____ 2003: 'Contextualism and the Problem of the External World.' *Philosophy and Phenomenological Research* 66: 1–32.
____ and Pritchard, D. H. forthcoming: 'McDowell and the New Evil Genius.' *Philosophy and Phenomenological Research*.
Nozick, R. 1981: *Philosophical Explanations* (Oxford: Oxford University Press).
Nuccetelli, S. (ed.) 2003: *New Essays on Semantic Externalism and Self-Knowledge* (Cambridge, Mass.: MIT Press).
Pritchard, D. H. 2000: 'Closure and Context.' *Australasian Journal of Philosophy* 78: 275–80.
____ 2002a: 'McKinsey Paradoxes, Radical Scepticism, and the Transmission of Knowledge across Known Entailments.' *Synthese* 130: 279–302.
____ 2002b: 'Radical Scepticism, Epistemological Externalism and Closure.' *Theoria* 69: 129–61.
____ 2002c: 'Recent Work on Radical Skepticism.' *American Philosophical Quarterly* 39: 215–57.
____ 2002d: 'Resurrecting the Moorean Response to Scepticism.' *International Journal of Philosophical Studies* 10: 283–307.
____ 2003: 'McDowell on Reasons, Externalism and Scepticism.' *European Journal of Philosophy* 11: 273–94.

——— 2004: 'Epistemic Luck.' *Journal of Philosophical Research* 29: 193–222.
——— 2005a: 'Contextualism, Scepticism and Warranted Assertibility Manoeuvres.' In J. Keim-Campbell, M. O'Rourke, and H. Silverstein (eds.), *Knowledge and Skepticism* (Cambridge, Mass.: MIT Press).
——— 2005b: *Epistemic Luck* (Oxford: Oxford University Press).
——— 2005c: 'Scepticism, Epistemic Luck and Epistemic *Angst*.' *Australasian Journal of Philosophy* 83: 185–206.
——— 2005d: 'The Structure of Skeptical Arguments.' *Philosophical Quarterly* 55: 37–52.
——— 2005e: 'Wittgenstein's *On Certainty* and Contemporary Anti-Scepticism'. In D. Moyal-Sharrock and W. H. Brenner (eds.), *Investigating On Certainty: Essays on Wittgenstein's Last Work* (London: Palgrave Macmillan).
——— 2007: 'Knowledge, Luck, and Lotteries.' In V. F. Hendricks and D. H. Pritchard (eds.), *New Waves in Epistemology* (Aldershot: Ashgate).
——— forthcoming a: 'Contrastivism, Evidence, and Scepticism.' *Social Epistemology* (special issue on *Contrastivism*), 20.
——— forthcoming b: 'McDowellian Neo-Mooreanism.' In A. Haddock and F. Macpherson (eds.), *Disjunctivism: Perception, Action, Knowledge* (Oxford: Oxford University Press).
Pryor, J. 2000: 'The Skeptic and the Dogmatist.' *Noûs* 34: 517–49.
——— 2001: 'Highlights of Recent Epistemology.' *British Journal for the Philosophy of Science* 52: 95–124.
Schaffer, J. 2004: 'From Contextualism to Contrastivism in Epistemology.' *Philosophical Studies* 119: 73–103.
——— 2005a: 'Closure, Contrast, and Answer.' Manuscript.
——— 2005b: 'Contrastive Knowledge.' In T. Gendler and J. Hawthorne (eds.), *Oxford Studies in Epistemology* (Oxford: Oxford University Press).
Sosa, E. 1999: 'How to Defeat Opposition to Moore.' *Philosophical Perspectives* 13: 141–54.
Stanley, J. 2004: 'On the Linguistic Basis for Contextualism.' *Philosophical Studies* 119: 119–46.
——— 2005: *Knowledge and Practical Interests* (Oxford: Oxford University Press).
Williamson, T. 1996: 'Knowing and Asserting.' *Philosophical Review* 105: 489–523.
——— 2000a: 'Scepticism and Evidence.' *Philosophy and Phenomenological Research* 60: 613–28.
——— 2000b: *Knowledge and its Limits* (Oxford: Oxford University Press).
Wittgenstein, L. 1969: *On Certainty*, ed. G. E. M. Anscombe and G. H. von Wright, tr. D. Paul and G. E. M. Anscombe (Oxford: Blackwell).
Wright, C. 1991: 'Scepticism and Dreaming: Imploding the Demon.' *Mind* 397: 87–115.
——— 2002: '(Anti-)Skeptics Simple and Subtle: G. E. Moore and John McDowell.' *Philosophy and Phenomenological Research* 65: 330–48.
——— 2004: 'Warrant for Nothing (and Foundations for Free)?' *Proceedings of the Aristotelian Society* 78 (suppl. vol.): 167–212.
Zagzebski, L. 1996: *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge* (Cambridge: Cambridge University Press).

5

Some Ins and Outs of Transglobal Reliabilism

David Henderson and Terry Horgan

1. INTRODUCTION

Our central concern in this chapter is epistemological. We have recently developed an account of what it is for an agent to be objectively justified in holding some belief. We call it *transglobal reliabilism*. Here, we present this view in outline and we show how it is motivated by various thought experiments. In a later section, we discuss how to understand transglobal reliabilism with respect to epistemological internalism and externalism. This matter warrants special attention, for while transglobal reliabilism is a variant of externalist epistemology, it also successfully accommodates certain concerns associated with internalist epistemologies.

One, but only one, of the motivating scenarios discussed here presupposes the falsity of what we will call *strong semantic externalism*: roughly and generically, the view that (1) all intentionality is grounded in certain past or present causal connections between states of the cognitive system and states of the external world, and (2) there can be no mental intentionality without some suitable kind of actual connection between what is going on 'in the head' and the wider environment. The motivating scenario, which involves a brain in a vat, is a version of what Sosa (1991) has dubbed 'the new evil demon problem' for reliabilism. The scenario as envisioned is intelligible only if strong semantic externalism is mistaken.

Transglobal reliabilism can be motivated by thought experiments that are intelligible even given strong semantic externalism—and we will set forth a number of such thought experiments here. There is ample reason to embrace transglobal realism, we maintain, even if one accepts some version of strong semantic externalism. Still, radical-deception scenarios involving envatted brains or Cartesian evil demons, if intelligible as usually construed, provide particularly compelling motivation for transglobal reliabilism. This fact, combined with the fact that we ourselves believe that such scenarios are intelligible (and thus that

strong semantic externalism is mistaken), is why our overall case for transglobal reliabilism rests in part on its capacity to handle the new evil demon problem in a natural and intuitively plausible way.[1]

Transglobal reliabilism is a species of reliabilism. It holds that it is necessary for being objectively justified that the belief in question be fixed by way of processes that are reliable—where reliability, as always, must be understood as relative to some reference class of environments. Transglobal reliabilism is largely distinguished by its conception of the relevant reference class of environments. For the most familiar versions of reliabilism, the relevant form of reliability is taken to be reliability relative to the agent's own global environment—global reliability. For transglobal reliabilism, on the other hand, the relevant form of reliability is reliability relative to the set of *experientially possible global environments—transglobal reliability*. A possible global environment is *experientially possible* just in case it is compatible with one's there having experiences of roughly the character of those that agents actually have.[2] The actual global environment is but one among a diversity of experientially possible global environments. Some experientially possible global environments would be extremely epistemically inhospitable—there would be few if any globally reliable processes to be had there. Demon-infested global environments, and those featuring agents as envatted brains, would be cases in point—provided one is enough of a semantic internalist to believe that such scenarios make sense. For the semantic externalist, while there still would be epistemically inhospitable global environments, the spectrum of such possible environments would not include such epistemically inhospitable extremes as demon-infested, or brain-envatting, environments. The actual global environment (which, we take it, is demon free) would seem to be a relatively moderate experientially possible global environment; there would seem to be both more and less hospitable global environments. In the yet more hospitable experientially possible global environments, there would be fewer sources of error which need be guarded against than there are in the actual global environment. Transglobal reliabilism, in a somewhat simplified nutshell, is the view that it is constitutively required for objective epistemic justification that a

[1] Philosophers like Kripke, Putnam and Burge are surely right in maintaining that *some* aspects of mental intentionality are constitutively dependent at least partly on externalistic linkages between the cognitive agent and the agent's wider environment—in particular, that such linkages figure constitutively in determining the referents (if any) of thought constituents that purport to refer to individuals or to natural kinds. But it bears emphasis that this comparatively mild form of semantic externalism does not begin to entail *strong* semantic externalism. For articulation and defense of a position that accommodates mild semantic externalism—while also wedding it to the claim that the most fundamental kind of mental intentionality (i) is narrow, (ii) is phenomenally constituted, and (iii) is shared in common between oneself, one's Twin-Earthly duplicate, and one's brain-in-vat duplicate—see Horgan and Tienson 2002 and Horgan, Tienson, and Graham 2004.

[2] Note well: An experientially possible global environment can be compatible with one's having such experiences within it even if these experiences are radically and systematically nonveridical.

belief be the product of processes that are transglobally reliable—reliable relative to the class of experientially possible global environments.[3]

In order for a process to be transglobally reliable, the process need not be globally reliable in *all* experientially possible global environments. Rather, and again somewhat crudely, what is required is that the belief-fixing process be reliable in a wide range of such global environments. Consider the parallel matter involving what might be termed *local* versus global reliability. Local reliability is reliability relative to some local environment that an agent might encounter within that agent's actual global environment. Within a global environment, there typically are relatively epistemically hostile and inhospitable local environments. Were there a fake-barn county within the agent's global environment, it would constitute an epistemically inhospitable local environment *vis-à-vis* the formation of perceptually based beliefs about barnhood. Obviously, within the actual global environment, there are local environments that are epistemically relatively inhospitable (thus, the fog of war, the fog of American politics, the fog of fundamentalism, the fog of hypoxia at altitude in a white-out, and so on). There are also relatively hospitable local environments (presumably one's own kitchen is relatively hospitable for perception of everyday objects, with its lack of camouflage, with its paucity of fakes, with the prevalence of familiar objects, and with the presence of oxygen). Now, importantly, a process can be globally reliable while failing to be locally reliable in some inhospitable local environments. Similarly, a process can be transglobally reliable while failing to be globally reliable in some experientially possible global environments.

2. A SERIES OF THOUGHT EXPERIMENTS: THE ROAD TO TRANSGLOBAL RELIABILISM

We now want to explore the motivations for transglobal reliabilism. We will do so by considering a series of thought experiments—and reflecting on the concern for a kind of epistemic safety which, these thought experiments suggest, is central to the concept of being justified in believing.

The motivation of transglobal reliabilism to be presented here proceeds in two stages. First, the thought experiments considered in section 2.1 reveal a pivotal concern for epistemic safety that figures in the notion of objective epistemic justification—a concern served by employing processes with robust forms of reliability. Second, in section 2.2, as further thought experiments are considered, it will become clear that the indicated concern for epistemic safety can best be accommodated by an account of objective epistemic justification that gives a central place to transglobal reliability. Thus, the thought experiments considered

[3] This simplified formulation will receive some refinement below.

in section 2.1 ultimately (when considered along with those to be considered in s.2.2) provide motivation for transglobal reliabilism.

First, however, we explain how the thought experiments in section 2.1 lend support to a refinement of more familiar forms of reliabilism—a refinement we call *neoclassical* reliabilism. The dialectical reason to proceed in this manner is that, ultimately, considerations much like those that initially motivate refining classical forms of reliabilism into the neoclassical version will also motivate going still further—namely, eschewing both classical and neoclassical reliablism in favor of transglobal reliabilism.

2.1. Classical and Neoclassical (or Global) Reliabilism

A generic form of reliabilism has become fairly standard. It emphasizes reliability within the global environment that the epistemic agent *actually occupies*—as opposed to various non-actual global environments such as that of an envatted brain. Adherents of this general position—call it *classical reliabilism*—have commonly not distinguished sharply, or not made much of, the distinction between local reliability and global reliability. In this section we consider scenarios suggesting that, were the choice between just these two forms of reliability, then reliabilist accounts of objective justification would better focus on global reliability. We call the refined reliabilism thus suggested *neoclassical reliabilism* or *global reliabilism.*

a. Athena and Fortuna

Suppose that Athena and Fortuna are driving from New York City to Memphis. In rural West Virginia, they drive through a county in which there happen to be numerous extremely realistic-looking fake barns within view of the highway—although neither of them has any inkling of this fact or any reason to suspect it. As it happens, in this local area all the real barns are yellow, and none of the fake barns or any other buildings are yellow. Again, they have no information to this effect. As they drive past a saliently presented yellow building, Athena, who has had reasonable experience with barns, gets a clear look at it, and on the basis of its barn-like visual appearance, she judges it to be a barn.

Fortuna gets only a very brief glimpse of the building. She saw her first barn just yesterday, elsewhere, and it happened to be yellow. She judges, on the basis of the briefly glimpsed building's yellow color, that it is a barn—even though she did not get a good look at it, and was thus unable to discern any features that are generally distinctive of barns as opposed to other kinds of buildings. It's not that she has a *general* belief that all and only yellow buildings are barns, or that all barns are yellow; she has never formed any such belief. Also, it's not that she has a *general* tendency to inductively extrapolate from old cases to new ones in a hastily generalizing way, a tendency she might otherwise be exhibiting here. Rather, it just happens that *in the present circumstances*, a psychological process is

present within her that takes as input both the brief glimpse of a yellow building and the yellow-barn memory, and generates as output the barn-belief about the briefly glimpsed object.

So Athena and Fortuna each form the belief that the building is a barn. And indeed it is.

First consider Athena. One's strong inclination is to say that her belief that the building is a barn is extremely well justified. After all, she has excellent perceptual evidence for the belief, and she formed the belief using a process of perceptual barn-categorization that generally works extremely well. On the other hand, as is commonly stressed in discussions of fake-barn scenarios, one is also strongly inclined to say that she does not *know* that the building is a barn, because her belief has been produced by a belief-forming process that happens to be *unreliable in this specific local environment*. Given all those fake barns around, any of which she would have mistakenly taken to be a real barn, the truth of her belief is too much a matter of epistemic luck to count as knowledge.

Now consider Fortuna. Here one has a very strong inclination to say that she *lacks* objective justification for her belief that the building glimpsed is a barn. This seems so despite the fact that this belief was produced by a belief-forming process that happens to be reliable in the local vicinity she then occupies. The trouble is that *this reliability is itself too fortuitous*, too much a matter of epistemic luck, for the belief to count as justified. And as a consequence, the truth of her belief is also too much a matter of luck for the belief to count as *knowledge*—the reliability of the belief-forming process notwithstanding.

b. Local Reliability, Global Reliability, and the Fake-Barn Scenario

If one takes these intuitive verdicts to be correct, would this mean that this a case in which one person's belief (Athena's) is justified even though it is *not* produced by a reliable belief-forming process, and in which another person's belief (Fortuna's) fails to be justified even though it *is* produced by a reliable belief-forming process? That would be too flat-footed a moral to draw. The reliability of the relevant processes is not such a simple matter—one cannot flatly say that the relevant process is reliable, or that it is not reliable.

Athena employs a barn-categorizing belief-forming process that is indeed highly reliable in one respect: it is globally reliable. Recall that she has had reasonable experience with barns, and it is plausible to suppose that she has thereby come to have a reliable perceptual ability to discriminate barns from non-barns. (In the world at large, there are few fake barns. Humans with reasonable experience typically do become reliable perceptual judges with respect to such generally non-tricky perceptual matters involving common middle-sized physical objects.) But her barn-categorizing process is locally unreliable, because of all those fake barns in the vicinity. Intuitively, the strong objective justification her belief possesses does indeed involve production by a reliable belief-forming process—namely, a globally reliable one. On the other hand, intuitively, the fact that this strongly

justified belief nonetheless fails to qualify as knowledge also involves a failure with respect to reliability of the belief-forming process—namely, a failure to be locally reliable. Both dimensions of reliability thus figure importantly: the presence of global reliability seems to figure in her belief being objectively well justified, whereas the absence of local reliability seems to figure in that belief nevertheless failing to qualify as a case of knowledge.

Fortuna is a converse case. Her belief is produced by a barn-categorization process that is locally reliable (for reasons beyond her cognitive ken) but is globally unreliable. One judges that Fortuna's belief fails to be well justified. Intuitively, what appears to figure importantly in its lacking objective justification is the fact that the belief-forming process is not globally reliable. Given this lack of global reliability, the fact that the process happens to be locally reliable strikes one intuitively as a *lucky accident*, epistemically speaking; and for this reason, the belief also does not count as a case of knowledge.

c. Global Reliability as a Robust Disposition

Call a disposition *robust* if it obtains relative to a fairly wide reference class of potential circumstances, situations, or environments. A robust disposition is one whose possession does not depend heavily upon certain unusual or atypical features that are highly specific to the particular circumstance or environment which the possessor of the disposition might happen to occupy; i.e. the disposition does not obtain only relative to a narrow reference class of environments in which those particular features happen to be present.

Now consider reliability, with robustness in mind. The tendency to produce (mostly) true beliefs must be understood as relative to a reference class of actual or possible environments. If a process is globally reliable, it has this tendency with respect to the wide reference class comprising the potential local environments to which an agent might be exposed (or which the agent might inhabit) within that agent's global environment. This is to have reliability in a reasonably robust fashion. For a process to be globally reliable is for its reliability not to depend heavily upon certain unusual or atypical features that are highly specific to the particular circumstance or environment which the possessor of the disposition might happen to occupy; its reliability then does not obtain only relative to a narrow reference class of environments in which those particular features happen to be present. This said, a process can be globally reliable while failing to be locally reliable with respect to some local environment afforded by the global environment (as illustrated by Athena's processes), and a process can be locally reliable without being globally reliable (as illustrated by Fortuna's processes). When a process is locally, but not globally, reliable, that process's reliability does depend heavily upon certain unusual or atypical features that are highly specific to the particular circumstance or environment which the possessor of the disposition occupies. Such merely local reliability is non-robust, because it

involves a narrow reference class: local environments with highly specific features exhibited by the particular local environment that an agent happens to be in.

Compare automobiles. An old car with its somewhat compromised cooling system, worn tires, and the like might well be reliable relative to a narrow range of potential environmental circumstances, but not relative to a wide range. It might be reliable in a local environment where the climate is temperate, it rains moderately and seldom snows, there are no long steep hills, and there is very little traffic. But it might be unreliable in local environments where there are temperature extremes, or lots of traffic, or demanding hills, or significant amounts of rain, snow and ice, etc. Although the car might be in service in an environment in which it happens to be locally reliable, it is not globally reliable. It thus is not *robustly* reliable. Its reliability depends heavily on the somewhat unusual combination of features particular to the environment where it happens to be employed. On the other hand, there are fine new cars that qualify as globally reliable—that is, would stand up well to the demands of providing daily transportation with respect to the representative environments to which an auto might be exposed or put to work. Of course, such cars would fail in a few local environments—for example, those characterized by significant flooding, high electromagnetic surges (associated with nuclear explosions) or volcanic flows. But, with respect to the global environment taken as a whole, as a composite of the range of potential local environments presented within the global environment, the relevant cars would tend to provide daily transportation on demand. Although a robustly reliable item—a car, a belief-forming process, or whatever—is reliable in a *wide range* of the potential circumstances within the reference class of the item's reliability, there can be *exceptional* local circumstances within which the item fails to be locally reliable. This is a ubiquitous feature of robust dispositions.

Here is a way of thinking about or intuitively gauging the robustness of the reliability of a process or item, a way of doing so that will prove useful as we proceed. Start with a process or item that is reliable in a given environment (for now, think of a particular local environment). Then think of varying the environment in various ways. To continue the automotive example, compare two vehicles: the old car mentioned above and a new car of a make that would be highly rated in measures like the *Consumer Reports* survey of automotive reliability. (Because respondents to that survey are widely distributed geographically, it may be taken to measure global reliability.) Both vehicles are locally reliable in the benign environment described above. So, begin varying that environment. As one considers variations in temperatures, across the range of temperatures that typically occur in the global environment, one finds that the older vehicle quickly ceases to be reliable, while the newer vehicle remains reliable through more variation. Similar results are obtained when considering variations in precipitation type and amount, or variations in the length or degree of inclines to be managed. The vehicle that retains its reliability across greater variations in environmental conditions is more robustly reliable. As we noted, robustness of

reliability is a matter of not being highly dependent on particular environmental conditions—a matter of accommodating more rather than less environmental variation.

d. Epistemic Safety: a Constitutive Requirement for Justification

Return again to Fortuna, and consider what seems intuitively objectionable about her belief that the yellow structure she fleetingly glimpsed is a barn. The problem is that, although the process that generated this belief does happen to be locally reliable, its being so is too much a matter of epistemic *luck* for her belief to be justified. Adopting the belief in the way she did—on the basis of a brief and fleeting glance in which the only salient feature she noticed was the structure's color, and in a way that arises somehow out of the fact that the one barn she has previously seen was yellow—is objectionably risky, epistemologically speaking. It is not epistemically *safe*. Somehow Fortuna has deployed a process that is not at all globally reliable, and rather accidentally or coincidentally, she just happened to do so in a rather special local environment in which that process was locally reliable. This seems an extraordinary lucky coincidence, for there is nothing about Fortuna that in any way 'tracks' the features of the local environment in virtue of which her yellow-building-triggered barn-identification-generating process is locally reliable. It is precisely because the manner of belief-formation is so unsafe that its local reliability is too much a matter of luck to count for much, insofar as justification is concerned.

Compare the epistemically more delightful Athena. She, like ordinary people in their everyday lives, employs the sorts of perceptual process that have been shaped by reasonable courses of experience with common enough objects. Experientially informed, or 'trained up,' perceptual processes having to do with familiar objects in clear view qualify as globally reliable processes. It strikes one as relatively safe to employ such a process in the absence of information suggesting that conditions are somehow exceptional.

Epistemic safety, then, is clearly associated with being a *suitably* truth-conducive belief-forming process. At least in the case of Athena and Fortuna, prima facie it is *global* reliability that makes for the difference that one senses in epistemic safety.[4] One might also put the contrast, and the difference in epistemic safety, in terms of the differential robustness of the reliability of the processes in play.

One can begin to appreciate why the epistemic safety—and the relative robustness of reliability—of belief-fixing processes would be epistemologically significant when one reflects on a prominent and pervasive characteristic of epistemic life: one's epistemic endeavor must be undertaken in the face of uncertainty. One's epistemic chores must be managed while possessing only

[4] Although global reliability is prima facie the key difference-making factor (and this idea is the core of what we are calling *neoclassical* reliabilism), ultimately we will be arguing that the real difference-making factor is *trans*global reliability.

a fallible understanding regarding one's global and local environment. Such epistemic uncertainty (or fallibility) regarding one's environment is paralleled by, or mirrored in, an uncertainty (or fallibility) regarding what processes will work in one's environment. In view of the uncertainty characteristic of the epistemic situation, consider two alternatives illustrated in the above discussion. On the one hand, one could employ a process that is reliable with respect to the wide reference class comprising the potential local environments to which one might be exposed (or which one might inhabit) within one's global environment. In other words, one might employ a globally reliable process; such a process is reasonably robustly reliable, as explained already. It is relatively safe insofar as its reliability then does not depend heavily upon certain unusual or atypical features that are highly specific to the particular circumstance or environment which the possessor of the disposition might happen to occupy; its reliability then does not obtain only relative to a narrow reference class of environments in which those particular features happen to be present.

Alternatively, one might employ processes that would be reliable only given certain unusual or atypical environmental features—only given features that are highly specific to particular circumstances or environments within the global environment that one happens to occupy.[5] *If*, in employing such a process one employs a reliable process, it is a merely locally reliable process. In light of the uncertainty that is a pervasive fact of epistemic life, it is clear what one should make of this general abstract choice. The safety afforded by globally reliable processes would be rationally desirable in preference to the risk one runs when using a merely locally reliable process. Globally reliable processes are safer and epistemically more valuable insofar as they afford one a certain margin of error (or margin for ignorance) regarding what local environment within one's actual global environment one happens to occupy. Because the reliability of a globally reliable process is relatively robust, not highly dependent on unusual features of one's local environment, it is relatively safe.

A natural line of thought, at this point, represents what we maintain is a significant insight into what is constitutively required for objective epistemic justification. The thought to be suggested is fully consonant with the above reflections on Athena and Fortuna, local and global reliability, and epistemic safety and risk—and furthermore, it is a line of thought that is well within the spirit of standard reliabilism:

First, embrace the idea that safety—a feature as yet to be more fully explained—is indeed a *constitutive requirement* for a belief's being objectively justified.

Second, look to give an account of safety as the belief's having been produced by some *suitably robustly reliable* belief-forming process.

[5] We are talking here about employing these processes indiscriminately, without possessing information to the effect that one is presently in one of the pertinently unusual environments. The more selective use of such processes is a matter we address in s. 2.1.f below.

We suggest that this line of thought provides the seeds of a very workable understanding of objective epistemic justification.

e. The Need for Further Refinement: the Cases of Diana, Delia, and Elena

How should one understand this idea of a suitably robustly reliable process? One might think, with an eye on the Athena/Fortuna cases, that global reliability *per se* is just the ticket. But a moment's reflection reveals that this proposal is too crude as it stands. Consider, for instance, Diana and Delia—another pair of characters in the fake-barn environment. Like Athena, both Diana and Delia have had reasonable experience with barns, and have trained up, generally serviceable, barn-perception processes. Suppose that Diana and Delia both have full knowledge that the particular local region in question is full of very realistic fake barns, and of the location of this region. They both have read all about it in the local morning paper, having spent the night in a hotel along the interstate located just a few miles from the region in question. They now find themselves driving through that peculiar local region (and they know this). It remains true of each of them that that her perceptual barn-categorization process is *globally* reliable; after all, the global environment is pretty much as it was before, so those processes remain globally reliable. As they drive along, they see a barn-looking object off in the distance. Suppose now that Diana refrains from forming the belief that it is a barn—because of the very real and salient epistemic possibility that it is a realistic-looking fake barn. And suppose that Delia forms the belief that the object is a barn—on the basis of its barnish appearance, and in spite of what she read this morning in the local paper. (Delia fails to bring to bear her knowledge that there are lots of fake barns in the vicinity.) Now, Delia's belief that the object is a barn is indeed produced by a globally reliable belief-forming process, namely, perceptual-appearance-based barn-categorization. (In this respect, Delia resembles Athena.) Nevertheless, Delia most certainly is *not* justified in coming to believe of the barn-looking structure off in the field, on the basis of its visual appearance, that it is a barn. Rather, in these circumstances she should refrain from forming that belief in that way—as does Diana. After all, *they each have excellent reason to believe—indeed, they each know full well—that, in their current local environment, the reliability of this belief-forming process is compromised (it is locally unreliable)*. Considered of themselves, those perceptual belief-forming processes remain globally reliable, here are locally unreliable, but in this case would not give rise to objectively justified beliefs (unlike in the case of Athena).

Likewise, consider the case of Elena. Suppose she knows, concerning that particular local area, both (a) that there are lots of barn facsimiles, and also (b) that all genuine barns, and no other structures of any kind (including either other kinds of buildings or barn facsimiles), are bright yellow. Finding herself in that very location (and knowing it), she catches a fleeting glance of a bright yellow structure, and promptly forms the belief that it's a barn. Of itself, such

a process of barn-categorization, on the basis of yellow color, would be globally *un*reliable. But, now, unlike the case of Fortuna, although Elena formed her belief on the basis of a globally unreliable process (namely, classifying briefly glimpsed structures as a barns on the basis of their bright yellow color), one judges that this belief of hers *is* justified, even so. Such processes here are locally reliable, globally unreliable, and yet here feature in the production of objectively justified beliefs (unlike in the case of Fortuna). What is going on?

This question reveals a task to be undertaken in forging a neoclassical version of reliabilism that will incorporate the idea of safety of belief-forming processes as a constitutive requirement of justification: one needs to give an account of such safety in terms of some suitably robust form of reliability. The cases of Athena and Fortuna reflected the importance of safety, and provided some reason to think of it in terms of global reliability. But, as just demonstrated by the cases of Diana, Delia, and Elena, global reliability *per se* will not fill the bill. But perhaps some refinement of it will. Notably, what seems different in the two sets of cases is how the processes in play—the perceptual barn-categorizing processes—themselves are related to further information and processes. So the suggestion to be developed is that epistemic safety might be understood in terms of suitably robust reliability, and that this will at least sometimes turn on how processes can be conditioned or modulated by certain other processes and the information thereby provided.

f. A Refinement: Suitable Modulational Control

An important general point emerges from these considerations: cognitive agents like humans deploy various belief-forming processes in ways that are *holistically integrated* within the agent's overall cognitive architecture. Very frequently, such processes are employed not in isolation, but rather under the modulating influence of various other or wider cognitive processes that are coupled to them and are poised to modulate them if and when certain forms of information become available to those wider processes. As we will put it, the given belief-forming process is under the *modulational control* of these associated processes. Such control can make for a selective application of the process or a selective inhibition, or can otherwise tailor its application to aspects of those local environments about which information is had—and thereby can enhance its reliability as so tailored. In principle, a whole host of different conditioning or modulating relations might be epistemically important. The wider processes might give rise to a narrower process—designing it or otherwise selecting or spawning it. They might trigger the conditioned process in ways that are fitting, or thought to be appropriate. (One's perceptual processes for spotting large, nonhuman, omnivores are primed when walking in known Grizzly territory.) They might inhibit it—making for a more selective use of the process. (One's perceptual processes for spotting large, nonhuman, omnivores are given less free rein when walking about the city.) In triggering or inhibiting the process, the conditioning of the narrower process draws upon

information that the wider processes generate or possess. All such kinds of modulation can and should be found among normal human cognitive agents.

Let us explain our terminology more explicitly. A belief-forming process P may be under the conditioning control of a wider set of processes—with or without those wider processes having yet come by information that prompts changes in, or modulations of, P. This wider set of processes may be termed *control processes* with respect to P. When there is such a functional relationship between processes, the process P may be said to be *under the modulational control of* the wider processes. Being under the modulational control of wider processes, and being a control process, are a matter of the dispositional or control relationship between wider processes and some narrower processes. This does not require that the control processes have produced, or have acquired, information prompting them to effect changes in the processes to which they are coupled, or that they ever will; it just requires that they are so 'positioned' in the agent's cognitive system as to stand ready to do this, i.e. they are so disposed. When the control processes turn up information (veridical or not) bearing on the reliability of a process P, and when P or its use thus comes to be spawned, tailored, selectively triggered, selectively inhibited, or in some like manner altered, P may be said to be *modulated by* those wider processes. So a belief-forming process P is *under the modulational control of* a wider set of processes S, within the agent's cognitive system, if S would tailor P (would trigger P, inhibit P, or the like), were S to come to generate or possess certain relevant information. P *is modulated by* S, provided that S has generated or come to possess results that prompt it to actually make some changes in how or when P is put into play.

What kinds of modulational control mechanisms there are in humans, and what kinds humans are *capable* of possessing and deploying, are empirical questions, principally within the purview of cognitive science. Although the examples of Diana and Elena both involved a kind of modulation that turned on certain articulate beliefs, there is no reason to think that all modulational control need be a matter of processes that work at the level of belief. Indeed, it may be useful to think of many processes to which philosophers have commonly made reference as somewhat complex and as involving an important modulation of the core processes. For example, one expects that a competent perceptual agent may (it seems must) make subtle and sensitive use of a great deal of information. Some of this information may be represented at the level of beliefs, of course, as in the case of Diana. But, some of it may be had by the cognitive system and condition perceptual processes without being so represented. A competent perceiver might be sensitive to much in the conditions of an observation without being able to represent those matters in a way even approaching conscious belief. Light levels, degrees to which an object is obscured by other things, and the like clearly are important for perceptual reliability, and these are matters to which a system might be sensitive without representing them via full-fledged beliefs about them—and perhaps without consciously noticing them at all. Such sensitivity

could condition perceptual processing by giving rise to occasional states that might be described as hesitations, or warnings, indicating that the perceptual processes were compromised by conditions in ways that might yet be somewhat vague. Still, this would make for a kind of inhibitory modulation built into the perceptual processes themselves, triggered by information processing below the level of conscious awareness. So, some modulation in humans probably relies on information that does not rise to the level of belief-like states, or of other states commonly recognized by epistemologists.

From a reliabilist point of view, the presence and efficacy of suitable control processes, within the cognitive system viewed as an integrated whole, thus emerges as extremely important in relation to objective justification. What count, though, as *suitable* control processes? What counts as being under suitable modulational control? This looks to be a question that cannot be answered in any great detail from the armchair. What ultimately makes for satisfactory, epistemically competent, belief-formation cannot be settled in detail without taking into account what processes can be tractably employed (perhaps innately, perhaps via learning) by cognitive systems of the relevant sort—adult human cognizers, in much epistemological enquiry. So, what counts as suitable modulational control depends in part upon the cognitive architecture of the given kind of cognitive agent, and upon the agent's susceptibility (by virtue of its cognitive architecture) for learning or internalizing various kinds of controlling processes that might not be innate to it. Also, whether or not *all* belief-forming processes are susceptible to modulation is itself an empirical question for cognitive science. One possibility is that all belief-forming processes are thus susceptible, at least to some extent; but another is that some belief-forming processes are not themselves so susceptible (though they might figure importantly in modulating *other* such processes). The notion of being under *suitable* modulational control is thus one that leads directly to a call for naturalization in epistemology.

g. Neoclassical Reliabilism

With the notion of modulation by control processes in hand, and also the notion of suitable modulational control, let us return to the question whether there is a form of robust reliability that is a plausible candidate for what constitutes *safety* of a belief-forming process. The preceding discussion points in a natural-looking direction. The suggestion is this:

> Safety is a belief's having been produced by a belief-forming process that is *globally reliable under suitable modulational control*.

This is the neoclassical reliabilist (or refined *global* reliabilist) proposal for explicating epistemic safety.

Several points need to be made about this proposal, by way of elaboration and explication of what we mean by the phrase 'produced by a belief-forming process that is globally reliable under suitable modulational control'. First, the claim

that a belief is produced in this way entails that the given belief-forming process is indeed under suitable modulational control *at the time it produces the belief*. Second, the claim entails that the given process, *when under suitable modulational control*, is globally reliable; this leaves open whether or not it would be globally reliable anyway, apart from such modulational control.[6] (Elena's special-purpose belief-forming process, whereby she classifies a local structure as a barn solely on the basis of brief glimpse of its yellow color, is certainly not globally reliable apart from suitable modulational control; but it is indeed globally reliable when it is under the modulational control of processes that selectively trigger it only when she has information to the effect that locally, all and only yellow structures are barns.) Third, a belief-forming process need not be undergoing actual modulation at a given time in order to be under modulational control at that time; rather (as already stressed above), being under modulational control is a matter of being coupled to other processes that are *poised* to do modulation, should pertinent information become available to the cognitive agent. Fourth, if (a) such pertinent information is indeed available at a time t, but (b) the agent's modulational mechanisms nonetheless fail to operate at t because of a performance error (e.g. as a result of fatigue, or emotional distress, or distraction, or the like), and (c) the given belief-forming process operates at t in an unmodulated way, then this counts as a case in which the process is *not* under suitable modulational control. (It counts as such a case even though modulational-control mechanisms are indeed presently in place, and would be presently operative if the agent were not currently suffering a lapse in cognitive competence—a performance error.) Fifth, although (as stressed already) what counts as suitable modulational control is partly dependent on facts about human cognitive architecture that are matters for scientific theorizing about human cognitive competence, nevertheless the following is (according to neoclassical reliabilism) a constitutive aspect of the notion of suitable modulational control: such control is globally reliable *itself*. That is, cognitive mechanisms of suitable modulational control *enhance* the global reliability of the belief-forming processes that these cognitive mechanisms modulate; they do so by reliably enhancing the cognitive agent's tendency both (a) to employ certain belief-forming processes when they are locally reliable, and (b) to refrain from employing them when they are not locally reliable.

From a reliabilist perspective, there is a clear and powerful general rationale for a construal of epistemic safety that incorporates the idea of suitable modulational

[6] How might further explication be given to the idea of a process-type being globally reliable when under suitable modulational control? One possibility might be to say that it is not a matter of an ordinary process-type P having some special form of global reliability, but rather is a matter of a special process-type P^+ having ordinary global reliability—where what's special about P^+ is that the presence of suitable modulational control is built right into the individuation conditions of P^+ itself. We will remain neutral here about this proposal. The pre-theoretic idea *reliability under suitable modulational control* seems clear enough already to serve our present purpose of articulating neoreliabilism; so we leave open whether this idea can be further explicated (and if so, how).

control. The cognitive agent *as a whole system* will be significantly more reliable in its belief-formation than it would otherwise be, provided that various belief-forming processes it employs are under suitable modulational control of other cognitive processes (rather than being employed in an unmodulated manner). What is wanted epistemically, from a reliabilist perspective, is that the cognitive agent should employ specific belief-forming processes *when they are apt to be reliable*—and likewise, that the cognitive agent should refrain from employing those processes when they are apt to be *unreliable*. Suitable modulational control *globally* enhances a cognitive agent's tendencies to employ belief-forming processes in such *locally* sensitive ways—because such control is itself globally reliable.

2.2. Transglobal Reliabilism

We now discuss a few thought experiments indicating that the reliability property featured by neoclassical reliabilism—namely global reliability under suitable modulational control (for short, *neoclassical* reliability)—does not yield an adequate account of objective epistemic justification. However, these thought experiments do not suggest that epistemic safety is not needed for objective justification, nor do they indicate that safety is not afforded by robust reliability. Rather, the lessons from the foregoing, about epistemic safety as robust reliability, are preserved and extended. The thought experiments that we will now discuss suggest that the kind of safety that is constitutively required for objective epistemic justification is a yet more robust form of reliability, different from neoclassical reliability.

a. Epistemic Safety without Neoclassical Reliability: The New Evil Demon Problem

The scenario to be presented here was set out by by Lehrer and Cohen (1985). and was labeled the 'new evil demon problem' for reliabilism by Sosa (1991). In keeping with philosophical mythology, suppose that there is an evil demon—malicious and very powerful—out to deceive the agent. Seeking to epistemically defeat the agent at every turn, the demon provides the agent with appearances or experiences that are compellingly consistent in the way they indicate an apparent world in which the experiencer is an apparently embodied, apparently perceiving, apparently acting, cognitive agent; these experiences are just as compellingly consistent as are the embodied, en-worlded, experiences of ordinary humans. As the agent undertakes to do things in that world, the demon responds by giving the agent the fitting appearances. But, the world that the agent and the demon inhabit is radically other than the world that the agent is led to believe in and theorize about—thus the deception. Of course, if the demon is really good at deception, then the agent is presented with sufficient

epistemic problems to 'solve'; these keep the agent engaged making sense of the seeming-world the agent experiences.

Objectively, there is no hope for an agent in a demon-infested environment. There are no reliable processes to be had in such an environment.[7] Whatever cognitive processes the agent employs, the demon will counter with input that is fittingly deceptive (that leads to false beliefs) given those processes.

These days, computers are commonly cited in place of evil demons when constructing skeptical scenarios. Typically, one supposes that the supercomputer has charge of a brain in a vat. Again, the envatted brain has no real chance of arriving at systematic true belief.

Consider several agents inhabiting such a demon-infested environment. To begin with, suppose that some agent, call her Constance, is remarkably like the intelligent, educated, and conscientious epistemic agents one would want to include in one's own epistemic community. She avoids fallacious ways of reasoning, both deductive and inductive. Constance avoids inconsistency as well as the best of us. She maintains a high degree of wide reflective equilibrium in her belief system(s). In keeping with the facts as just stipulated, she only generalizes when samples are large enough for statistical confidence at some high level, and then only when the samples are either random or characterized by a diversity that seems to match the distribution of likely causal features in the population. Further, Constance has taken note of where her observations have seemed untrustworthy in the past, and discounts certain observations accordingly. In crucial respects, she is like the best perceptual agents among us. She is like those who, on the basis of long experiential refinement, prompted by successes and failures, have come to be sensitive, careful, and discriminating perceivers of everyday things. Constance has become trained up through a long process of refinement in her perceptual experience, drawing on apparent successes and failures, to be sensitive and careful in how she forms beliefs via perceptual experience.[8] Put simply, if one were picking epistemic teams to play in our actual global environment, one would not hesitate to pick Constance for one's own team, as long as she could join us in our actual global environment. In the actual environment, Constance would be a model epistemic citizen.

Suppose then that Constance, in her demon-infested environment, hears familiar noises clearly emanating from the phone on her desk (or so things appear), and that she then forms the belief that someone is calling. Of course,

[7] At least this holds for empirical beliefs that the agent might generate. The case of *a priori* beliefs is not addressed here. Also, we are here assuming that empirical beliefs pervasively of the form 'The Deceiver is providing me excellent evidence for the false statement that . . .' are non-starters as candidates for being objectively justified (even though they happen to be true), since agents in the envisioned scenario have not a shred of evidence for such radically paranoid empirical beliefs.

[8] Of course, because she is subject to demonic deception, her training has of necessity been reflective of false successes and false failures. But otherwise it reflects the sort of training and shaping of perceptual processes that makes for systematic success in our actual global environment.

the belief is false and arises by way of a highly unreliable process. Virtually all processes are both locally and globally unreliable in an environment where a powerful and resourceful evil demon (or analogous supercomputer) is at work on the inhabitants. Still, there is a very strong tendency to judge that Constance is objectively justified in holding this belief. Of course, one does not find this scenario to be a happy or epistemically desirable scenario. But one is strongly disinclined to find fault with the agent. The problem lies with the agent's extremely inhospitable environment—its demon-infestation. There, no agent and no process will help. So, one judges that the problem is in with the environment, and not the agent. Fine agent, lousy environment. In keeping with this, one judges that Constance is justified and has done nothing epistemically wrong or inappropriate in believing that someone is calling. [9]

This judgement, if based on a genuinely intelligible scenario, poses a clear challenge to standard reliabilist accounts of justification. If it is honored, then such accounts must be abandoned. Neoclassical reliabilism is not immune to the challenge. This refined position takes global reliability under suitable globally reliable modulational control to be constitutively required for objective justification—to be necessary for justification. Notice that Constance has in place very significant conditioning processes, exhibiting significant modulational control (of a sort that in less inhospitable environments would be globally reliable themselves, and would suitably enhance the global reliability of the processes they modulate), but these cannot contribute to global reliability in her global environment, for no processes will avail agents in this environment. Constance's processes lack the property that neoclassical reliabilism says is a necessary condition for objective justification: the property of being globally reliable under modulational control.

Consider, by way of contrast, a different agent who is also beset by a powerful deceptive demon (or computer). This agent, call her Faith, is provided with the appearance of a community that holds certain epistemic standards that are quite at odds with those that most of us have come to approve. For example, folk, or rather apparent folk, in Faith's epistemic community engage in the gambler's fallacy, and consistently approve of such inferences. They have no notion of the representativeness of samples, to take another example, and do not have evaluative practices that would lead to the associated caution in forming generalizations from instances. It is as if Faith has been raised in such a community—and has come to have the predictable inferential tendencies. Faith is not unreflective, however, and attempts to get at the truth as best she can. She conscientiously applies her epistemic standards—such as they are. (When bets or strategies informed by instances of the gambler's fallacy lead to disappointments or disaster, this is written off as

[9] Note well: One judges that Constance's beliefs are *objectively* justified—not merely that they are subjectively justified relative to her own epistemic standards. Her epistemic standards are objectively just fine, even though she's in a damn lousy global environment.

cursed luck.) One day, Faith notes that it has been quite a while since the fair die used in a game has turned up a six. So, she forms the belief that a six is due—and that the probability of a six on the next toss is rather higher than 1/6—let us say she believes it is greater than 0.5. What is one to make of Faith's belief here?

Of Faith, one is no longer inclined to say (as one was of Constance), 'Fine agent, lousy environment.' One is inclined to judge that there is something about Faith's processes themselves (and not just the lousy environment) that makes them objectively inappropriate, and that makes her belief objectively unjustified.[10]

Of course, Faith's disposition to commit the gambler's fallacy is no less reliable in the demon-infested environment than were the processes that Constance employed.[11] Even so, one is inclined to judge that there is something wrong—objectively wrong—with Faith's processes, *and not with Constance's.* One is inclined to judge that Faith is not objectively justified in believing as she does, *although Constance is*. Constance is objectively and subjectively justified, while Faith seems only subjectively justified.

Apparently, the global reliability of belief-fixing processes is not a necessary condition for being objectively justified in believing. At least this is what seems indicated by one's judgements about Constance in her demon-inhabited environment.

b. The 'Truman Show' Scenario, with Harry and Ike

This scenario is structurally parallel to the Athena/Fortuna scenario, and elicits intuitive judgements suggesting that the property *being globally reliable under suitable modulational control* is not constitutively required for justification. Although epistemic safety does indeed appear to be constitutively required for justification, an alternative account of this feature is needed.

In the 1998 movie, *The Truman Show*, the main character has been raised from childhood on a movie set where he has lived his entire life being the subject of a massive deception by skillful actors. The result has been a 24 hour a day television program in which viewers are treated to Truman's life from birth (which, one supposed, wasn't staged) to his being a young adult. Suppose there is a competing show in which two friends, Harry and Ike, have been similarly deceived and observed. Harry and Ike spend their entire lives in carefully contrived circumstances in which everyone else they ever interact with is an actor playing a part in an elaborate conspiratorial deception. The conspiracy is so well

[10] Faith may be *subjectively* justified in believing, and one tends to evaluate her conscientiously formed beliefs accordingly. Given her observation of a long enough run of non-six tosses, and her beliefs or standards, she is subjectively justified in her belief that a six is due. Still, one is far less inclined to judge that Faith is objectively justified.

[11] Constance's dispositions are better than Faith's as a basis for predicting the future course of first-person experience, of course. But Constance's are just as bad as Faith's with respect to the truth values of the external-world beliefs they generate.

orchestrated that it has rich 'counterfactual depth': for various potential actions that Harry or Ike might engage in, including actions such as undertaking to change jobs or to travel to distant lands, the conspirators (the actors, directors, and stage-hands) stand ready and able to accommodate themselves in such a way as to maintain the ongoing deception. Thus, they are capable of doing their deceptive work within whatever alternative local environments Harry or Ike might venture into. Put simply, the deception of Harry and Ike extends potentially to such an extensive set of local environments that it is global.

Harry is highly paranoid, in one specific way. He believes that everyone else around him except Ike is an actor playing a part, and that every local environment he ever finds himself in is an elaborate stage-set being manipulated by designer/controllers behind the scenes. Harry possesses not a single shred of evidence for these sweeping paranoid beliefs, and he never obtains any evidence for them throughout the entire course of his life. Ike, on the other hand, exhibits no such paranoia. He takes his ongoing experience at face value, and has no tendency at all to believe or even suspect that he is being duped by a bunch of actors or that all his local environments are elaborate stage-sets. Ike believes that his friend Harry is just hopelessly paranoid, and he considers Harry's paranoid beliefs far too silly to take seriously for even a moment. (Ike too never receives even the slightest positive evidence in support of Harry's beliefs.)

Harry's curious belief-forming processes are not the product of more pervasive generalization practices in which he fails to pay adequate attention to sample size and the like. Rather, he has a very specific form of paranoia about what those around him are up to. In him it came about as the effect of the enormous psychological trauma he suffered upon having learned as a child that his parents and all other adults were lying to him about Santa Claus. His being so upset provoked a one-time sweeping generalization—that in which he came to his belief that everyone but his friend Ike (who has been his best pal since early childhood) is constantly play-acting an elaborate deception.

Harry's paranoid belief-forming processes produce an excellent track record of true beliefs over the course of his life. Moreover, objectively, these processes have an extremely high propensity to produce true beliefs within Harry's actual global environment—since all those actors and stage-set designer/controllers do indeed stand ever-ready to accommodate themselves to whatever choices Harry might make about moving from one local environment to another. They maintain the deceptive conditions flawlessly, and yet Harry believes that he is being deceived in all those things that he would otherwise tend to believe.

Harry's specific form of paranoia thus constitutes an effective modulational-control mechanism. Whenever P is a potential belief-content that he is otherwise inclined to believe that involves other people's actions or purposes, his paranoia both (a) inhibits the belief-forming process that would otherwise generate a belief that P (thereby preventing this belief from arising), and (b) generates the belief that *those people are pretending that P*. Because of the peculiar nature of his global

environment, with that troop of play-actors systematically deceiving him and Ike and standing ready to continue doing so whatever either he or Ike might do and wherever either he or Ike might go, this paranoid form of modulational control is wonderfully conducive to global reliability: it prevents him from forming *false* beliefs that he would otherwise form about what people around him are up to, and it converts his erstwhile inclinations to form such false beliefs into *true* beliefs about what they are up to. So, by the neoclassical yardstick of global reliability, Harry's paranoia counts as a (very effective) form of *suitable* modulational control. So the belief-forming processes that produce Harry's systematically paranoid beliefs are neoclassically reliable: given the neoclassical understanding of suitable modulational control, Harry's belief-forming processses are indeed globally reliable under suitable modulational control. Thus, Harry's paranoid beliefs count as *safe* according to neoclassical reliabilism—and hence count as objectively justified.

Intuitively, however, Harry's paranoid beliefs are epistemically unjustified even so: they are highly *risky*, rather than being safe. Unlike in the case of Fortuna, this lack of safety cannot be understood in terms of a lack of global reliability in the processes that form these beliefs (or in the modulating mechanisms that spawned those processes). The local reliability of Harry's processes does not depend on special or peculiar features of some particular local environment. All local environments are ones into which Harry and Ike's deceivers would freely follow (or even precede) them. So, all local environments are ones in which these agents would be subjects of deceptive presentations. For any potential local environment E within Harry's global environment, if Harry *were* to go into E then his processes *would be* reliable in E. The cumulative effect is that, inhabiting a global environment with able deceivers as his constant traveling companions, Harry's processes are globally reliable (as are the modulating mechanisms that spawned them). Still, the paranoiac belief-forming processes strike one as unacceptably risky or unsafe.

Recall that Fortuna's locally reliable processes seemed risky, despite being locally reliable, because that local reliability seemed dependent on peculiar aspects of her environment that her cognitive processes did not track and for which she had no informational basis, even no apparent informational basis. Similarly, Harry's processes seem risky, despite being globally reliable, because that *global* reliability is highly dependent on peculiar aspects of his *global* environment that his cognitive processes seem ill suited to track,[12] and for which he has no real indication. Such thoughts begin to come into better focus when one deploys, for contrast, the idea of a process that is yet more robustly reliable: a

[12] Although his paranoid form of modulational control does make for processes that are globally reliable, this itself seems an extraordinary bit of luck. It is not as though this control process really 'tracks' anything about Harry's environment—as is evident when one imagines how it would fare across possible global environments.

transglobally reliable process would not be highly dependent on peculiar aspects of the agent's global environment—because it would be reliable with respect to the class of experientially relevant possible global environments.

Ike's belief-forming processes fare differently. In the numerous cases in which his non-paranoid beliefs are at odds with Harry's paranoid ones, Ike's beliefs are systematically mistaken. But despite this very poor track record with respect to the goal of systematic true belief, intuitively these beliefs of Ike's are well justified nonetheless. This is because the processes that generate the non-paranoid beliefs are intuitively epistemically *safe*—even though they happen to be systematically non-veridical because of all that conspiratorial play-acting and behind-the-scenes set designing/controlling. As in the case of Athena, this epistemically laudable safety seems closely tied to the fact that Ike's belief-forming processes here have a kind of robust reliability—but clearly, this is not to be understood as a matter of being globally reliable. Harry's processes are globally reliable (as are the modulational mechanisms to which they are subject), yet these processes lack the relevant epistemic safety, while it seems that Ike's processes lack global reliability and yet possess the relevant epistemic safety.

3. LESSONS: TRANSGLOBAL RELIABILISM

For reasons of space, we will not continue proliferating scenarios, and we will quickly suggest some lessons. We ourselves find all the scenarios discussed here decidedly intelligible. The advocate of strong semantic externalism will find all but the evil demon (or brain in vat) scenarios intelligible. All the scenarios suggest that some notion of epistemic safety of belief-fixing processes is central to judgements deploying the concept of being objectively justified in believing. The judgements prompted by the cases of Athena, Fortuna, Diana, Delia, and Elena provide reason to believe that epistemic safety could be understood in terms of processes that are robustly reliable. Globally reliable processes are more robustly reliable than are merely locally reliable processes. So, this encouraged formulating and considering neoclassical reliabilism—which turns on global reliability under suitable modulational control. However, the new evil demon problem, in the persons of Constance and Faith, seems to indicate that such neoclassical reliability is not an adequate measure of the needed epistemic safety. Constance lacks it—and seems epistemically safe and objectively justified. She should count as better off from the point of view of objectively justified belief than Faith—although neither uses processes that are neoclassically reliable. None of this suggests that the needed form of epistemic safety cannot be understood in terms of robustly reliable cognitive processes. Rather, it suggests that neoclassical reliability—i.e. global reliability under suitable modulational control—does not make for the required robustness of reliability. Now, the fan of strong semantic externalism may write off as unintelligible the scenarios involving Constance

and Faith. But the cases of Ike and Harry do much the same work, and should be uncontroversial. Ike's processes are globally unreliable, but safe. Harry's are globally reliable under what the global reliabilist should think was suitable modulational control—but unsafe.

Things fall into place when one takes seriously the suggestion that robustness of reliability provides the needed key to what makes for epistemic safety and objective justification. As global reliability stands to local, so transglobal reliability stands to global—transglobal reliability is a more robust form of reliability than global. In all the scenarios discussed here, those who are judged to be justified in their beliefs deploy processes that are transglobally reliable—and those who are judged to be unjustified deploy processes that are not. If one were to consider yet further cases, one would find that the concern for suitable modulational control needs to feature in transglobal reliabilism in a role parallel to the role envisioned in neoclassical reliabilism. (The parallel form of suitability would include the transglobal reliability of the control processes themselves.) So, our suggestion is this:

Safety is a belief's having been produced by a belief-forming process that is *transglobally reliable under suitable modulational control.*

This is the transglobal reliabilist proposal for explicating epistemic safety. Earlier remarks about how to understand the expression 'globally reliable under suitable modulational control' all carry over, *mutatis mutandis*, to the expression '*transglobally* reliable under suitable modulational contol.' (Hereafter we will refer to the feature picked out by this latter expression as *postclassical* reliability.)

When the demand for postclassical reliability is applied to the cases of Athena, Fortuna, Diana, Delia, and Elena one finds that the common judgement tendencies are readily accommodated. Postclassical reliability provides a powerful and unitary explanation for the full range of scenarios we have considered in this chapter.

Although neoclassical reliabilism does deliver the intuitively correct verdicts about questions of objective justification regarding these cases, it does so for the wrong reasons. An obvious strategy for making this claim plausible is to consider envatted-brain variants of some of these respective fake-barn cases. Suppose that Athena and Elena each has a lifelong experiential duplicate who is an envatted brain: Envatted Athena and Envatted Elena.[13] These latter two characters each have a barnish visual experience, while apparently driving along a highway after apparently having read in the morning papers about numerous realistic-looking fake barns in the immediate vicinity. For both Envatted Athena and Envatted Elena, the envatment set-up works in such a way that the following counterfactual-involving facts obtain: (1) if the envatted agent has an experience

[13] We focus on Athena and Elena because we want cases in which neoclassical reliabilism will assign a different justification-status to the envatted agent's barn-belief than it assigns to the unenvatted agent's barn-belief.

as of a yellow building, then if she were to have experiences as of carefully investigating whether the building is a barn, she would have experiences as of its turning out to be a barn; and (2) if she has an experience as of a non-yellow building, then if she were she to have experiences as of carefully investigating whether the building is a barn, she would have experiences as of its turning out *not* to be a barn.

Intuitively, Envatted-Athena's barn-belief has the same justification status as does Athena's corresponding barn-belief: both are epistemically safe, and in exactly the same way. Transglobal reliabilism accommodates this intuitive judgement just fine, because in both cases the belief is caused by a belief-forming process that is transglobally reliable under suitable modulational control. Neoclassical reliabilism, on the other hand, cannot accommodate the intuitive judgement; instead, it must claim that Athena's barn-belief is objectively justified whereas Envatted Athena's is not (because Athena's is neoclassically reliable whereas Envatted Athena's is not). Likewise, *mutatis mutandis*, for Elena and Envatted Elena.

So much the worse for neoclassical reliabilism, insofar as accommodating intuitive judgements about cases is concerned. This underscores the moral already drawn earlier: that the *epistemic safety* constitutively required for justification is that provided by postclassical reliability rather than neoclassical reliability. Athena's barn-belief is justified not because the process that produced it is neoclassically reliable, but rather because this process is postclassically reliable. Envatted Athena's barn-belief is justified in exactly the same way as Athena's, even though Envatted Athena's belief is produced by a process that is not neoclassically reliable at all. Likewise, *mutatis mutandis*, for Elena and Envatted Elena. Epistemic safety is the crucial requirement, and epistemic safety is a matter of postclassical reliability.[14]

Earlier, when contrasting the greater robustness of reliability had by globally reliable processes in comparison with merely locally reliable processes, we recommended a useful way of understanding degrees of robustness of reliability. The general idea was as follows: The robustness of reliability possessed by a process is a matter of the range of environments in which the process would be reliable. So, consider a process and an environment in which it is reliable; then consider variations in that environment; a robustly reliable process will retain its reliability through significant variation in environments, before ceasing to be reliable. It works as follows when contrasting *globally* reliable versus *merely locally* reliable processes: Consider a local environment relative to which a process is a merely locally reliable process; then notice that as one 'varies the local environment' by considering other local environments afforded by the agent's global environment, one does not need to 'move to' very dissimilar local environments before getting

[14] There could also be envatted counterparts of the other three characters—Fortuna, Diana, and Delia. Here the just-mentioned lesson does not emerge as starkly, because neoclassical reliabilism delivers the same, correct, verdict for Fortuna and for Envatted Fortuna (and likewise for each of the other two pairs)—rather than treating the members of the pair differently. Still, such cases are to be counted as further successes for postclassical reliabilism.

to ones in which the process is no longer locally reliable. In contrast, for a process that is both locally and globally reliable, one could 'vary the local environment' much more—could put the process to work in rather more dissimilar local environments—before the process would become unreliable. This should serve to make vivid the idea of a process being reliable in a wider range of local environments afforded by an agent's actual global environment.

Now, *transglobally* reliable processes are more robustly reliable than merely globally reliable processes, because they are reliable across a wide range of experimentally possible *global* environments. Again, thinking of varying environments—now global environments—can help understand what is at issue. The global reliability of the merely globally reliable process could be dependent on the particulars of the specific global environment in which the process is employed, in such a way that the process would fail to be reliable in global environments that are not highly similar. Starting with a global environment in which it is clearly reliable, think of that process being deployed in varying experimentally possible global environments. If the process is *merely* globally reliable, then as one considers increasingly dissimilar global environments, the process will quickly become unreliable. On the other hand, the global reliability of a transglobally reliable process is not so sensitive to, or so highly dependent on particularities of, a given global environment—and would be retained across greater variations in global environment.[15]

To provide a quick and relatively simple illustration, consider two inductive processes. One incorporates sensitivity to sample bias, size, and the like, while the other is without such sensitivity. Then imagine a global environment with sufficient homogeneity of populations and causal structure that both processes would be globally reliable. The first is transglobally reliable, while the second is merely globally reliable. As one imagines global environments with increasing heterogeneity of populations and causal structure, one quickly comes to global environments in which the inductive processes without sensitivity to sample properties become globally unreliable—while the processes incorporating sensitivity to sample properties would continue to be globally reliable. The transglobal reliable process is the safer.

A more thorough development of transglobal reliabilism would make more articulate the concern for the real tractability of cognitive processes for a class of cognitive systems—a concern that has not been highlighted in this chapter (although it is reflected in remarks on the empirical issues surrounding the

[15] An earlier version of this paragraph led one reader to a misunderstanding that we hope now to have forestalled. The reader understood us as weighting some global environments more heavily in gauging transglobal reliability, perhaps in proportion to their similarity to the actual world, or to the agent's world. We do neither. Instead, we write of similarity in global environments as a way of understanding or gauging the range of environments across which a process would be globally reliable. It is not as though any global environment is more important or weighty than any other in determining transglobal reliability—what is important is the range, the extent to which the degree of global robustness of a process is or is not dependent on peculiarities of a particular global environment in which it is employed.

matter of what constitutes 'suitable modulational-control processes' for human cognitive agents). What has been highlighted here is the concern for epistemic safety and the sense one has that it is intimately connected with the reliability of one's belief-fixing processes. Reflecting on the modal dimensions of the concept of *reliability*, one is led to associate epistemic safety with adequate robustness of reliability. Thus, the motivation for transglobal reliabilism can be recognized as being continuous with the motivation for reliabilist positions generally. This said, scenarios such as those involving Constance and Faith, and Harry and Ike, are variations of scenarios that have motivated *internalist* epistemologists in their misgivings regarding reliabilism. Transglobal reliabilism thus has the capacity to respond sensitively to such internalist concerns, accommodating them naturally within an externalist, reliabilist, account of objective epistemic justification.

We turn next to the question of how internalist and externalist epistemology look from the perspective provided by transglobal reliabilism.

4. EPISTEMOLOGICAL INTERNALISM AND EXTERNALISM

Reliabilist epistemologies are typically externalist in character. The status of transglobal reliabilism on this score is rather a subtle matter, however. Relative to dominant construals of internalism in epistemology, on one hand, transglobal reliabilism counts as an externalist account of objective justification. On the other hand, transglobal reliabilism shares certain important features in common with epistemological internalism—features which are lacking in more standard forms of reliabilism, including neoclassical reliabilism. Let us address these two sides of the matter, in turn.

An epistemological position regarding some epistemic-evaluative status—in this case, the status *being justified in holding a given belief*—is epistemologically internalist just in case the position would make that status depend only on matters that are accessible to the individual epistemic agent. This is to understand epistemological internalism as what we will henceforth call 'access internalism'—i.e. the internal is what is directly, introspectively, accessible to the agent. We take it that what is commonly called 'internalism' in epistemology amounts to some form of access internalism.[16]

Transglobal reliabilism holds that a constitutively necessary condition for a belief's being objectively justified is that the belief has been generated or maintained by a process that is postclassically reliable—i.e. is transglobally

[16] Our characterization of access internalism does not build in the requirement that the belief's epistemic status *itself* must be directly accessible to the agent, even though it does require that this status is entirely determined by factors that are directly accessible. One might call a position that builds in this additional requirement '*industrial-strength* access internalism'.

reliable under suitable modulational control. Whether or not a belief is generated or maintained by such a process does not depend only on matters that are accessible to the agent. There are several reasons why not, most having to do with the limited access one has to the character of one's own cognitive processes, particularly as these are instanced in an episode.

First, in general *the nature of the belief-forming process itself* need not be accessible to the agent. That is, the agent might not be able to tell, directly and introspectively, anything very specific or detailed about the cognitive operations that generate or maintain the belief in question. One way of specifying the process would be in terms of a theoretical account, at the cognitive-science level of description, of the pertinent cognitive architecture and the pertinent information-processing procedures subserved by that architecture. But in general that is a matter for empirical cognitive science to discover, rather than something that is fully ascertainable by introspection. Another, more abstract and generic, way of 'specifying' the process might be as some kind of input-output function—say, where (i) each item in the domain of the function comprises the current total experiential state of the agent together with all the agent's background beliefs, and (ii) each corresponding item in the range of the function is a set of new beliefs (the ones that the given process would generate from the given input plus the given background beliefs).[17] But such a gigantic input-output function, with such sprawling items in its domain, certainly could not be consciously entertained all at once, and hence certainly is not an introspectively accessible item. So, the detailed nature of one's belief-forming process is not something that normally is accessible. But according to transglobal reliabilism, the nature of the belief-forming process is one of the factors that determine whether or not the belief in question is objectively justified. Since access internalism requires that all such factors be accessible to the agent, and since this process factor is normally not accessible, transglobal reliabilism thus runs contrary to access internalism.

Further conflicts with access internalism arise from transglobal reliabilism's requirement that the belief-forming process be under suitable modulational control. In general, whether or not this requirement is satisfied will not be accessible to the agent. For one thing, the detailed *nature* of one's modulational-control processes (if any) typically is not accessible. For another, whether or not such processes are (or were) fully *in play* in an episode need not be accessible. For a third, the factors that determine what counts as a *suitable* degree of modulational control typically will not be accessible either (since such factors will include aspects of cognitive architecture that are inaccessible themselves). So, the requirement of suitable modulational control generates several further respects in which transglobal reliabilism runs contrary to access internalism.

[17] The generic belief-forming process specified by such an input-output function will be multiply realizable by various, more specific, kinds of processes specifiable at the cognitive-science level of description—including the specific one that plays this realizing role in the given agent.

But although transglobal reliabilism is an externalist position, insofar as epistemological internalism is construed (in the usual way) as access internalism, nevertheless there are aspects of commonality between transglobal reliabilism and epistemological internalism—aspects not shared with more familiar forms of reliabilism (e.g. neoclassical reliabilism).

To begin to explore the commonalities between postclassical reliabilism and epistemological internalism, it will be fruitful to articulate a second, less constraining, form of internalism. Access internalism about justification entails, but is not entailed by, the following thesis, which we will call *agent internalism*:

A belief's being justified depends only on psychological features of the agent who holds the belief.[18]

According to this thesis, psychological facts about the agent constitute a full supervenience base for a belief's justification-status, independently of any further facts about the agent's global environment.[19] Agent internalism about justification is an important thesis, marking a significant divide between accounts of justification that conform to it and those that do not. It surely does deserve the label 'internalism', since its claim is that justification depends entirely on matters *internal to the agent's own psychology and psychological history*. Admittedly, it is a weaker or more liberal form of internalism than access internalism. All access internalist positions are agent internalist, but not all agent internalist positions are access internalist.

Transglobal reliabilism, unlike neoclassical reliabilism, is an agent-internalist position.[20] To appreciate this, one need only consider the various factors that constitute postclassical reliability of a belief-forming process—i.e. transglobal reliability under suitable modulational control. First, there is the nature of the belief-forming process itself; this is clearly psychologically internal to the agent. Second, there is the nature of the modulational-control processes (if any) that are connected to the given process; this too is psychologically internal to the agent. Third, there are the features of the agent's cognitive architecture that are pertinent

[18] The claim that access internalism entails agent internalism assumes that any item that is directly introspectively accessible by an agent is a psychological state of that very agent.

[19] The 'further facts about the agent's global environment' that are held to be irrelevant to justification, according to agent internalism, are facts over and above whatever external facts (if any) figure constitutively in the supervenience bases of the agent's psychological states themselves. Agent internalism is neutral about 'content externalism' in philosophy of mind—the view that the content of an agent's psychological states is partly constituted by certain relations the agent bears to the actual global environment. Perhaps the supervenience base for the agent's own psychological states includes certain aspects of the agent's wider environment, in content-fixing roles; or perhaps not. Be that as it may, what agent internalism claims is that nothing *else* about the agent's wider global environment plays a constitutive role in determining the justification-status of a given belief.

[20] At any rate, postclassical reliability is an agent internalist feature. According to transglobal reliabilism, postclassical reliability of the belief-forming process is a *necessary* condition for justification, but there might be other necessary conditions too that the belief-forming process would have to satisfy—e.g. speed, richness of outputs, and the like. But it is very plausible that any other such necessary conditions would be agent internalist features too.

to what counts as *suitable* modulational control—again, features psychologically internal to the agent. These factors having to do with the cognitive processes at play in a given episode, all of them psychologically internal, jointly suffice to determine whether or not the given belief-forming process employed by the agent is postclassically reliable.[21] Thus, postclassical reliability itself is a an agent internalist feature of a belief-forming process.

Needless to say, this fact marks a very significant difference between transglobal reliabilism and neoclassical reliabilism. The latter repudiates agent internalism in a striking way: it makes the justification-status of a belief depend upon various features of the agent's actual global environment that are independent of the agent's own psychology—namely, features in virtue of which the pertinent belief-forming process is, or is not, neoclassically reliable (i.e. globally reliable under suitable, globally reliable, modulational control).

Let us turn now to a second important aspect of similarity between transglobal reliabilism and access internalism, over and above their common allegiance to agent internalism. First-person conscious experiences are paradigmatic examples of psychological states that are directly accessible to an agent. These kinds of accessible states are not unimportant within standard forms of reliabilism, of course, because often they figure as important inputs to globally reliable belief-forming processes—and thus, often they are important *elements* in the total set of factors that determine the justification-status of a given belief. In this respect, transglobal reliabilism and global reliabilism are on a par. However, there is a further role for experiences in transglobal reliabilism, and a very significant one: namely, that the pertinent reference class for transglobal reliability is the class of *experientially possible* global environments. (Recall that these are global environments whose nature is compatible with the presence of agents with experiences much like those of actual humans—including demon-infested and vat-infested global environments in which the experiences of such agents are systematically nonveridical.)

This central theoretical role for accessible psychological states constitutes a substantial—albeit only partial—accommodation of the spirit of *access* internalism. Access internalists are right (according to transglobal reliabilism) in playing up the extreme importance of accessible psychological states in the justification of belief, even though they go too far in insisting that justification-status is determined only by matters that are accessible to the agent. Given a sufficiently

[21] There is a case to be made on this point, but not in this chapter: the transglobal reliability of a cognitive process in all its psychological richness is determinate, and thus is agent internal, given the character of everyday experience, which is itself agent internal. Central here are two thoughts. First, the set of relevant possible global environments is determined with respect to the general character of everyday experience—and that is agent internal. Second, what is agent internal—including the rich psychological chararacter of the processes in play, and including their input-output profiles—is enough to settle the matter of whether the processes in question would be reliable in a wide range of experientially possible global environments.

robust form of content internalism in philosophy of mind, access internalists are right to say, with respect to an agent who is a psychological duplicate of oneself but who is systematically deceived by a Cartesian demon (or who is a brain in a vat), that this agent's beliefs are just as well justified as are one's own beliefs. The agent's beliefs are equally well justified because of two correlative factors, both involving matters experiential (and neither of which involves non-psychological aspects of the agent's actual global environment). First, the agent's beliefs rest on the same *experiential basis* as yours, feeding into the same belief-forming processes; and second, these belief-forming processes are reliable relative to the reference class comprising the *experientially* possible global environments (and are under suitable modulational control by control processes that are likewise reliable relative to that reference class). Thus, the access internalist is right to think that (i) the agent's experiences render the agent's external-world beliefs objectively justified, and (ii) the justification-status of these beliefs is not at all impugned by the fact that the agent's belief-forming processes are utterly lacking in global reliability.

On the other hand, neither you yourself nor your deceived duplicate has full introspective access to the full body of psychological facts—including facts about the nature of the belief-forming processes themselves—that collectively constitute the agent internalist supervenience base for the justification-status of your beliefs. Full-fledged access internalism is repudiated by transglobal reliabilism, even though access internalism's emphasis on the role of experience in objective justification is given its proper due.

5. STRONG SEMANTIC EXTERNALISM AND TRANSGLOBAL RELIABILISM

We ourselves both reject strong semantic externalism: we both believe that an adequate account of mental intentionality and reference should have each of the following features. First, it should be mildly externalist, in the sense that it accommodates the ways that mental reference, for some thought constituents, is partly dependent on actual externalistic relations between thought constituents and the thinker's actual environment. But second, it should also have a significant internalist dimension too, so that it yields the intuitively natural results about the mental life of brains in vats and radically Cartesian-deceived beings—namely, that these beings have beliefs with extremely rich external-world intentional content, that those external-world beliefs are systematically false, and that many of the constituents of such beliefs that purport to refer (in particular, the constituents that purport to refer to concrete particulars or to natural kinds) fail to refer to anything at all. Constance's mental life, for instance, includes an auditory experience as of a phone ringing, and also includes a belief that someone is calling her and is thereby causing her phone to ring. Her belief that her phone is ringing

includes a thought constituent (expressible via inner speech as 'my phone') that purports to refer to a telephone, yet fails to refer to anything at all. (Phone-ish singular referential purport is an aspect of the intentional content of the belief.)

So we ourselves have no concerns regarding the intelligibility of any of the thought experiments employed in the present chapter. That said, however, we understand that advocates of strong semantic externalism may well claim to find the case of Constance and Faith to be unintelligible. They will perhaps maintain that one cannot make sense of the pervasive deception therein envisioned. To address their worries, we would like to make two points in conclusion.

First, at least for many kinds of strong semantic externalism, it should be possible to vary the pervasive deception scenarios in certain simple ways and thereby placate the intelligibility worries. Thus, one might imagine agents who get a good running start in life in a global environment with which they causally interact in ordinary ways, but who then get kidnapped in their sleep by Martians and spend the rest of their lives as envatted brains whose continuing mental lives fit together seamlessly with their pre-envatment mental lives. On the strong externalist account, this might well allow the agents to get enough concepts hooked up with referents that subsequent demon activity or envatting would be quite deceptive. If the deception comes into play with enough dogged persistence, one can reasonably say (as we said earlier) that there are no globally reliable processes to be had in such a global environment. Perhaps the case of Harry and Ike can be seen as a relatively mild variation on the family of pervasive deception scenarios.

Second, and more importantly, transglobal reliabilism can be motivated by thought experiments that are intelligible even to fans of of strong semantic externalism. Setting aside the case of Constance and Faith, other scenarios we have employed here are unproblematically intelligible by strong externalist standards—and these cases themselves already yield a strong case. Still, while not strictly *necessary* for our epistemological story, our repudiation of strong semantic externalism does help with *one,* but only one, of our motivating cases. Scenarios involving envatted brains or Cartesian evil demons, if intelligible as usually construed, do provide particularly compelling motivation for transglobal reliabilism. But even if one forgoes these cases, transglobal reliabilism can be motivated by the remaining thought experiments anyway. Indeed, we have emphasized how all the thought experiments discussed above reflect a concern for epistemic safety, and how epistemic safety is associated with robustly reliable processes (and suitable, associated, modulational control). Furthermore, without recourse to demon infestations and envatted brains, the case of Harry and Ike shows that the needed safety is not to be afforded by processes that are merely globally reliable under suitable modulational control. More robustly reliable processes—transglobally reliable ones—are constitutively required for objective epistemic justification.

REFERENCES

Henderson, D. and Horgan, T. 2000: 'Iceberg Epistemology.' *Philosophy and Phenomenological Research* 61: 497–535.

Horgan, T. and Tienson, J. 2002: 'The Intentionality of Phenomenology and the Phenomenology of Intentionality.' In D. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings* (Oxford: Oxford University Press).

—— —— and Graham, G. 2004: 'Phenomenal Intentionality and the Brain in a Vat.' In R. Schantz (ed.), *The Externalist Challenge: New Studies on Cognition and Intentionality* (Berlin: de Gruyter).

Lehrer, K. and Cohen, S. 1985: 'Reliability and Justification.' *The Monist* 68: 159–74.

Sosa, E. 1991: 'Reliabilism and Intellectual Virtue.' In E. Sosa, *Knowledge in Context* (Cambridge: Cambridge University Press).

6

Entitlement, Opacity, and Connection

Brad Majors and Sarah Sawyer

Central to debate over the nature of representational content is the question whether such content can be individuated without essential reference to nonintentional relations between the thinker and the objects or properties represented. Semantic externalism is the view that it cannot be so individuated. Its proponents hold that part of what it is to be a perceptual representation is to bear causal or other sorts of nonrepresentational relations to the relevant subject matter.[1] Semantic internalism denies this claim, and thus insists that a thinker's representational properties, states, and events supervene upon the intrinsic nature of the thinker. Physically type-identical subjects, on this view, necessarily share all the same semantic or representational attributes.

A leading issue in contemporary epistemology concerns the nature of epistemic justification or warrant. Epistemic externalism requires of a justified subject that her belief or beliefs be connected in the right kind of way to truth. According to these theorists, the subjective resources of the believer, though perhaps not wholly irrelevant epistemically, are of derivative importance. The various forms of reliabilism are the most prominent versions of this sort of doctrine. Epistemic internalism, by contrast, typically emphasizes the primacy and significance of that which is available, in some special way, to the believing subject. Here too internalism can be formulated as a supervenience thesis: Two believing subjects who do not differ in those features to which they have special access, in the relevant sense, cannot differ with respect to justification.

There is interesting, philosophically significant overlap and intersection between the semantic and the epistemological debates. It can be approached in different ways. We have argued previously that an adequate theory of justification must be externalist in character. Only thus can the constitutive connection between justification and truth be preserved. But refining epistemic externalism in such a way that it is immune to the most powerful objections requires that one be committed to the truth of semantic externalism. Reliability yields justification

[1] Attention will be confined here to perceptual representational content. While we think that the general externalist view may extend to all types of representation, this restriction will leave us with more than enough to discuss.

only when coupled with the sort of constitutive connection between content and world that semantic externalists have insisted upon.[2]

Here we take an alternative but complementary approach. We set the semantic and the epistemic debates within a framework structured by two general problems, which we label the 'Connection Problem' and the 'Problem of Opacity' respectively. The former is faced most immediately by internalist doctrines, while the latter is faced most directly by externalist doctrines. What is needed, and what we aim to provide, is an overall theory that combines both semantic and epistemic elements, and which steers a middle path between the general problems. The Connection Problem concerns the question of how a thinking subject relates to the outside world. In its semantic guise this is the problem of explaining how representation is possible. It is a potentially serious difficulty for semantic internalism to explain this possibility, given that it denies any essential connection between representation and that which is represented. Semantic externalism does not face this difficulty, given its insistence on just such an essential connection.

In an epistemological context it is the problem of how justification relates to truth. There are reasons for thinking that the relation is a constitutive one. The Connection Problem is a special problem for epistemic internalism, given that proponents of the view deny the necessity of any real connection between justification and truth. Epistemic externalists do not face this difficulty, inasmuch as they require (by definition) such a connection for justification.

The Problem of Opacity concerns the question of how a cognitive subject relates to herself. Here it is versions of externalism that appear to be on the defensive. In the semantic debate this is the problem of explaining how we know our own thought contents. For most or all versions of semantic internalism, the nature of content is taken to be transparent, in some strong sense or other, to the thinking subject. There would thus appear to be no special difficulty concerning access to such content. Semantic externalism, by contrast, insists upon a possible—and very often actual—gap between individual conceptions and the actual concepts with which one thinks. It is thus at least prima facie difficult to see how such a view can account for the authority with which we know our thoughts, or the privileged sort of access we have to them.

In the epistemic realm the Problem of Opacity appears as the problem of explaining how justification relates to that which is subjectively accessible. Again, internalists appear to be relatively well off here. For it is precisely endorsement of the claim that that which justifies one is and must be subjectively accessible, in some way or other, and on most versions of the view, which qualifies one as an epistemic internalist. Externalists, on the other hand, have not typically required any particular positive contribution from the subjective resources of a thinker to her epistemic status. What matters is reliability. It is therefore a problem for

[2] Majors and Sawyer 2005. The argument of that paper owes a great deal to Burge 2003a.

such theorists to explain how that which constitutes, or contributes to, epistemic standing can be outside the thinker's ken.

The key to unifying the debates, and to providing an adequate and principled solution to the problems, we shall argue, lies in Tyler Burge's notion of entitlement. An entitlement is an epistemic right possession of which does not consist in having reasons, or indeed in anything which is specially accessible to the believing subject. This notion is part of a more general rationalist conception of the way in which norms function to govern our cognitive (and perhaps practical) activity. After looking at the Connection Problem and the Problem of Opacity in more depth, we turn to the task of explaining how the notion of entitlement—or the more general view of norms that supports it—promises to illuminate and, in the end, provide the solution to the two problems. In the concluding section we suggest briefly that the strategy employed here has application as well to the pair of internalism/externalism distinctions within moral philosophy.

Throughout the discussion our aim is breadth rather than depth. By treating the relevant issues from a high level of abstraction, we inevitably skate over details that mark out particular theories; but in so doing we hope to make visible the central interconnections between the various forms of internalism and externalism.

1. THE CONNECTION PROBLEM

The Connection Problem is the problem of how the world and the subject relate. Within the theory of content it manifests itself as a challenge to internalism to explain how representation of an independent, objective subject matter is so much as possible, in the absence of any necessary or constitutive connection between that which is represented and the representational states of the thinking subject. As indicated above, *semantic externalists* face no special challenge here, because of their insistence upon precisely such a connection.[3]

For the *semantic internalist*, all of a subject's representational states and events could have been exactly as they are, even if she had never been in causal contact with any of the objects or properties that the states and events designate. If thought and reference are determinate, then, it must be the case that an individual's own nonrelationally specified cognitive resources—her phenomenology, so-called 'narrow' descriptions of entities she is disposed to offer, and so on—suffice in themselves to specify the referents of her terms and concepts, as well as the way in which she thinks of these referents. This must be the case if there is no necessary connection between world and thought.[4]

[3] Classic externalist works include Putnam 1975; Burge 1979, 1986. Of course, Putnam is not in this early paper an externalist about mental content. But he came to embrace the view.

[4] Versions of semantic internalism—we include two-factor theories under this general rubric—are defended by Dennett 1983; Fodor 1987; Loar 1988; and Segal 2000.

It is quite dubious, however, whether this set of ideas is so much as coherent. Set aside for the present the many thought experiments which are widely thought to refute semantic internalism. There are more basic conceptual reasons for thinking that the internalist's picture of the relation between world and mind is radically mistaken.[5] Suppose that a subject represents something, under a certain mode of presentation, as being a certain way. Perhaps she thinks that water is wet. If internalism is correct then something about the subject's own cognitive resources must make it the case that her representation is indeed of water, rather than something else. But now suppose that there is a liquid, twater, which the subject cannot discriminate from water. The crucial question for the semantic internalist is this: What makes it the case that our subject is thinking about water rather than twater? Since, by hypothesis, the subject cannot discriminate between the two liquids, and since internalism is committed to there being something in a thinker's cognitive repertoire that accounts for all the facts of representation, there appear to be only two options. Either the internalist admits that such a thinker does not in fact represent water rather than twater, or she denies the possibility of the sort of indistinguishability we have been describing.

Neither move is plausible pre-theoretically. We take the latter to be obviously untenable. So far from being impossible, there are actual cases of such indistinguishability. Jadeite and nephrite provide one example. And the former gives up altogether on the idea of objective reference. A thought about water is not a thought about the infinitely many possible liquids which are superficially indistinguishable from water. It is about water. Representation is representation *as such*.

A related point is that this internalist move makes a fundamental error about the *function* of representation. In all minded animals, nonhuman as well as human, representational systems evolved to allow the organism to deal with its environment. The environment, in every case, contains particular kinds of things. Our environment, for example, contains water but not twater. Thought has a representational function in presenting truth. It is not true that our environment contains twater. It is implausible, therefore, that in everyday thoughts ascribed with 'water' in oblique position, we represent twater—or any of the infinitely many other distinct liquids superficially indiscernible from water.[6]

Semantic internalism is thus poorly placed to solve the Connection Problem. If there were no necessary or constitutive connection between world and mind, thinkers would be unable to represent objects and properties in the world as such. The Connection Problem does not arise for semantic externalism. The reason our subject thinks about water rather than twater, for the externalist, is that she has been in causal contact with (or bears some other relevant nonintentional relation to) water, but not twater. On this view, which was the dominant one

[5] For reasons not mentioned here see Sawyer (forthcoming).

[6] Most of these points against internalism have been stressed by Burge. See, among other papers, his 2003a, 2003b.

in philosophy until Descartes's reorientation, thought is not independent of the way the world is.[7]

There is an important Connection Problem for *epistemic internalists* as well. This is the problem how justification or warrant relates to truth. There are strong reasons—conclusive reasons, in our view—for thinking that the two are essentially or constitutively connected. We will mention two of them. First, this relation to truth is at least a large part of what distinguishes epistemic from other sorts of justification, such as moral or prudential justification. If a consideration does not bear in any way on the truth of a content, then it does not provide epistemic warrant for belief in that content. Second, it is a commonplace that belief aims at truth. Truth provides the representational norm for the propositional attitude of belief. Justification, or warrant, is a property of belief, and provides its epistemic norm. It plausibly accrues to a particular belief only insofar as the latter is well placed to achieve its representational norm of truth. This is because belief's primary epistemic norm is a measure of how well an organism is doing, given its capacities and limitations, with respect to its function of accurately representing the world. But its representative function just is to present truth. Any adequate account of epistemic justification or warrant must therefore respect the constitutive connection between epistemic norms and truth.[8]

Epistemic internalism is, roughly, the doctrine that justification supervenes upon the subjective resources of the believer, or that to which the believer has special access. Some versions require that a justified subject have, and base her belief on, reasons; others that she have evidence, and base her belief on it; still others merely—for perceptual belief, anyway—that her belief be based on the relevant sort of experience. Many versions require only that a subject be acting, in forming her belief or making a judgement, in an epistemically responsible manner.[9] The problem for internalism is that none of these conditions suffices to place belief in the necessary relation to truth, the relation claimed above to amount to fulfillment of belief's central epistemic norm. The reason for this is that a believer's subjective resources, or those things to which she has special access, do not in general, and taken by themselves, suffice to put her in an objective relation to truth.

Basing one's belief that p upon reasons is insufficient for justification. The reasons might be quite poor. They may have no tendency whatever to show that p is true. It is not open to the internalist, furthermore, to require that one base

[7] Actually it is not at all clear that Descartes himself was a semantic internalist; see Burge 2003c. Nevertheless he did inaugurate the internalist tradition, both in philosophy of language and mind, and in epistemology.

[8] Cf. Burge 2003a: s. I.

[9] Versions of epistemic internalism, all of which combine one or more of the elements we have separated out here, are put forward by Bonjour 1985; Feldman and Conee 1985; Moser 1985; Steup 1988; and Pryor 2000.

one's belief on *good* reasons. For either one can specify what counts as 'good' without invoking alethic considerations, or one cannot. If one can, then having good reasons will not suffice to put a (putatively) justified believer in the requisite relation to truth. And if one cannot, then the view will have lost its qualifications as internalist, inasmuch as a relation to truth is required, after all, for justification.

It is important to notice that requiring that a subject *believe* her reasons to be truth-conducive is inadequate. This is for two reasons. First, it makes higher-order cognitive capacities—the ability to think about reasons, beliefs, truth—essential to having justified beliefs. It is neither philosophically nor empirically plausible that this is the case. Second, it is widely acknowledged in the contemporary literature that epistemic responsibility is insufficient for justification.[10] But being epistemically responsible just is believing on the basis of what one takes to be truth-conducive reasons.

It is worth taking a moment to reflect here. While many today, including some internalists, acknowledge that epistemic responsibility is insufficient for justification, no internalist to date has offered an adequate explanation of why this is the case. We think that the explanation lies in the constitutive connection between justification and truth. To have an epistemically justified belief is to have a belief that is well placed to achieve the representational norm for propositional attitude content—truth. But being epistemically responsible does not suffice for being thus well placed. This is because one's cognitive or perceptual faculties, including one's ability to reason, might be seriously defective. This would certainly affect, in a negative fashion, one's ability to form true beliefs. But one might nevertheless be doing as well as could reasonably be expected, given one's circumstances. Therefore one could be as epistemically responsible as possible and still be in a poor position to achieve truth, even in one's normal environment. This is why epistemic responsibility does not suffice for justification.

If this explanation is correct then epistemic internalism is inherently problematic. This is not because all versions of internalism require only epistemic responsibility for justification. Many versions do not focus on the notion of responsibility. But the explanation offered above appeals to an objective sort of connection between justification and truth that the internalist cannot countenance. Offering a detailed defense of the explanation and its constituent claims is not our purpose here. What we do wish to emphasize is the need for internalists to explain why epistemic responsibility does not suffice for justification. Mere acknowledgement of the fact is not enough. We think that epistemic internalism, at least as classically conceived, cannot survive recognition of the fact that epistemic responsibility is insufficient for justification.

The same basic problem afflicts views that require for justification only that one's belief be based upon evidence, or perceptual experience. Justification

[10] See Plantinga 1993: chs. 1 and 2: and also Pryor 2001: 112–14. Pryor's paper contains additional relevant bibliographic information.

requires that one be connected in the right sort of way to truth. But having evidence, or perceptual experiences, does not suffice to put one into any objective relation to truth. This is because the evidence could be poor, or radically misleading; and the perceptual experience might be wildly inappropriate to that which causes it. Again, while it might be the case that requiring that the evidence be good, or the perceptual experience appropriate, suffices to overcome this difficulty—and thus to put one into the right sort of connection to truth—this is not a move open to the epistemic internalist. For what it is to be 'good,' or 'appropriate,' in the relevant senses, is to link up in the right sort of way to truth. To require such a link is precisely to abandon internalism.

We think that these facts about justification have been obscured by a fallacious line of reasoning. It is quite tempting to argue as follows: My recently envatted twin and I are both justified in many of our perceptual beliefs; but she is almost completely unreliable in forming such beliefs; therefore justification cannot require any objective connection with truth. We have called this the 'Internalist Fallacy' (2005: 268). It is a fallacy because requiring of justified belief an objective connection with truth does not imply that one's belief-forming mechanisms must be reliable in any environment into which one might be placed. No relevant belief-forming mechanism is reliable in all possible worlds. Some proper subset of these worlds must be privileged. The challenge to the theorist of justification is to say which subset is crucial for justification.

Needless to say, the various forms of *epistemic externalism* do not suffer from the Connection Problem; for they require reliability, of one form or another, for justification.[11] They do, however, face the Problem of Opacity, to which we now turn.

2. THE PROBLEM OF OPACITY

The Problem of Opacity concerns how the cognitive subject relates to herself. In the semantic realm it takes the form of the problem of accounting for our knowledge of our own thoughts, in particular their intentional contents. It has been common at least since Descartes to hold that one's knowledge of one's own thoughts is special, in at least two ways. It is authoritative, since there are definite limitations on the sorts of errors that one can make.[12] There are no such limitations with respect to one's knowledge of the empirical world. Further, one appears to have a privileged kind of access to one's own thoughts. One does not normally have to observe one's behavior in order to know what one thinks. One can simply reflect on the matter. One's knowledge of the contentful mental states of others, by contrast, has almost always been taken to be necessarily based upon behavior.

[11] Forms of epistemic externalism are defended by Goldman 1986; Sosa 1991; and Burge 2003a.
[12] See Burge 1988, in particular the discussion of brute error at 657 ff.

It is widely held that there is a special problem for *semantic externalism* here. This view, as noted, holds that what determines the nature of thoughts about an extensive range of entities and properties lies partly outside the thinking individual's own physical make-up and discriminatory abilities. It implies that two physically and phenomenologically (intrinsically) indistinguishable thinkers can differ with respect to their intentional mental states, given the appropriate sorts of variation in the relevant speech communities or physical environments. But if what determines the nature of one's thoughts is often not something accessible, in the relevant respect, to the thinking subject herself, then the special nature of our knowledge of our own thought contents seems to be put in jeopardy. It is thought to be unclear how, on the externalist conception, knowledge of one's own thoughts can be authoritative in the sense mentioned above. For one is not authoritative about the external determinants of these thoughts. And it is often claimed to be mysterious how one could have the sort of privileged access to one's thoughts that much of philosophical tradition has thought to obtain, given that one's access to that which partly determines the nature of one's thoughts does not have this sort of privilege. Externalism appears to make thought content opaque to the thinker.

It is sometimes taken for granted that *semantic internalism* does not face the Problem of Opacity. For this view is precisely the denial that anything outside the milieu of the subject herself can be relevant to determining the contents of her thoughts. But matters are not nearly so straightforward. For one thing, internalism is often formulated as a local supervenience thesis: No difference in thought content without a difference in intrinsic physical properties. But one does not have any special authority over, or any privileged access to, one's intrinsic physical properties. A more important problem is that no viable internalist model of self-knowledge has ever been offered. There are very serious difficulties facing any attempt to construct such a view.[13] So it may be granted that the Problem of Opacity is particularly urgent, at least at first glance, for the semantic externalist; but until a coherent internalist model of self-knowledge is made available, the internalist herself has a good deal of work to do here.

Epistemic externalism also faces the Problem of Opacity. Here the difficulty is understanding how a subject's belief can be justified or warranted when the subject herself has no reasons supporting the belief. More generally, it is the problem of explaining how elements outside of a thinker's ken can make it the case that her beliefs are justified. The problem is at its starkest for pure reliabilism, the view that all that is required for epistemic justification is the reliability of one's belief-forming processes. Cases in the literature show that pure reliability is neither necessary nor sufficient for epistemic justification.[14] It is not necessary, because one's recently envatted twin presumably shares one's epistemic status

[13] See e.g. the discussions in Burge 1988 and Davidson 1987.
[14] See for example Bonjour 1985: ch. 3.

with respect to perceptual beliefs; but she is almost completely unreliable with respect to such beliefs. It is not sufficient, because one can imagine a case in which things are set up in such a way that beliefs about a given subject matter simply pop into one's head, as it were. It may be the case that one is completely reliable about the subject matter; perhaps a meddling super-scientist sees to it that one is always right. Nevertheless one is intuitively not justified with respect to these beliefs. In each case the problem for pure reliabilism is that putatively epistemic elements which are opaque to the believing subject appear not to make the relevant sort of epistemic difference (or perhaps any epistemic difference at all). Whether these objections apply to other forms of reliabilism depends upon the details.

Epistemic internalism, at least as classically conceived, does not face the Problem of Opacity. For it is precisely the view that only elements within one's ken make an epistemic difference. Some recent forms of internalism bear little obvious relation to the classical version, of course, and so it should perhaps be left open whether they might not face a version of the problem. But we will not pursue the matter here.

We have not attempted to conduct a dispassionate survey of the various problems facing internalist and externalist accounts of content and justification. The reader will see where our sympathies lie. And we have looked only at those problems which seem to us most basic, and (relatedly) which are most important in seeing how the two internalism/externalism controversies intersect. Of the two problems we have discussed, the Connection Problem is the most fundamental. Only if it is solved will a theorist have accounted for the *very possibility* of representational content, in the semantic realm, or of epistemic justification, in the epistemic realm. The Problem of Opacity concerns how a subject relates to her representational contents, or, in the epistemic realm, how that which is accessible to a subject bears on her epistemic status. Without the relevant sort of connection to the world in either case there can be no such thing as representational content, on the one hand, or epistemic justification, on the other. Consequently there can be no problem of explaining how a subject's own perspective relates to these things.

This means that externalism, both semantic and epistemic, is at a clear dialectical advantage. Only the semantic externalist, we believe, has any chance at all of accounting for the possibility of representational content; and only the epistemic externalist has any chance of accounting for epistemic justification. Problems of opacity are certainly important; but they arise only secondarily.

3. ENTITLEMENT

We have said that we take Burge's notion of entitlement—or, better, the objectivist conception of norms which underlies it—to be the key to solving the two problems. We hold in addition that reflection upon the issues raised by

this conception points the way toward an illuminating general framework for understanding the ways in which the two internalism/externalism controversies interact and impact one another. It is the business of this section to explain these claims in more detail.

We begin by distinguishing the notion of entitlement we will be employing from the various other conceptions which have appeared in the literature in recent years. The term 'entitlement' is used in a number of distinct, even incompatible ways. One must be careful to understand what precisely is the property or relation that a given thinker associates with it, lest confusion arise. Discussion of the ways in which these alternative conceptions fall short of grounding the genuinely epistemic property we are interested in also leads naturally to consideration of the way in which our view handles the Connection Problem.

Conceptions of entitlement are most naturally divided into those which purport to be epistemic—and thus to relate the believing subject to truth, in some way or other—and those which do not. While we do not take seriously conceptions of entitlement which are not epistemic, they are mentioned here for the sake of orientation and completeness.

Paul Boghossian has employed a conception of warrant, which he sometimes calls 'entitlement', that does not require the possession of reasons or a rationale on the part of the believing subject (Boghossian 2000, 2003). His primary concern is with our justification for using basic rules of inference. Familiar Carrollian and Quinean considerations suggest that our most fundamental and basic patterns of deductive inference cannot themselves be deductively justified. It is obscure, furthermore, how there could be an inductive (or even abductive) justification for our use of them. Yet we are intuitively justified in such use.

Boghossian's idea, the main elements of which he takes from earlier work by Christopher Peacocke (1992), is that we need not possess a classical internalist sort of justification for such practice, if the practice bears the right sort of intimate relation to possession of the concepts involved in reasoning. More specifically, one is entitled to a given pattern of inference if employment of that pattern is constitutive of possessing the associated concepts, concepts we necessarily employ in reasoning.

> What I am urging is that that entitlement is precisely what flows naturally from a conceptual-role account of the meanings of our logical words . . . [that] our problem about our entitlement to employ a rule of inference reduces to that problem [of what makes a rule meaning constituting], a problem that any conceptual-role semantics faces. (Boghossian 2000: 250)

There is more wrong with this line of thought than we have the space to go into here.[15] The central problem is that the fact—if indeed it is a fact—that we must employ certain patterns of inference in order to possess the relevant concepts in

[15] Two further problems: It is highly questionable whether conceptual-role accounts of meaning and content, for whatever sorts of terms or concepts, are defensible; and it is doubtful whether there are many (or even any) concepts the mere possession of which requires that one be disposed

no way shows that we are epistemically warranted in such employment. Epistemic warrant bears an essential and constitutive connection to truth. An epistemically warranted belief must be well placed, in some sense to be specified, to achieve belief's representational norm of truth. An epistemically warranted deductive practice must be well placed to achieve the preservation of truth. But being forced, by the nature of the concepts involved, to engage in a given practice in no way shows that the practice leads, or even is likely to lead, to the preservation of truth. It can at most show that one is not being epistemically irresponsible in engaging in the practice. And epistemic responsibility, as we have already argued, and as is now widely recognized, is quite insufficient for epistemic warrant.[16]

More or less the same holds for several other recent accounts of entitlement, including those by Fred Dretske, Crispin Wright, and Martin Davies. Dretske follows Burge in taking an entitlement to be an epistemic right that does not require the possession of reasons.[17] But his explanation of why we are entitled to the relevant range of beliefs has nothing to do with truth. Dretske is impressed by the fact that many of our perceptual beliefs are formed spontaneously and automatically. They are not under our control. He infers from this that we are blameless in holding them, despite the fact that we have no reasons for them, and can offer no relevant rationale. A conception of warrant on which being warranted does not entail being well placed to achieve belief's representational norm, truth, is simply not epistemic. It can have nothing directly to do with epistemic warrant or justification.

Wright has developed several ostensible conceptions of epistemic entitlement (2004a, 2004b). He thinks that when we act in accordance with certain dominant practical strategies; when we accept propositions which are fundamental to developing important cognitive projects; when we accept that which is indispensable to decision making and acting; and when we accept that without which no sense can be made of our enquiries as investigations into an objective subject matter—we are entitled to do so despite our lack of justification in the traditional sense. As these descriptions of the conceptions indicate, it is the apparent fact that without acting or accepting in the ways in question we cannot engage in cognitively and practically fundamental projects that, for Wright, warrants our so acting and accepting. But again, none of this has anything to do with truth. It is easy to imagine creatures who cannot but judge and behave in certain

to make certain specific inferences. On these points see especially Burge 1986 and Peacocke 2004: s. 6.2. Cf. also Williamson 2003.

[16] In related work Boghossian explicitly affirms that his topic is justification in the sense of epistemic responsibility, despite the fact that mainstream epistemology largely rejects the view; see Boghossian 2002: 40–1. He does not discuss any of the reasons for this rejection. Wright 2002: 60–1 follows him here, which partially accounts for the problems with Wright's view of entitlement mentioned below. The general difficulty is mentioned by Williamson 2003: 250.

[17] In his 2000, M. Williams takes Dretske to mean by 'entitlement' R. Brandom's notion (briefly discussed below); this despite the fact that Dretske explicitly claims to be following Burge's very different understanding. The error renders much of Williams's critique impotent.

ways—and who, consequently, must so judge and behave in order to carry out their cognitive projects—but who are nevertheless so poorly adapted to their normal environments, and who reason so poorly, that they completely lack epistemic warrant. Wright's various conceptions of entitlement do not therefore have any epistemic relevance.[18]

Davies is sensitive to some of the problems in Wright, and he has tried to salvage what he can by developing a notion of negative entitlement. For Davies the notion of epistemic warrant is closely tied to that of cognitive achievement. He thinks that there are entitlements which are not warrants, because they do not stem from any achievement on the part of the believing subject. We are entitled, more specifically, not to doubt those propositions which we lack antecedent reason or warrant for doubting.[19] The problem with this is that the mere fact that we lack reasons for calling into question a certain belief or judgement by no means shows that we have an epistemic right to it. Here is one way to make the point: Either we have epistemic warrant, in the first place, for the belief or judgement, or we do not. If we do, then whatever grounds the warrant is the source of our epistemic right—not the fact that we lack reasons for doubting it. If we do not, then we have no right to the belief or judgement in the first place, and the fact that we have no articulable grounds for doubting it is wholly beside the point. Like appeals to judgement-patterns which partially constitute the possession of certain concepts, and to cognitive behavior which is not epistemically irresponsible, or that we cannot avoid, Davies's conception does not connect the relevant beliefs or cognitive practices of subjects to truth. It therefore has no bearing on the issue of specifically epistemic entitlement which is our concern.[20]

It bears emphasizing that we do not dispute the particular claims these writers make concerning which are the beliefs, practices, and judgements to which we are entitled. We agree that thinkers are typically entitled to basic logical inferences, and to certain beliefs for which they can offer no justification

[18] Wright shows some sensitivity to this problem, which partially accounts for the fact that he focuses on the attitude of accepting that p, rather than believing that p; see 2004b: 175 ff. But apart from the fact that this immediately renders his proposal largely epistemically irrelevant, it is a confusion to infer from the fact that one lacks evidence for believing that p, that one cannot be epistemically warranted in believing that p. Wright, in common with others (including Davies; see below), has failed to understand the force of Burge's conception of epistemic entitlement. We elaborate upon this point later in the section.

[19] Davies 2004. Davies attempts a kind of synthesis of the views of Burge and Pryor 2000; see Pryor 2000. Though Pryor does not use the term 'entitlement' (see n. 24), his notion of immediate justification is similar in certain respects to that of entitlement. Like the other putative explanations of entitlement surveyed here, however, Pryor's has nothing to do with truth. He thinks their peculiar phenomenology is what explains our immediate justification for certain perceptual judgements. We explain why this is inadequate, and also go into more detail on Dretske and others, in Majors and Sawyer 2005: s. IV.

[20] The same problem afflicts H. Field's notion of 'default reasonableness'; see Field 2000: 119 ff. A default reasonable proposition, for Field, is a proposition which can be reasonably believed without any evidence. His explanation makes clear that the notion has nothing to do with truth. It is therefore not a conception of warrant, or of any other epistemically significant property.

or rationale. We agree that we are entitled to certain very basic beliefs and practices which we have no reasons or rationale to support. Our disagreement concerns the *explanation* of the various entitlements cognitive subjects possess. Explaining epistemic entitlements perforce requires showing how the beliefs or practices in question tend toward truth, or truth-preservation, at least in normal circumstances. Non-epistemic accounts of entitlement, such as those we have just reviewed, precisely fail to do this.

There are, apart from the Burgean notion considered below, two extant conceptions of specifically epistemic entitlement; that is, conceptions which connect entitlement with truth, in a constitutive way. Robert Brandom uses the term 'entitlement' to pick out a certain sort of epistemically deontic status. His view is, nearly enough, that one is entitled to that which one's linguistic cohorts will let one get away with saying. 'Entitlement is, to begin with, a social status that a performance or commitment has within a community' (1994: 177). One is entitled to a given belief or judgement just so long as one is not challenged to justify it. This is an understanding with affinities to Davies's notion of negative entitlement. It is of course a question why we ought to consider it a specifically epistemic understanding, given that, as we have emphasized, epistemic warrants or entitlements must connect belief with truth. From a pre-theoretic standpoint, the issue of what one can get away with saying has precious little to do with what is true. The answer is that Brandom has a radically anti-realist conception of truth. To a first approximation, being true for him just is being agreed upon in a certain sort of way. There is therefore a technical sense, perhaps, in which his conception is epistemic. But we mention the view only to set it aside. We think this sort of anti-realism indefensible.[21]

The other alternative to Burge's understanding of specifically epistemic entitlement is put forward in recent work by Christopher Peacocke. The necessary connection between entitlement and truth is enshrined in Peacocke's first principle of rationalism, according to which a transition is one to which a thinker is entitled only if it is truth-conducive, in a specific sort of way that Peacocke develops (2004: 6 ff.). This view of entitlement is taken initially from Burge, and is largely compatible with our Burgean view. But there are difficulties. The most important of these is Peacocke's insistence that judgements or transitions to which we are entitled must be rational from a thinker's own point of view. This claim is incompatible with the fact that some higher non-human animals, and most very young human children—members of both of which groups lack the concepts necessary to think contents concerning that which is rational—are entitled to many beliefs for which they have no reasons. Nor is it compatible with Peacocke's own insistence that entitlement accrues to a thinker through a

[21] Brandom may prefer to put the point by saying that 'true' functions in a prescriptive, rather than a descriptive, way. But the problems remain. For a recent critical treatment, with which we are in sympathy, see Peacocke 2004: s. 1.2.

particular sort of truth-conduciveness in her beliefs or transitions. That which one takes to be rational has, in many cases, nothing to do with what is actually rational or warranted, because it has nothing to do with what is truth-conducive in the relevant kind of way.[22]

We have said that entitlement on Burge's understanding is an epistemic right possession of which does not consist in having reasons for one's belief. The sense in which reasons warrant belief is fairly clear. To have a reason for one's belief that p is to have another belief, or set of beliefs, which lend support to the former; support is provided when it is probable (or, in the deductive case, necessary) that p is true, given the truth of the supporting belief or beliefs. Because epistemic entitlement does not involve reasons, there must be some alternative way of showing, in a given case, that the beliefs to which one is entitled tend toward truth. There must be another way of showing, that is, that one has an *epistemic right* to the range of beliefs in question. What this way is differs from case to case.[23]

Notice that it is a mistake to think that those who distinguish entitlement from justification are attempting to deal with difficult epistemological problems by linguistic fiat or stipulation. The work done in an account which appeals to entitlement is not done merely through the introduction of the term 'entitlement'. The term is introduced to mark the important fact that not all warrants or epistemic rights are conferred upon beliefs in the same way. Some warrants involve reasons and evidence; others do not. The various kinds of warrant are different enough that it is helpful to mark the distinction linguistically. But the real work done by an account of entitlement comes in explaining why it is that we are entitled to the class of beliefs in question. It comes in connecting those beliefs with truth.[24]

[22] These and other problems with Peacocke's most recent book are discussed in Majors 2005.

[23] The term 'entitlement' first appears in Burge's work in Burge 1993. But the notion developed out of earlier historical work Burge had done on Frege, in particular the essays collected in part III of Burge 2005. See also Burge 1996, 1998a, 2003a. Burge uses the term 'warrant' to cover both entitlements and justifications—where the latter essentially involve reasons. Unlike Plantinga's conception, Burgean warrants are defeasible.

[24] Pryor complains about the popularity of the term 'entitlement.' He says, more specifically, that 'Burge, Peacocke, Davies and Dretske . . . prefer the terms "entitlement" and "warrant" to "justification." I insist upon "justification." It ought not to have the undesired associations they hear it to have.' (See Pryor 2004: 371–2.)

There are a couple of problems with this. First, while the claim may be true of Peacocke, it is false of the others that Pryor lists. Dretske, following Burge, reserves the term 'entitlement' for a specific kind of epistemic warrant. Each of them recognizes justification as a distinct, though still perfectly useful, concept. There is no sense in which they *prefer* 'entitlement.' Davies differs from Dretske and Burge in thinking that entitlement is not a form of epistemic warrant. But even for him it is simply false to say that he prefers 'entitlement' to 'justification'. The terms have disjoint and non-empty extensions. (We think that Pryor has confused the views of these thinkers with the very different sort of claim, associated with W. Alston, and also Plantinga, to the effect that the term 'justification' has outlived its usefulness.) Second, whether a term *ought* to have undesired associations is surely not to the point. What matters is that it does (when it does).

Our present interest is less in the details of the particular legitimizations or deductions of our entitlement to various classes of beliefs, than in the objective view of norms which supports the general idea of an entitlement. How can it be possible to have warranted belief in a given content when one has no reasons, and can offer no rationale, for the belief? The key point to notice is that, as beings who represent an objective subject matter, we are necessarily subject to representational and epistemic norms. These norms are objective in character. It is an a priori truth that representations function accurately to depict a subject matter (whatever other functions they may be given). All representational contents have veridicality conditions. The sort of veridicality at issue with propositional content is truth. The representational norm—or correctness condition—which governs belief, as a cognitive propositional attitude, is therefore truth. Beliefs fulfill this norm only insofar as they are true.

It is widely agreed, however, that a belief can be correct, in another sense, without being true. One can believe what one ought epistemically to believe without one's belief being true. Epistemic warrant does not entail truth. Epistemic warrant is a measure of how well placed one's beliefs are, given one's limitations and cognitive capacities, to meet their representational norm.

Meeting the norms of belief, therefore, both representational and epistemic, is a matter of relating in the right sort of way to truth. But one can bear the right sort of relation to truth, in many cases, without having reasons for one's belief. The objective norms which govern representation determine whether one is warranted in a given belief. And one can fulfill these norms, in at least many cases, without having reasons for one's belief.[25] We say more about how this works, in the perceptual case, just below.

Objective norms also have a place, we believe, in the individuation of intentional mental states and events. Ways of talking and thinking about objective matters are necessarily responsible, in at least many cases, to the ways in which others employ terms and concepts; and, more fundamentally, to the natures of the entities and properties represented. This shows up not only in the possibility that an individual can be mistaken about what one of her words means, or in the explication of one of her thoughts or concepts. It is manifest also in the possibility that entire communities can make errors concerning the natures of many kinds of concepts.[26] These points will be elaborated upon below.

4. SOLUTIONS

We now set out our solutions to the Connection Problem and the Problem of Opacity against this background picture of the way in which our thinking is

[25] Cf. Burge 2003a as well as the other papers mentioned in n. 23.

[26] See Burge 1986 and 'Frege on Sense and Linguistic Meaning', in Burge 2005.

governed by objective norms. In the next, penultimate section, we will discuss the sense in which the picture provides a unified way of viewing the internalism/externalism debates in semantics and epistemology. More speculatively, we will suggest in closing that the other two main internalism/externalism debates in philosophy—those in ethical theory—can themselves be seen to conform to the schema we set out for understanding the corresponding debates in semantics and epistemology.

Because our view is externalist both in semantics and epistemology, the Connection Problem is not a serious difficulty. Take first the semantic case. We think that representation of any objective empirical subject matter requires causal, or other sorts of nonintentional relations to obtain between representer and that which she represents. Paradigmatic cases of perceptual representation involve causal contact between the thinker and the subject matter of her representations. Now the possibility of representing that which does not exist shows that not all cases fit this simplest sort of model. One might perhaps be mistaken about the existence of water, for example, despite the fact that one has many thoughts containing the concept *water*. But it is plausibly necessary that one bear the requisite sorts of nonintentional relations to *some* entities in order to possess the concept *water*. One could presumably not have the concept were one not in causal contact either with water, or with some thinker or thinkers who have theorized about the existence of water.[27] Empirical representation as such, then, requires the obtaining of causal relations between the thinker and that which is represented, or at least between the thinker and some other related entities or properties. A deep, individuative connection is therefore a necessary condition upon the very possibility of perceptual representation.

In the epistemic realm externalism is equally well placed to solve the Connection Problem, though here certain difficulties arise. The problem is to account for the constitutive connection between warrant or justification and truth. As we saw earlier, only such a connection allows us to distinguish between epistemic and other kinds of justification; and only such a connection allows us to account for the fact that belief necessarily aims at truth. To say that some constitutive connection between belief and truth is necessary for justification is to say that justification requires reliability. But it is not a simple matter to say in what sense a belief-forming process must be reliable in order for its deliverances to be epistemically justified.

As already indicated, the problem is that reliability in the world the subject happens to inhabit is neither necessary nor sufficient for justification. It is not necessary, because the perceptual beliefs of one's twin, who was recently transported into a world controlled by a malevolent demon, are intuitively as justified as are one's own. She has all the same perceptual experiences, and appears to have the same evidence more generally. But she is almost completely

[27] Cf. Burge 1982.

unreliable in forming perceptual beliefs. And this sort of reliability is not sufficient for justification, because it is easy to imagine a subject who is accidentally reliable, and therefore not justified.

There are a number of ways of attempting to overcome these difficulties. We will not review them here.[28] Our contention is that the difficulties for pure reliabilism can be overcome only by invoking anti-individualistically individuated representational content. It is true that reliability in the world the subject happens to inhabit is not necessary for justification. What is necessary, however, is that a subject be reliable in the kind of world relative to which the natures of her perceptual intentional contents are individuated and explained. And it is true that reliability is not sufficient for justification. What is sufficient, though, is reliability coupled with the sort of constitutive connection between representational state and world that the semantic externalist insists upon.

A potential problem for our view is the Problem of Opacity. Indeed, the more easily a view overcomes the Connection Problem, the more difficult it is to account for the relation between the subject and herself. Within semantics the problem is to explain how we could have privileged and authoritative access to our intentional mental states, when the very natures of those states are dependent upon relations to the environment for their individuation; relations to the obtaining of which one has no special or authoritative access. Our Burgean response comes in two parts. First, there is a species of self-knowledge, called 'basic self-knowledge', which we are guaranteed to have on purely formal or structural grounds.[29] These cases are analogous to the cogito cases that so interested Descartes. Take, for example, the judgement 'I hereby judge that writing requires concentration.' This is a second-order claim about one's mental states. One is not only thinking the relevant thought, but judging that one is thinking it. Now the notions of writing, requirement, and concentration are all plausibly externally individuated. There is no guarantee that a subject, in thinking contents containing these notions, has an adequate or complete understanding of their essential natures. She may well make mistakes about them. The question thus arises how she can know what she thinks, when she thinks with these notions, given that she has no privileged or authoritative access to their real natures.

In the ordinary sense of self-knowledge, one knows what one thinks when one knows the contents of the relevant mental states. And with cases of basic self-knowledge, one's knowledge of one's first-order mental state contents is logically guaranteed by one's very exercise of the second-order abilities. One cannot make the second-order judgement that one is judging that writing requires concentration without thinking the content that writing requires concentration.

[28] These ways are discussed in Majors and Sawyer 2005: s. V. That paper contains a more complete statement of the argument for our version of epistemic externalism. Here only a brief summary is provided.

[29] For this thesis, and the arguments to come, see Burge 1988; cf. also Sawyer 2002.

The second-order judgement would not be what it is, were its subject matter a different first-order content.

Of course, one can imagine variations in one's social and physical environment which would, after a while, occasion changes in one's first-order propositional contents. But if the circumstances are different enough to make it the case, for example, that one does not, in uttering the word-forms 'Writing requires concentration', judge that writing requires concentration, then they are different enough to change the character of one's second-order judgement as well. It is quite impossible for there to be disconnection between first- and second-order contents, in cases of basic self-knowledge; this is, again, because the natures of the first-order contents partially constitute the second-order contents. There could be no variation in the former without variation in the latter. They are logically locked onto one another.

Not all self-knowledge is basic in this sense, however. The majority of cases of self-knowledge are not logically self-verifying. We can and sometimes do make mistakes about the contents of our thoughts. But it does not follow from this that nonbasic cases of self-knowledge are empirical, in the sense that we have no authority over, or privileged access to them. The explanation of our entitlement to nonbasic cases of self-knowledge has to do with the necessity of such knowledge to the very enterprise of critical reasoning.[30] If we were not normally *correct* about our intentional mental contents; if we were not normally *entitled* to the higher-order claims and judgements we make about them, and with them; and if, further, we did not normally have *knowledge* of our first- (and, more generally, lower-) order intentional mental states—then we could not engage in the practice of critical reason: The practice of subjecting one's reasoning to rational review, and of acting upon such review. Since no coherent form of skepticism denies the possibility of critical reasoning, this suffices to ground nonbasic cases of self-knowledge.

Some have apparently held that in order properly to be said to know one's thoughts one must be able to distinguish the components of these thoughts from all possible counterfeits. But externalism seems to foreclose on this possibility. For there is a sense in which one cannot tell, as it were from the inside, whether one is thinking with a concept C, or a concept C* which picks out entities or properties subjectively indistinguishable from those in the extension of C. If therefore it is indeed a condition upon the possibility of self-knowledge, as traditionally construed, that one be able to distinguish one's thought components from all possible counterfeits, then externalism seems to rule out the possibility of such knowledge.

We think it is clear that this is not a reasonable requirement. No one thinks that our perceptual knowledge is impugned by the fact that we cannot distinguish the objects of our perception from all possible subjectively indiscernible yet

[30] See Burge 1996 and also 2000.

distinct alternatives. The objectivity of perception guarantees that there are such alternatives. More to the point, though, is a simple distinction between what our thoughts are, and that which makes them so. One normally knows the identity of one's thoughts, for reasons covered a few paragraphs back. In cases of basic self-knowledge, one knows the identity of one's thoughts in the very act of thinking the relevant higher-order contents. The former partially constitute the latter. One cannot therefore make a mistake about them. In nonbasic cases, while error is possible, knowing the contents of most of one's thoughts is necessary to the very functioning of critical reason. It is simply a confusion to think that one must know, in order to know what one is thinking, all of the background conditions that must obtain in order for one to think one's thoughts. This is made manifest by considering knowledge of kinds in special sciences other than psychology. It would be ludicrous to insist that one must know what sorts of functional relations individuate hearts, in order to know what a heart is, or that one has a heart. Knowing one's thoughts does not require either being able to discriminate them from all possible counterfeits, or knowing all of the background conditions which make possible the thinking of the thoughts.

We have already touched upon our solution to the Problem of Opacity for epistemic externalism. The general worry here is that, in locating the source of epistemic warrant outside the thinking subject, externalism makes these epistemizing factors opaque to her. How can considerations which lie outside the ken of a believer epistemically warrant her beliefs? We take two interrelated lines of response here. First, the Problem of Opacity is at its most serious for pure reliabilism. But we have allowed that reliability as such is neither necessary nor sufficient for epistemic warrant. When the view is modified to overcome the most powerful objections in the literature it becomes much less clear whether there is a real worry along these lines. We require for perceptual warrant that a subject's perceptual states—and, by extension, her perceptual beliefs—be individuatively connected to the environment in which the states are formed; and that the subject be reliable in this sort of environment.[31] It is far from clear that the counterexamples to pure reliabilism touch this sort of view.

Second, we follow Burge in holding that every warrant has a psychological element over and above that trivially bestowed by the fact that warrant is a property of belief. All perceptual entitlements, for example, essentially involve perceptual states. Perceptual states are, in an intuitive sense, internal to a thinker; though it bears emphasizing that not all thinkers are conceptually sophisticated enough to think about their perceptual states—so it would be misleading to say that all warrants involve elements to which one has privileged or authoritative access.[32] Other entitlements rest partly upon understanding, which is, or crucially

[31] Again, see Burge 2003a: esp. 530–7; cf. also Majors and Sawyer 2005: section VI.

[32] For this reason we think it is misleading to view Burge's conception of epistemic warrant as a compromise between epistemic internalism and externalism. It is true that justification, on his

involves, an internal propositional state of the subject. Again, however, the crucial fact is that the view differs in significant ways from pure reliabilism. It is not subject to the counterexamples which plagued earlier versions of the doctrine.

5. OBJECTIVE NORMS

Our two central goals in this chapter are to articulate, and point toward solutions to, the Connection Problem, and the Problem of Opacity, on the one hand, and to set out a new and helpful way of viewing the two internalism/externalism disputes, on the other. In the previous section we argued that the forms of externalism do not in fact face the Connection Problem, and can overcome the Problem of Opacity. In the present section we try to set out an illuminating way to view the disputes in more general and abstract terms—an 'organizing principle'.

We think that the battles waged today between internalists and externalists, in both the theory of content and epistemology, are best viewed as the central contemporary manifestations of a perennial contest in philosophy, that between empiricism and rationalism. The key issue concerns *the source of the norms which govern thought and judgement*. Empiricists have traditionally been inclined to hold that such norms, to the extent that they are genuine, are grounded in features of the thinking subject herself, or the human species at most; whereas rationalists have been apt to insist that the relevant norms are grounded at a much deeper level, and that their normative force does not depend upon any sort of ratification by the individual thinking subject, or even a community of thinkers. And it does not depend upon specifics of our natures as animals. It is only a relatively slight exaggeration and oversimplification to put the point by saying that for the rationalist the norms are objective, and for the empiricist they are subjective.[33]

It is what Burge has called 'intellectual norms' which are at issue in the debate over the nature of representational content. Our ways of thinking and talking about objects and properties are answerable to the ways in which the most competent speakers of our language use the corresponding terms; and, more

view, is in certain respects closely associated with internalism (as entitlement is with externalism). But the similarity is rather superficial. For internalists have traditionally been prone to require that a warranted cognitive subject be able to justify her beliefs to the skeptic. But on the Burgean view we favor, being justified, or having reasons, does not guarantee even having the ability to think about justification, belief, or reasons. Higher nonhuman animals, and young human children, have justified beliefs, but lack the concepts necessary to think about their epistemic status, much less to defend it against the skeptic. So we think that the view offers little real solace to the epistemic internalist.

[33] Needless to say, we are not here concerned with defending historical claims concerning the views of empiricists and rationalists. But we do think that a historical case for our claims can be made.

fundamentally, to the natures of the entities themselves. This is the source of semantic externalism. My belief that I have arthritis in my thigh is mistaken, because the term 'arthritis' is not used in such a way, in the relevant linguistic community, that it is correctly applied to rheumatoid ailments outside the joints. This is a case of incomplete understanding. My belief that sofas are religious artefacts, rather than pieces of furniture meant for sitting, on the other hand, need not stem from incomplete understanding of the term 'sofa.' I may simply have a deviant theory about what sofas are. In this case the error is due to failure to grasp, not the way in which a term is used, but the actual nature of the entities referred to by the term. In both cases the norms are grounded in something more objective than the thinker's own considered practice. And in the second kind of case the norms are grounded even more deeply than the practice of the linguistic community as a whole. It is primarily in this sense that semantic externalism is a form of rationalism.

Semantic internalism, by contrast, disowns any commitment to normative authority, of the relevant sort, beyond that which is certified by the practices and commitments of the individual thinking subject. There is a sense in which an individual, on this view, cannot make a mistake about the contents of her thoughts and words. She is herself taken to be the source of such norms as govern her thought about (ostensibly) objective matters.

The dependence of thought upon the way the world is, and the possibility of incomplete understanding (both of meaning and of cognitive value), have arguably always been an important part of the rationalist picture. The Connection Problem, in its semantic guise, arises precisely because thought and world are not independent, and cannot be taken to be so if we are to account for the possibility of representation. And the key to the solution to the Problem of Opacity here is, in part, to recognize that there are different levels of understanding of thought content. The fact that one does not recognize one or more essential features of a concept one thinks with has no tendency to show that one does not know one's own thoughts, in the sense relevant to understanding first-person authority and privileged access. No more does the fact that one misunderstands the use experts make of the corresponding term threaten these features of self-knowledge. One can think with a concept, and know what one thinks when one employs it, without having attained mastery of the nature of the concept, or even of the associated linguistic meaning. This is because the intellectual norms which govern our thinking about objective matters are not of our own making. We are responsible to them. And we severally meet this responsibility in differing degrees.

The connections between externalism and rationalism—and internalism and empiricism—are if anything even more clear in the epistemic case. The classical epistemic internalist thinks that it is completely up to the individual whether she meets her epistemic obligations. The source of the norms is effectively thought to be located within her. If she does all that she can, or all that can reasonably be

expected of her, then she cannot fail to meet them. The externalist, by contrast, recognizes that belief and judgement are governed by norms ultimately grounded in the function of representation, which is to present veridicality or truth. Being epistemically responsible is neither necessary nor sufficient for being true, or even being well placed to achieve truth. It is not necessary, because not all believing creatures have the conceptual wherewithal to think about things like reasons, truth, and belief. And there can be no question of epistemic responsibility where these concepts are absent. It is not sufficient because doing the best that one can cognitively does not suffice to place one in the sort of relation to truth that meeting representational or epistemic norms consists in. If one is poorly adapted to one's normal environment, if one is malfunctioning perceptually or cognitively, then one will not meet the epistemic norms which necessarily govern one's beliefs and judgements.

The Connection Problem arises within the epistemic milieu because of the function of representation, and the nature of epistemic warrant. This function, and this nature, are such that only a view which can account for the constitutive connection between warrant and truth will be satisfactory. And the Problem of Opacity, here, is to show how cognitive elements without the ken of thinkers can nevertheless warrant their beliefs. Again the key, for the externalist, is to recognize that, because the norms which govern belief and judgement are not of our own making, satisfying the norms requires more than merely that everything be 'in order' on the inside of the thinker. Thus the objectivity of epistemic norms guarantees that not all epistemically significant elements will reside within the thinker herself. In this way the debate between epistemic internalists and externalists can be viewed as part of the Homeric struggle between *empiricists*—who, if they acknowledge epistemic norms in any substantive sense at all, locate them within the resources of the thinking subject—and *rationalists*—who view such norms as grounded in representational function, or a function of reason, which is not itself something over which the individual has authority or control.

6. MORAL PHILOSOPHY

Before closing we would like briefly to illustrate how the organizing principle we have suggested illuminates and unifies the internalism/externalism debates in semantics and epistemology can be applied as well to the corresponding disputes in moral philosophy. There are two internalism/externalism debates in moral philosophy. The first concerns the relationship between moral judgement and motivation. The second concerns the relationship between moral reasons and an agent's 'subjective motivational set.' In each case we think that the most profitable, general way to view the conflict is by seeing it as part of the larger battle between rationalism and empiricism. Furthermore, the Burgean conception of

norms as objectively governing our cognitive activity suggests particular positions with respect to each debate.[34]

It should be noted at the outset that the Connection Problem and the Problem of Opacity do not have clear application to the internalism/externalism debates within moral philosophy. Our focus here will be on the way in which the Burgean conception of norms illuminates the debates, and points toward solutions to the attendant problems.

Judgement internalists see a conceptual connection between moral judgement and motivation. On their view, no one who sincerely judges that she ought to ϕ can fail to be motivated, to some extent at least, to ϕ. Certainly the motivation may be outweighed by other considerations, or the agent may be weak of will. But there must be *some* motivation, if she genuinely makes the judgement. Proponents of this view often appeal to the phenomenology of moral judgement. When I recognize that I ought to ϕ, this is not like recognizing that the weather is fair. I *feel* that I ought to ϕ, and this just is to be motivated to do it. Moral obligations seem necessarily to force themselves upon us.[35]

Judgement externalists deny this. They agree that in most cases there is motivational commitment on the part of those making moral judgements. But they think that the connection is not a necessary one. Most people one knows are motivated to act in accordance with their moral judgements, perhaps. But this is because one is not acquainted with many moral freaks or monsters; not because there is any sort of conceptual connection between moral judgement and motivation.[36] Externalists often support their position by claiming the conceivability of the *amoralist*. This is a person who seems sincerely to judge that he ought to ϕ, yet nevertheless is not motivated in the slightest actually to ϕ. Both sides seem to agree that if the amoralist is possible, then judgement internalism is false. As we say, the externalist presses for such a possibility. But for the internalist the case is misdescribed. Certainly we can imagine a person uttering the word-forms 'I ought to ϕ', and yet not being motivated at all to ϕ. But to move from this to the claim that it is possible sincerely to judge that one ought to do such and such, without being motivated to do it, begs the question at issue. The issue is precisely whether, in uttering the relevant word-forms, such an agent would indeed be making a sincere and literal moral judgement.

[34] In this section we go beyond that to which Burge has officially committed himself, though we think that the views we outline are congenial to his general perspective. We do make clear use of his demonstration of the possibility of incomplete understanding of words and concepts; see the papers cited in n. 26. And it ought to be noted that there are some remarks in Burge 1998b: 250 ff., which suggest the sort of account we set out concerning moral judgement and motivation. Our remarks here also seem consistent, finally, with the brief discussion of practical norms in Burge 2003d: 327.

[35] Classic defenses of judgement internalism include McDowell 1979; Hare 1981; and Smith 1994

[36] Externalist accounts are advanced in Brink 1989; Svavarsdóttir 1999; and Shafer-Landau 2000.

This is the most difficult case for our suggested framework, in large part because there is no obvious way to characterize judgement internalism as a supervenience thesis. It is the only one of the four distinctions which resists such treatment. And to make matters worse, we think that both sides to the dispute are mistaken. In every other case, we believe, it is clear that externalism is the most plausible view. But the debate between judgement internalists and judgement externalists arguably rests upon a false presupposition.

As we have seen, internalists insist upon a conceptual connection between moral judgement and motivation. Externalists deny the existence of such a connection, because they think it is possible—as in the case of the amoralist—to judge that one ought morally to ϕ, and yet be completely unmotivated to ϕ. Yet the claims, first, that there is a conceptual connection between moral judgement and motivation, and, second, that the amoralist is conceivable, may not be genuinely incompatible. The tacit assumption shared by both parties to this dispute is that if there is a necessary connection between two concepts, then anyone who possesses the two concepts must (in some manner or other) recognize or manifest this fact. This appears to be false. Knowledge is factive. It is a necessary part of the concept *knowledge* that if one knows that p, then p. Yet one sometimes has to debate the point with students. These students are making a mistake about the concept, which is possible only if they possess the concept. That is to say, there is a conceptual connection between knowledge and factivity; yet it is perfectly possible for one to possess both of the relevant concepts and fail to recognize, in theory and in practice, the obtaining of the connection. Examples abound. Causation is either transitive or not transitive, and by conceptual necessity. Yet it is possible (indeed actual) to debate the matter. Therefore the obtaining of a conceptual connection does not necessitate its recognition by anyone who possesses the relevant concepts.

The upshot is that it seems possible that both sides are correct in their primary contentions. Indeed, we think that this may be the most plausible view to take of the matter. It is true that there is a conceptual connection between moral judgement and motivation. But it is also true that it is possible sincerely to make a moral judgement, and yet fail to be motivated, to even the slightest degree, to act in accordance with it. The reason for this is that it is possible, and indeed quite common, incompletely to grasp concepts. One possesses the concept *knowledge* just in case one thinks thoughts of which it is a constituent. Yet one can think such thoughts without recognizing, again in theory or practice, the factivity of knowledge. And this despite the fact that it is a conceptual truth that knowledge is factive. The relevance of this to the present dispute is clear. One can possess the concepts which make up the moral judgement that one ought to ϕ, without understanding them sufficiently that one is automatically compelled or motivated to act in accordance with it. Anyone with a complete grasp, or fully reflective understanding, of the concepts involved in the judgement would be motivated to ϕ. But complete grasp of the concepts involved is not a necessary

condition upon making the moral judgement. And that is why the amoralist is possible.

Despite the fact that this particular internalism/externalism dispute fits only awkwardly into our organizational schema, it is clear that our central idea has application to it. As is the case in semantics, here the central mistake of the internalist is to fail to see the possibility of an incomplete understanding of notions or concepts. It is just that here the externalist makes the same mistake. On the more promising view which we have adumbrated, there are degrees of understanding.[37] At the lowest level is the simple ability to think with the relevant notion. At an intermediate level, one can not only think with the notion or concept, but one can offer illuminating and correct explications of the concept. Finally, at the highest level of understanding, one gets it right. One sees, or otherwise respects, the constitutively necessary aspects of the notion involved. What is true about judgement externalism is that one may judge, at a lower level of understanding, that one ought to ϕ, without being in any way motivated to ϕ. Such a being is conceptually deficient, to be sure; but this does not prevent him from sincerely and literally making the relevant judgements. What is right about judgement internalism is that one cannot possess a full, reflective understanding of the notions involved in a moral judgement without being motivated to act in accordance with it. Needless to say, there is no space here adequately to set out, much less defend this view. For the present we note merely that it is fully in the spirit of classical moral rationalism—at least in certain cases, failure to (be motivated to) act in accordance with the dictates of morality is a failure of reason and understanding.

Reasons internalism is the view that one has moral reason to ϕ only if one's 'subjective motivational set,' or immediate extensions of it, provides one with a motive to ϕ.[38] Thus one has reason to save the drowning child immediately before one only if one has, broadly and loosely speaking, some (potential) conative element which favors performing the requisite actions. *Reasons externalism* is the denial of this claim. For the externalist, one might have a moral reason to ϕ despite the fact that one has no interest, no desire, no conative element whatsoever which favors ϕ-ing. Reasons stem directly from obligations, or from objective values, on this kind of view; and there is no need to take a detour through the vagaries of one's conative make-up.[39]

Reasons internalism can be formulated as a supervenience thesis: No two agents who do not differ with respect to their subjective motivational sets can differ with respect to their reasons for action. This view is the direct descendant of one long associated with Hume—reason has as such no power to motivate; therefore where there is practical reason, or reason to act, there must be desire.

[37] See Burge 1986: 713.

[38] See Williams 1981, both for the classic defensce of reasons internalism, and for the 'subjective motivational set' terminology.

[39] For defence of reasons externalism, see Parfit 1996; McDowell 1998; and Shafer-Landau 2003: chs. 7 and 8.

Contemporary proponents of internalism are willing to be more flexible than Hume was about the kind of conative element necessary for the presence of moral reasons. But the basic structure of the view is the same. Reasons externalism can be seen as the denial of the supervenience thesis.

There is a direct parallel between reasons internalism and epistemic internalism. Each view holds that what makes something—belief, judgement, action—rational, or provides a reason for it, must be internal to the thinking subject. To put the views crudely: Theoretical rationality is necessarily provided by reasons, and practical rationality is necessarily provided (at least in part) by desires, or other conative elements. But we saw above that this restrictive view of theoretical rationality is implausible. It does not do justice to the aim of belief and judgement, which is truth. Epistemic internalism would be appropriate were the aim of belief and judgement *perceived truth*. But the fact that epistemic responsibility is insufficient for justification shows that this is not in fact the relevant aim. Belief and judgement aim at truth in an objective sense—they are representationally or epistemically as they ought to be only if they relate in the right kind of way to actual truth.

Likewise we hold that reasons internalism fundamentally distorts the nature of practical rationality. It is no more plausible that only a desire, or other conative element, could make an action rational—that is, provide reason for an action—than that only other beliefs or reasons, on the part of the subject, could provide epistemic warrant for her beliefs or judgements. If a person's epistemic status is a measure of how well she is doing with respect to the function of representing veridically, then a person's moral status is arguably a measure of how well she is doing with respect to doing the morally right or good thing. In neither case are the agent's subjective resources, whether cognitive or conative, what is crucial.

What is crucial is how well the subject is doing with respect to the constitutive aim of the activity in question. We saw that the constitutive aim of belief is derivable from the function of representation, along with the fact that belief is a propositional or conceptual attitude. We will not try here to attempt a parallel derivation of the aim of action. What is crucial to note at this point is that ordinary moral practice does not in fact relativize either moral obligations or reasons for action to the subjective motivational sets of moral agents. It is a revisionary proposal that one has reason to help the drowning child out of the lake only if doing so furthers some end one has. Reasons internalists have, we think, felt forced into holding this view because they have failed to see how elements outside the moral agent's desiderative profile could bear positively on the agent's reasons for action. But the version of epistemic externalism we have defended here in effect shows how this is possible: External elements bear on rationality when they bear on the agent's ability to meet the categorical norms—or the associated constitutive aims—which necessarily govern her as a representational or rational agent.

Here again, and in large part because of the extended parallel with the epistemological debate, the way in which our central organizing principle bears on the debate is fairly clear and straightforward. Reasons internalists are the empiricists, defending the deflationary view of norms as indexed or relativized to the subjective (this time conative) resources of the agent. And reasons externalism, in the form at least in which we have suggested it be defended, is precisely the view supported by the rationalist's conception of norms as objectively governing our cognitive and practical activity.

We have looked at four internalism/externalism distinctions, and argued that each of them is most profitably viewed as part of a larger battle between empiricism and rationalism. Furthermore, we think that the Burgean conception of objective norms which we have defended, and tried to elucidate, provides the key to finding the most plausible position within each debate. With respect to the nature of representational content, and also the controversy over the connection between moral judgement and motivation, the important point is to recognize the possibility of incompletely understanding meanings, and imperfectly grasping concepts. This is made possible, in each case, by the fact that our understanding and explication is answerable to external and objective intellectual norms. The norms are not of our making. With respect to the nature of epistemic warrant, and the nature of reasons for action, the key is to see the link between having warrant, or having reasons, and fulfilling the norms set by the constitutive aim or function of the activity in question. Again it is the objectivity of the norms which underwrites the possibility of reason's outrunning our subjective cognitive and conative natures.

No doubt work remains to be done in understanding how the various internalism/externalism disputes relate to one another and are most profitably viewed in abstract terms. What we have tried to do here is to suggest an organizing principle which illuminates central features of each such dispute, as well as their interconnections. On the conception we have recommended each can be seen as a microcosmic manifestation of the great and ancient debate between empiricism and rationalism

REFERENCES

Boghossian, P. 2000: 'Knowledge of Logic.' In P. Boghossian and C. Peacocke (eds.), *New Essays on the A Priori* (Oxford: Oxford University Press).

____ 2002: 'How are Objective Reasons Possible?' In J. Bermudez and A. Millar (eds.), *Reason and Nature* (Oxford: Oxford University Press).

____ 2003: 'Blind Reasoning.' *Proceedings of the Aristotelian Society* suppl. vol. 77: 225–48.

Bonjour, L. 1985: *The Structure of Empirical Knowledge* (Cambridge, Mass.: Harvard University Press).

Brandom, R. 1994: *Making it Explicit* (Cambridge, Mass.: Harvard University Press).

Brink, D. 1989: *Moral Realism and the Foundations of Ethics* (Cambridge: Cambridge University Press).
Burge, T. 1979: 'Individualism and the Mental.' *Midwest Studies in Philosophy* 4: 73–121.
____ 1982: 'Other Bodies.' In A. Woodfield (ed.), *Thought and Object* (London: Oxford University Press).
____ 1986: 'Intellectual Norms and Foundations of Mind.' *Journal of Philosophy* 83: 697–720.
____ 1988: 'Individualism and Self-Knowledge.' *Journal of Philosophy* 85: 649–63.
____ 1993: 'Content Preservation.' *Philosophical Review* 102: 457–88.
____ 1996: 'Our Entitlement to Self-Knowledge.' *Proceedings of the Aristotelian Society* 96: 91–116.
____ 1998a: 'Computer Proof, Apriori Knowledge and Other Minds.' *Philosophical Perspectives* 12: 1–37.
____ 1998b: 'Reason and the First Person.' In B. Smith, C. Wright, and C. Macdonald (eds.), *Knowing our own Minds* (Oxford: Clarendon Press).
____ 2000: 'A Century of Deflation and a Moment about Self-Knowledge.' *Proceedings and Addresses of the American Philosophical Association* 73: 25–46.
____ 2003a: 'Perceptual Entitlement.' *Philosophy and Phenomenological Research* 67: 503–48.
____ 2003b: 'Perception.' *International Journal of Psychoanalysis* 84: 157–67.
____ 2003c: 'Descartes, Bare Concepts, and Anti-Individualism: Reply to Normore.' In M. Mahn and B. Ramberg (eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge* (Cambridge, Mass.: MIT Press).
____ 2003d: 'Memory and Persons.' *Philosophical Review* 112: 289–337.
____ 2005: *Truth, Thought, Reason* (Oxford: Oxford University Press).
Davidson, D. 1987: 'Knowing one's Own Mind.' *Proceedings and Addresses of the American Philosophical Association* 60: 441–58.
Davies, M. 2004: 'Epistemic Entitlement, Warrant Transmission and Easy Knowledge.' *Proceedings of the Aristotelian Society* suppl. vol. 78: 213–45.
Dennett, D. 1983: 'Beyond Belief.' In A. Woodfield (ed.), *Thought and Object* (Oxford: Oxford University Press).
Dretske, F. 2000: 'Entitlement: Epistemic Rights without Epistemic Duties?' *Philosophy and Phenomenological Research* 60: 591–606.
Feldman, R. and Conee, E. 1985: 'Evidentialism.' *Philosophical Studies* 48: 15–34.
Field, H. 2000: 'Apriority as an Evaluative Notion.' In P. Boghossian and C. Peacocke (eds.), *New Essays on the A Priori* (Oxford: Oxford University Press).
Fodor, J. 1987: *Psychosemantics* (Cambridge, Mass.: MIT Press).
Goldman, A. 1986: *Epistemology and Cognition* (Cambridge, Mass.: Harvard University Press).
Hare, R. M. 1981: *Moral Thinking* (Oxford: Oxford University Press).
Loar, B. 1988: 'Social Content and Psychological Content.' In R. Grimm and D. Merrill (eds.), *Contents of Thoughts* (Tucson: University of Arizona Press).
McDowell, J. 1979: 'Virtue and Reason.' *The Monist* 62: 331–50.
____ 1998: 'Might There be External Reasons?' In J. McDowell, *Mind, Value and Reality* (Cambridge, Mass.: Harvard University Press).
Majors, B. 2005: 'The New Rationalism.' *Philosophical Papers* 34: 289–305.

____ and Sawyer, S. 2005: 'The Epistemological Argument for Content Externalism.' *Philosophical Perspectives* 19: 257–80.
Moster, P. 1985: *Empirical Justification* (Dordrecht: Reidel).
Parfit, D. 1996: 'Reasons and Motivation.' *Proceedings of the Aristotelian Society* suppl. vol. 71: 99–130.
Peacocke, C. 1992: *A Study of Concepts* (Cambridge, Mass.: MIT Press).
____ 2004: *The Realm of Reason* (Oxford: Oxford University Press).
Plantinga, A. 1993: *Warrant: The Current Debate* (New York: Oxford University Press).
Pryor, J. 2000: 'The Skeptic and the Dogmatist.' *Noûs* 34: 517–49.
____ 2001: 'Highlights of Recent Epistemology.' *British Journal for the Philosophy of Science* 52: 95–124.
____ 2004: 'What is Wrong and Moore's Argument?' *Philosophical Issues* 14: 349–78.
Putnam, H. 1975: 'The Meaning of "Meaning".' In his *Mind, Language and Reality: Philosophical Papers*, ii (Cambridge: Cambridge University Press).
Sawyer, S. 2002: 'In Defense of Burge's Thesis.' *Philosophical Studies* 107: 109–28.
____ forthcoming: 'There is No Viable Notion of Narrow Content.' In *Contemporary Debates in Philosophy of Mind* (Oxford: Basil Blackwell).
Segal, G. 2000: *A Slim Book about Narrow Content* (Cambridge, Mass.: MIT Press).
Shafer-Landau, R. 2000: 'A Defence of Motivational Externalism.' *Philosophical Studies* 97: 267–91.
____ 2003: *Moral Realism: A Defence* (Oxford: Oxford University Press).
Smith, M. 1994: *The Moral Problem* (Oxford: Basil Blackwell).
Sosa, E. 1991: *Knowledge in Perspective* (Cambridge: Cambridge University Press).
Steup, M. 1988: 'The Deontic Conception of Epistemic Justification.' *Philosophical Studies* 53: 65–84.
Svavarsdóttir, S. 1999: 'Moral Cognitivisim and Motivation.' *Philosophical Review* 108: 161–219.
Williams, B. 1981: 'Internal and External Reasons.' In B. Williams, *Moral Luck* (Cambridge: Cambridge University Press).
Williams, M. 2000: 'Dretske on Epistemic Entitlement.' *Philosophy and Phenomenological Research* 60: 607–12.
Williamson, T. 2003: 'Understanding and Inference.' *Proceedings of the Aristotelian Society* suppl. vol. 77: 249–93.
Wright, C. 2002: 'On Basic Logical Knowledge.' In J. Bermudez and A. Millar (eds.), *Reason and Nature* (Oxford: Oxford University Press).
____ 2004a: 'Intuition, Entitlement and the Epistemology of Logical Laws.' *Dialectica* 58: 155–75.
____ 2004b: 'Warrant for Nothing (and Foundations for Free)?' *Proceedings of the Aristotelian Society* suppl. vol. 78: 167–212.

7

Content Externalism, Entitlement, and Reasons

Anthony Brueckner

1. INTRODUCTION

I want to explore the question of whether there is some connection between externalism about mental content and the concept of epistemic justification. One quick, naive, negative thought concerning the possibility of such a connection runs as follows. Externalism is, in part, a thesis about the *truth conditions* of a believed content, which are distinct from *justification conditions*, conditions under which a belief of the content is epistemically justified. So it is prima facie hard to see how externalism about content could tell us anything deep about justification. A quick, not-so-naive response to the prima facie worry proceeds from the idea that justification is *truth-conducive*. There is some sort of conceptual connection, it has been held, between justification and the likelihood of truth. So it seems quite possible that a thesis about truth conditions could have bearing on questions about justification. Further, if one is an externalist about justification who holds a pure reliabilist theory, there is all the more reason to expect a connection between justification and a view about how the truth conditions of a state's content are determined. If a justified belief is just one produced by a reliable belief-forming process, namely a process possessing a sufficiently high truth-ratio, then a thesis about truth conditions might well have a deep connection with the concept of justification.

These musings are way too abstract to be of any use in thinking about my guiding question about content externalism and justification. I would like to begin by looking at a theorist—Tyler Burge—who is not a pure reliabilist but who nevertheless accepts theses about epistemology that are in a certain way externalist (Burge 2003). He argues that externalism about *perceptual content* helps to explain why we are epistemically *entitled* to hold perceptual beliefs,

where this entitlement is not readily accessible to the non-philosophers among us. After examining Burge's view, I will look at the theory of Bill Brewer (1999). It appears that Brewer wishes to argue from a form of externalism about mental content to the conclusion that perceptual experiences provide reasons for beliefs. But the notion of *reason* figuring in Brewer's conclusion is *internalist* in character. A reason, in Brewer's conception, is always a reason *for* a person, in such a way that the reason is *accessible* to him.

2. BURGE

2.1. The Connection between Perceptual Content and Epistemic Entitlement

Burge distinguishes between two kinds of *epistemic warrant*: *entitlement* and *justification*. Entitlement is epistemically externalist, in that it is a form of warrant which need not be fully accessible, even on reflection, to the entitled subject S. S may even lack the concepts which are required to think the content that formulates the warrant. Justification, by contrast, is 'conceptually accessible', on reflection, to S. Burge's main goal in his (2003) is to elucidate the connection between anti-individualism about mental content, or content externalism, and perceptual entitlement.

For Burge, intentional, or representational, states have two different sorts of content: propositional and non-propositional. Beliefs have propositional content, while perceptual representations have non-propositional content. Both states can be *veridical*: beliefs are veridical when their propositional *truth conditions* are satisfied; perceptual experiences are veridical when their non-propositional *correctness conditions* are satisfied (as we find in the cases of good maps and accurate photographs).

The notion of perceptual entitlement which Burge is seeking to elucidate attaches to *perceptual beliefs*, beliefs which are formed on the basis of perceptual experiences. For Burge, the relation between a perceptual experience and an associated perceptual belief is not like the relation between a reason for belief and a belief. Reasons have propositional content, and perceptual experiences do not. When I believe that there is fire on the basis of the reason that, as I believe, there is smoke and where there is smoke, there is fire, my reason-based belief is *justified*. Things are different in the case of perceptual belief, according to Burge. When I believe that a black cat is approaching on the basis of a perceptual experience of a cat, I have an *epistemic entitlement* to rely upon the non-propositional perceptual representation in forming the belief. The relation between the experience and the belief is not like that between a justifying reason and a justified belief, according to Burge. To think otherwise, he says, is to think of the move from perception to belief as being too much like *inference*.

What is the source of this sort of perceptual entitlement? How do perceptual states contribute to perceptual entitlement? Burge's answer is that the thesis of anti-individualism about perceptual content entails that our experiences afford entitlement to our perceptual beliefs. According to the anti-individualist thesis, perceptual states have the representational content they do in virtue of causal interactions between the relevant perceptual system and the types of perceptual referents of the states. Burge says:

> A condition on particular perceptual representational states' having the content that they have is that there have been both causal-formative interactions (which are not in themselves representational) and representationally successful interactions between instances of types of relevant perceptual referents and aspects of the individual's perceptual system (in either the individual's history, or in evolution of the system in his evolutionary ancestors, or in some other way). (2003: 531)

One way of seeing the point is to consider the *shadow world* in which S's perceptual states of a given neural kind k are systematically caused by shadows (and there are few cracks). By contrast, in the *crack world*, S also has perceptual states of kind k, but they are systematically caused by cracks (and there are very few shadows). According to Burge, it would make no sense to attribute systematic error to S by attributing the perceptual content *crack* to S's perceptual states in the shadow world while attributing the content *shadow* to S's perceptual states in the crack world. Instead, the right way to ascribe content to S is to attribute mostly correct *shadow*-representations in the shadow world and mostly correct *crack*-representations in the crack world.[1]

Burge presents an argument linking anti-individualism about perceptual content with perceptual entitlement. I reconstruct Burge's argument as follows:

(I) Anti-individualism holds for perceptual content. (Assumption)

(II) An individual's perceptual state types are reliably veridical in his perceptual system's normal environment. (I)

(III) This reliable veridicality is explained by the nature of the perceptual states. (I)

(IV) If (i) an individual's perceptual state types are reliably veridical in his perceptual system's normal environment, and (ii) this reliable veridicality is explained by the nature of the perceptual states, then the individual is entitled to rely upon his perceptual states, and he is entitled to hold beliefs appropriately based upon the states. (Assumption)

(V) Perceptual entitlement exists. ((II), (III), (IV))

Before critically assessing this argument, let us note one key feature of Burge's overall position. Suppose that S is landed in an abnormal environment that

[1] See Burge 1986.

is quite different from the normal environment 'by reference to which the nature of [his] . . . perceptual states is explained' (1986: 33). Suppose that S's perceptual state types are *not* reliably veridical in the abnormal environment. Still, Burge wants to maintain that S has an entitlement to form perceptual beliefs on the basis of his experiences while in the abnormal, unfriendly environment.

2.2. Analysis of Burge's Argument

The last point helps to answer a prima facie worry about Burge's view. Given his emphasis on the claim that the truth of anti-individualism about perceptual content ensures that perceptual state types are reliably veridical in the normal, content-establishing environment, one might suppose that for Burge, the epistemic warrant flowing from perception is of the sort championed by the pure reliabilist. Burge is careful to dispel any such appearance. First, the perceptual system of a perceiver landed in a hostile, abnormal environment is highly *un*reliable in that environment, and yet on Burge's view, the perceiver's perceptual entitlement does not lapse. A recently envatted individual, for example, is still entitled to form ordinary beliefs upon the basis of his experiences. So reliability is seen not to be *necessary* for warranted perceptual beliefs.

Second, suppose that all we know about S is that his perceptual state types are reliably veridical in the environment in which he happens to find himself. According to Burge, we cannot say whether he is entitled to the perceptual beliefs he forms there. Why not? Why isn't this reliable veridicality *sufficient* for epistemic warrant? Burge's answer is the key to his argument. He says:

> The idea behind this view of entitlement is that reliance on perceptual states is warranted partly because the very identity of the states is constitutively and explanatorily associated with veridical representation. (2003: 532)

He continues:

> If the reliability of a perceptual representation is not grounded in the individuation and nature of the state, then the reliability cannot yield entitlement. (532)

Going back to my reconstruction of Burge's argument, the foregoing point is highlighted in premiss (IV): (i) and (ii) are *together* sufficient for perceptual entitlement. Reliability is not held to be by itself sufficient.

We now see that Burge is not a pure reliabilist about the warrant of perceptual beliefs. We also see that (IV) is the heart of his argument from anti-individualism to perceptual entitlement. Is (IV) plausible?

If one is a pure reliabilist, then the reliable veridicality of perceptual state types that is guaranteed by the anti-individualist account of content-establishment will suffice for the epistemic warrant of perceptual beliefs in normal friendly environments. End of story. At least, for the purposes of epistemology, we need

not enquire into the explanation of the existence of the reliable connections between beliefs and the world, on the pure reliabilist approach. Perhaps God has set the connections in place. It doesn't matter. But Burge, as we have just seen, is not at all a pure reliabilist about perceptual warrant. What exactly is the role of (III) in Burge's explication of perceptual entitlement? He says,

> Reliable connections to the world that are accidental relative to the conditions that individuate the individual's perceptual states . . . contribute nothing to empirical epistemic entitlement. (534)

So it is crucial to perceptual entitlement that the reliable connections flow from the nature of the perceptual states themselves, since this entails that the existence of the connections is not *accidental*. Thus it appears that it is the *necessity* attaching to the normal-environment-reliability of perceptual beliefs that, for Burge, sets up perceptual entitlement.

Suppose that a Cartesian God exists in every possible world and that in every possible world in which there are humans with perceptual systems, their perceptual states are reliably veridical (otherwise God would be a deceiver). Then it is *necessary* that perceptual states are reliably veridical. Would it then follow that we are entitled to our perceptual beliefs? Pure reliabilists would say 'Yes,' but if one is not a pure reliabilist, what should one say? It is not obvious. What does the element of *necessity* add to the epistemic status of the reliably produced perceptual beliefs?

According to Burge, it follows from the very concept of warranted belief that the formation of *true* belief is an *epistemic good*. Burge's argument shows that given anti-individualism about perceptual content, the very nature of perceptual states guarantees their connection with truth. So from anti-individualism we can deduce that perceptual beliefs satisfy the truth-conduciveness condition that is built into the concept of warranted belief. So there is a contrast between the necessary reliability flowing from anti-individualist perceptual content-establishment and the imagined necessary reliability in the case of the non-deceiving Cartesian God. In the former circumstance, the truth-conduciveness attaching to warranted belief is shown in an a priori manner to follow from the nature of the perceptual states that are related to the pertinent beliefs. This is not so in the imagined Cartesian circumstance. It is *God's* non-deceiving nature, and not that of the perceptual states themselves, which allows us to know a priori (under the assumption of a non-deceiving God) that the states are connected to beliefs that satisfy the truth-conduciveness required for warranted belief.

There are two problems with this line of thought. First, even if the contrast between the scenarios *is* relevant to the question of epistemic warrant, there still remains the question of why the *necessity* of the reliability of perceptual belief (whatever its source) is itself relevant to our entitlement to hold them (given that reliability *tout court* is held to be insufficient). Second, it is not obvious why

the *source* of the reliability in question *is* relevant to warrant, if there is necessity attaching to the reliability in both cases.[2]

So Burge's argument connecting his anti-individualism about perceptual content and perceptual entitlement is problematic, owing to the problematic status of the argument's key premiss, (IV).

3. BREWER

3.1. The Connection between Perceptual Demonstrative Content and Reasons for Belief

Bill Brewer presents an argument for the following thesis:

(R) Perceptual experiences provide reasons for empirical beliefs.

He wants to argue that the content-determining role of perceptual experiences vis-à-vis associated empirical beliefs requires that experiences provide reasons for the beliefs. Before considering Brewer's master argument for this thesis and his defense of it, I will first discuss his conception of perceptual experience.

In contrast to Burge, Brewer holds that perceptual experiences have *representational content* and that the content is *conceptual*. For Brewer, conceptual content is *propositional content*: it is the sort of content possessed by beliefs, which can serve as premisses and conclusions in inferences. On this view, perceptual experiences can function as *reasons* in the same way in which beliefs function as reasons. Rather than saying, as Brewer does, that experiences *provide* reasons for belief, we could say more precisely that experiences *are* reasons for belief. But I will follow Brewer's usage.[3]

Earlier in the book, before trying to establish (R), Brewer argues that purely descriptive reference to mind-independent spatial particulars is impossible. But we *do* refer to such particulars. So, Brewer argues, we must have experiences with *perceptual demonstrative content*. Brewer's argument for this claim derives from Strawson's famous discussion in *Individuals* (involving the possibility of massive reduplication of the scene to which purely descriptive reference is attempted).[4] For my purposes here, it is not necessary to discuss the details of Brewer's argument. So I will sketch what I take to be his picture of perceptual demonstrative content.

I look at my cat, Marco, and have a visual perceptual experience at t whose content we could express in language by uttering the sentence 'That is thus'

[2] Thanks to Aaron Zimmerman for helpful discussion of the role of premiss (III) in Burge's argument.

[3] One might hold that *believed propositional contents* and *the propositional contents of experiences* are reasons for beliefs. Then Brewer's usage would be appropriate. Thanks here to Aaron Zimmerman.

[4] See Strawson 1959.

in an appropriate way at t. The experience, it seems, has a content with two demonstrative elements corresponding to the *linguistic* demonstratives in the foregoing sentence, according to what seems to be Brewer's picture.[5] One perceptual demonstrative element (the analog of 'that' in the foregoing sentence) refers to Marco, and the other (corresponding to 'thus' in the foregoing sentence) refers to a way of being black, on the understanding of Brewer that one first achieves in reading the earlier parts of the book. When I have an experience with such demonstrative content, I typically will have a related belief whose content is also expressible by 'That is thus.' Call this a *demonstrative belief about a spatial particular*. Surely not all of my empirical beliefs are such *DBSP*s. For example, my belief that Marco is black is not a DBSP.[6] Though he is not altogether clear on this point, Brewer's view seems to be that in order to have any beliefs about spatial particulars (*BSP*s), I must have some DBSPs. Brewer presumably does not hold the stronger thesis that for every BSP that I have, I must also have some *suitably related* DBSP. I believe that Marco is now eating grass some miles away from where I sit. According to the stronger thesis, what would be the suitably related DBSP?

For our purposes, we need not settle this question, and we can simply grant for the sake of argument that if I have BSPs, then some of them must be DBSPs. Two points are crucial for our purposes, though. First, Brewer's argument for (R) seems to be restricted to DBSPs. I will return to this point later. Second, the argument hinges upon the claim that the content of a given DBSP is *determined by* the content of an associated experience having perceptual demonstrative content. Brewer, surprisingly, does not explain how this determination is supposed to work. One might suppose that Brewer is thinking of such content-determination along the following lines. Supposes that S has a perceptual experience at t whose content is expressible by an appropriate utterance at t (or just later) of 'That is thus'. The demonstrative elements involved in the experience's content refer to an object *o* and a property *F*, we will suppose. Given that the DBSP is formed on the basis of having the experience, the demonstrative elements in the belief's content will also refer to *o* and *F*.

How is perceptual content *itself* determined, on Brewer's view? Part of Brewer's story is that perceptual *demonstrative* content is determined by the objects of perception in virtue of such contents' *object-involving* character. For example, my cat Marco is a constitutent of the perceptual demonstrative contents of experiences caused by the cat. If I have a hallucinatory experience which I take to have the same content as a veridical experience of Marco, then I am simply mistaken about the comparative nature of the content. It is not a content of the same type as that had by my seeings of Marco, because it does not involve

[5] This is my reconstruction of Brewer's picture; he is surprisingly unforthcoming on the crucial details.

[6] We will discuss Brewer's views about such empirical beliefs below.

the cat. Given that the content of a veridical experience is, on Brewer's view, propositional, he seems to be committed to the existence of singular propositions having cats as constituents, which serve as contents of veridical perceptions.

But now we can see that the earlier account of how the contents of DBSPs are determined by the contents of associated experiences is not Brewer's. This is because on that earlier account, perceptual demonstrative contents were thought to involve, e.g. a demonstrative element that refers to an object. But as we just saw, Brewer holds that perceptual demonstrative contents contain objects as constituents. It would be odd to hold that a perceptual demonstrative content contains *both* a cat *and* a demonstrative element that refers to the cat. Maybe this *is* Brewer's view; unfortunately, he does not tell us enough to sort this out. Another question now arises: how does the singular, non-demonstrative content of 'Marco is black' differ from that of 'That is thus'?[7]

Let us now turn to Brewer's argument for his central thesis (R). He calls this the *switching argument*. Suppose that S believes that p on the basis of perceptual experience e. Suppose further that we have a case in which the content of S's belief is determined by e. So, given the sparse discussion of how this is supposed to work, and given the starring role of perceptual demonstrative content up until the presentation of the switching argument, we can fairly assume that we are dealing with a DBSP whose content is fixed by that of an associated experience also having demonstrative content. Here is the switching argument, as I reconstruct it.

1. The content of S's belief that p is determined by e. (Assumption)
2. S does not believe that q, which differs in content from p and whose content is such that it would have been determined by experience e′ whose content differs from that of e. (Assumption)
3. Suppose that e is *not* reason-giving, so that e does not favor p over q (i.e. e gives S no more reason to believe that p than to believe that q).
4. S has no more reason to believe that p than to believe that q. (3)
5. The alleged difference between believing that p and believing that q is 'nothing to S'. (4)
6. S does not understand p and q as alternatives; for S, *believing that p = believing that q*. (5)
7. It is not the case that e determines the content of S's belief.

I do not understand the logic of Brewer's argument. Rather than inferring (7), it would seem to make more sense to infer not-(3), regarding (3) as an assumption for reductio. (6) allegedly follows from (3) and the stipulated assumptions (1) and (2), and (6) contradicts our stipulated assumption (2).

[7] Thanks here to Aaron Zimmerman.

3.2. Analysis of Brewer's Argument

(4) is just an unfolding of what I am taking to be the assumption for reductio, (3). The move from (4) to (5) is clearly the heart of the argument. But this move seems at first glance to be a non sequitur. Why would the fact that S has no more reason to believe one perceptually based content than to believe an apparently different one entail that the contents *do not differ for S*? Suppose that I believe that *that is a goldfinch* on the basis of a bird-sighting. It might well be that given my poor discriminatory powers, I have no more reason to believe that that is a goldfinch than to believe that *that is a goldcrest*. It obviously does not follow that for me, there is no difference between believing the one content and believing the other. Another example: a skeptic, waiting to be impressed by an argument for (R), would protest that my present visual experience of a hand gives me no more reason to believe that *that is a hand* than to believe that *that is a pseudo-hand*, part of what Putnam calls 'the image' generated by a computer for the consideration of brains in vats. Those two contents, though, are clearly understood by me as being distinct.

Brewer might reply by holding that his switching argument is meant to apply only to what I will call *pure* DBSPs, whose content is expressible by an appropriate utterance of the ultra-minimal 'That is thus.' Our assumption for reductio, (3), says that e is not reason-giving. We are to understand this assumption as entailing that e gives S no more reason to believe the pure demonstrative content in question—p—than to believe some other pure demonstrative content q, determinable by experience e′, presumably in the same way that p is determined by e. However, since the belief that q would be a pure DBSP, it is hard to imagine the scenario we are being asked to evaluate. That is, while having e at t (or shortly thereafter, as S continues to entertain e's purely demonstrative content in memory), S forms the pure DBSP that p. Is it true that believing that p at t (or shortly thereafter) is no more reasonable than believing a *different* perceptual demonstrative content—q—which S is *incapable* of entertaining at t in virtue of lacking the appropriate experience e′ at t? Since it appears that S is incapable of believing that q at t, due to q's demonstrative character, it clearly cannot be just as reasonable for S to believe that q at t as to believe that p. But it seems that the denier of the reason-giving status of e is being set up as a straw man: he is being represented by Brewer as holding that S has no more reason to believe p than to believe some different content—q—that he is deeply incapable of believing at the pertinent time, in virtue of lacking at the time the pertinent content-determining experience e′.

Another worry about the switching argument is engendered by its apparent restriction to pure DBSPs. If the argument is indeed so restricted, then it does not show that my current perceptual experience provides (or is) a reason for me to believe that Marco is black, or that that is a cat, or even that that is

black. Surely these are the sorts of perceptual beliefs whose justification is of primary concern.[8] So if perceptual experiences do indeed provide reasons for such ordinary perceptual beliefs, then the switching argument does not establish this result, given the restricted scope we are presently assuming, i.e. a restriction to pure DBSPs whose content is expressible by 'That is thus.' I will return to this point later.

In order to better understand Brewer's overall case for (R), we need to look at two things. First, we need to look at Brewer's critique of what he takes to be some views according to which perceptual experiences are *non-reason-giving*. Second, we need to look at Brewer's account of *how* experiences provide reasons for belief. He thinks that there is a sense in which the switching argument is *non-constructive* (though he does not put it this way): it is a highly abstract argument to the effect that experiences are reason-giving, but it fails to elucidate the details of how this works.

Let us turn to what Brewer deems non-reason-giving views. His first example is what he calls the 'familiar view' of color. On this view, the redness of an object consists in some microphysical property of objects which causes *red′ experiences*, i.e. those experiences which are phenomenally like the experience I have when seeing a lobster under normal perceptual conditions. Beliefs that objects are red acquire their content in virtue of their relation to red′ experience. The familiar view is said to be a non-reason-giving view because having a red′ experience, according to Brewer, gives S no more reason to form beliefs about the microstructure X that actually constitutes redness than to form beliefs about some *other* microstructure. This is because if some other microstructure Y had been present in the world and had systematically caused red′ experiences, then S's beliefs 'about redness' would have been 'about Y,' and S would have been 'none the wiser'. Given S's ignorance about his supposed color-beliefs' truth conditions—about their 'semantic values,' as Brewer puts it—S cannot be said to believe that the lobster is *red*, as opposed to some other color. So we reduce the non-reason-giving view to absurdity: S clearly *does* believe that the lobster is red.

This is a very peculiar objection to the 'familiar view' of color. First off, in the counterfactual situation in which objects having microstructure Y cause red′ experiences, S does not, contra Brewer, have any beliefs about redness. The correct thing to say is that there is nothing red in the counterfactual situation; when S says, in that situation, 'The lobster is red,' he is expressing a different belief from that which he actually expresses by uttering the sentence. Depending upon what Y is, he may well be expressing the correct belief that we would express by saying 'The lobster is blue.' (Suppose that Y is the microstructure that typically causes blue′ experiences in the actual situation.) Does contemplation of the counterfactual situation show that S is ignorant of the actual truth conditions

[8] It is not at all clear to me that I ever form perceptual beliefs with contents expressible by 'That is thus.'

of his sentence 'Lobsters are red'? Well, he knows that 'Lobsters are red' is true if and only if lobsters are red. Just because he does not know the microphysical nature of redness, it hardly seems to follow that he is ignorant of his sentence's truth conditions in such a way as to preclude him from believing that lobsters are red.

Further, it is not at all obvious that a proponent of the 'familiar view' of color is committed to holding that it is a non-reason-giving view in Brewer's sense. If S correctly believes that being red is having whatever microstructure in fact typically produces red′ experiences, and S believes that he is having a red′ experience, then it would seem that S has a (defeasible) reason for believing that he is seeing something red!

Brewer turns to Burge-style content externalists and attributes to them the following view: non-reason-giving content-determining relations between perceptual experiences and beliefs give rise to pervasive, perfectly acceptable *ignorance* about semantic value, i.e. ignorance about truth conditions which is compatible with *possessing* the naturally attributed beliefs. So, on this interpretation of content externalism, in the case of the 'familiar view' of color, Brewer was wrong to suggest that ignorance of semantic value entails that one does not genuinely possess the beliefs at which the non-reason-giving view is targeted. I believe that there is water nearby, and if I were unwittingly switched to Twin Earth, I would unwittingly come to believe on occasion that there is *twater* nearby after inhabiting my new environment (while in ignorance of my new beliefs' semantic values).

Brewer ultimately rejects Burgean content externalism about natural kinds as a failed non-reason-giving view. Since I would be 'none the wiser' in the switching case, Brewer finds it implausible for the Burgean to claim that I genuinely possess the pertinent natural kind beliefs. Without going into the details of Brewer's analysis of the situation, we can see that the Burgean content externalist is a straw man in the debate over (R). Burge-style content externalism is just not obviously or explicitly a non-reason-giving view at all. Indeed, as we have seen, Burge argues that content externalism *entails* that experiences provide a form of epistemic warrant for perceptual beliefs. Even if Burge's argument for this fails, it seems pointless to try to expose the Burge-style view as itself *entailing* the *denial* of (R).

3.3. The Explication of (R)

The switching argument is a sort of non-constructive existence proof of *reason-giving* experiences. Let us now turn to Brewer's explication of (R): his discussion of *how* exactly experiences *do* provide reasons for belief. This explication consists mainly of Brewer's unfolding of the *objective content* of perceptual experience. He begins by maintaining that experiences represent spatial particulars by displaying their spatial location. He maintains further that experiences present objects as

located *relative to the perceiver*. So perceptual content is *egocentric*. Further, particular locations are represented in experience as being in a world of places and things that are independent of the perceiver's actual experiences of them. So there is an element of *objectivity*, or mind-independence, built into the content of the perceptual representation. Putting this all together, we can unpack the demonstrative content of a given perceptual experience as being expressible by an appropriate utterance of

(O) That thing there is thus, in virtue of features of the thing which might have appeared differently from different spatial points of view.

The element of content corresponding to 'there' provides both the egocentricity of the representation and its spatial-locating character; the element corresponding to the 'might have . . .' provides the objectivity of the representation. At first pass, we tried to capture the demonstrative content of perceptual experience as being expressible by an appropriate utterance of 'That is thus.' But, we now see, things are much more complicated: when I see my black cat Marco, my experiences represent Marco as a demonstrated thing located in space in a certain way with respect to me, as appearing in a certain demonstrated way, which way is but one of a number of different possible ways for the thing to appear, depending upon the spatial point of view from which the appearance is obtained.

This is Strawson gone wild: this is quite a complex content, which unfortunately cannot be entertained in perception by small children and apes, not to mention cats. Thus, on Brewer's view, adult humans have perceptual experiences whose contents are radically different from those of all other animals' experiences. Small children cannot entertain the modally flavored objective component of the complex content, and neither can non-human animals. The egocentric component is presumably available to small children, but not to non-human animals (though perhaps some attenuated analog of 'to the left of me' is available to non-human animals).

My aim here is not to present a detailed challenge to Brewer's highly intellectualized account of perceptual content as being objective and egocentric. Rather, my aim is to see how this account figures in Brewer's explication of (R). Before doing that, I will simply state in passing a prima facie worry about Brewer's account. It is very difficult to see how a *perceptual* representation could have the sort of *modal* content that Brewer claims. Granted, when one grasps the full concept of a physical object, one comes to understand the objective component of the concept, which seems to involve the modal conceptualization of various possible points of view upon a single physical thing, famously highlighted by Strawson in *The Bounds of Sense*.[9] But it is hard to get a grip on the idea that the content of a perceptual representation has this complex modal conceptualization somehow built into it. Further, it is very hard to square this view about the

[9] See Strawson 1966.

complex content of perceptual representation with Brewer's idea that such contents are singular in nature.

Brewer suggests several different ways of linking (R) with the foregoing unpacking of perceptual demonstrative content. Here is the first.

> [The perceiver] . . . understands that his current apprehension that things are thus and so is in part due to the very fact that they are. His grasping the content that that is thus is in part due to the fact that that *is* thus. He therefore recognizes the relevant content *as* his apprehension of the facts, his *epistemic openness* to the way things mind-independently are out there. (1999: 204)

Brewer also says that the subject's grasping of perceptual demonstrative contents 'as the objective contents they are involves his recognition of them as his apprehension of the relevant empirical facts' (204). Again, the subject 'necessarily recognizes that his entertaining . . . [of a perceptual demonstrative content] is a response to that thing's being thus, given his location and present circumstances' (205). Therefore, the subject's perceptual demonstrative contents 'give him a reason to endorse those very contents in belief' (204). This is (R).

The following picture emerges from these remarks. S entertains and grasps an objective perceptual demonstrative content C. It follows that S recognizes C as his apprehension of, or response to, a way W that mind-independent things are. Thus C provides S with a reason to believe that W obtains.

In order to understand this picture, we need to know what 'recognize' is supposed to mean. What is it for S to recognize x as being F? Here are two possibilities: S correctly believes that x is F; S knows that x is F. On the first interpretation, (R) is explicated in the following way. In virtue of entertaining C, S *correctly believes* that C is his apprehension of W, and *therefore* S has reason to believe that W obtains. This seems implausible. It is tantamount to the false claim that if S has a correct belief that entails that P, then S thereby has *reason to believe* that P. I can correctly but unjustifiably believe that Seabiscuit is the winner without having *reason to believe* that my favorite horse has won. Let us now consider the second interpretation of 'recognize.' According to this interpretation, in virtue of entertaining C, S *knows that* C is his apprehension of W, and thus, a fortiori, S *has reason to believe* that W obtains. This seems question-begging. We were seeking understanding of how experiences provide reasons for belief. We are now told that in having an experience, we entertain its content and *thereby* know that this content is our apprehension of mind-independent reality. Thus, experiences provide reasons for belief regarding external reality.

A related though more promising line of thought is suggested by some other remarks of Brewer's. He compares perceptual demonstrative contents to contents which are knowable a priori in the sense that 'understanding them is sufficient for knowledge of their truth' (206). Perceptual demonstrative contents are knowable a priori in this sense, according to Brewer, 'or at least they are a priori reasonable'

(206). In 'Epistemic Openness and Perceptual Defeasibility,' M. F. Martin characterizes Brewer's view as follows:

> Brewer's claim is that in entertaining the content of an experience (i.e., a veridical perceptual experience) one thereby necessarily recognizes its truth and hence has a reason to endorse the content. In a way, his account echoes what some have wanted to say about examples of self-evident propositions, such as may be expressed by sentences like 'I am here now', or, 'Whoever is tall, is tall', where one's mere entertaining of what the sentence expresses reveals to one that it must be true, given that one entertains it, and hence gives one reason to believe it. (2001: 442)

Martin considers a case in which I entertain the demonstrative content that *that is a cube* as a hypothesis, after catching a glimpse of an object through a haze. Here I entertain a perceptual demonstrative content without recognizing its truth. Martin's point is that merely *entertaining* a perceptual demonstrative content is not sufficient for recognizing its truth; Brewer must add that the content is in fact *experienced*. However, even if we take Martin's point into account and suppose that the pertinent content *is* experienced, it could well be that it is a *false* content (the demonstrated thing is non-cubical) and so not guaranteed to be recognizably true. It seems that, as in the case of the switching argument, Brewer must retreat to *pure* demonstrative contents: we are back to 'That is thus.' Such contents are so minimal that they *are* candidates for a sort of a priori knowability: when one recognizes that the content is merely to the effect that a certain demonstrated thing has a certain demonstrated property, one recognizes that there is no room for error.

There are three problems for this explication of (R). First, Brewer's claims about the egocentric and objective features of perceptual demonstrative content play no role whatsoever in the explication. Second, there can be cases in which one *mistakenly* thinks that one's experiential content involves a demonstration of an object and a property it possesses. When I have a hallucination of a black cat, I may well mistakenly believe that my experience has a perceptual demonstrative content expressible by 'That is thus'. In such a case, I may well mistakenly believe that I can recognize that there is no room for error regarding the truth of my experience's content.

Brewer holds that in what Timothy Williamson calls 'the bad case,' in which I have a hallucinatory experience, my experience does *not* provide me with a reason to endorse it in belief (and obviously does not yield perceptual knowledge). This is compatible with the view that my veridical experience in 'the good case' *does* provide me with a reason to endorse it in belief. In the good case, Brewer can say, I recognize my perceptual content for what it is: a pure demonstrative content whose minimal character ensures its truth. This rejoinder raises a host of difficult issues surrounding the question of what sort of discriminative ability is required of a perceiver whose experiences give him genuine reasons for belief. If a perceiver in the bad case believes that he is in the good case, then there is a sense in which

he cannot discriminate the good case from the bad case. How can it then be said that when he is in the good case, he *recognizes* that he is in the good case and *recognizes* that he is entertaining a perceptual demonstrative content whose truth is ensured by its minimal character?

The third problem for the current understanding of why (R) is true concerns that minimal character. At the least, we want an account of why an experience whose content is expressible as 'That is cubical' provides one with a reason to endorse it in belief. Brewer considers non-inferential knowledge that *a is F*, where neither 'a' nor 'F' functions as a perceptual demonstrative. So we might, for example, be dealing with a case of non-inferential knowledge that *Marco is black*. Brewer says,

> . . . a person's non-inferential knowledge that a is F . . . is not just a matter of his being drilled by training and practice to say something when and only when it is true, or to form a belief that a is F just when a is indeed F in his vicinity. For he has learnt that that is a and that being thus is (a way of) being F, by learning that he can just recognize such things. Thus it is not that there are simply the theorist's reasons for his (the subject's) believing that a is F, namely that a is indeed F, right there in front of him, but the subject himself has reasons for this belief: that is thus, that is a, and being thus is (a way of) being F. (1999: 247)

Brewer's picture is that once one has learnt that one can recognize that *that is a* and that *thus is a way of being F*, then one can non-inferentially recognize that a is F.

I will ignore the puzzling idea that one's non-inferential knowledge that Marco is black depends upon the *higher-order* knowledge that one can recognize that that is Marco and that thus is a way of being black. I will proceed directly to a related worry of Martin's. Martin objects that the account embodied in the foregoing remarks is incompatible with Brewer's claim that one's knowledge that a is F is *non-inferential*. It would seem that on Brewer's account, one is justified in believing that a is F only if one is justified in believing that that is a and that thus is a way of being F. Brewer replies as follows:

> [The subject's knowledge that a is F] . . . is not the product of the subject's perceptual knowledge of some more basic 'evidential' fact (or facts), such as that the thing in question is white with black dots, whose obtaining she has independent knowledge is a *reliable indicator* that the thing in question is a cube. The dependence which I insist upon here is constitutive rather than evidential in this way.[10]

The first problem for this response to Martin is that it is unclear what Brewer means by distinguishing between a constitutive and an evidential dependence. In the previously quoted passage, Brewer says that the perceiver has three reasons for believing that a is F: that is thus, that is a, and thus is a way of being F. If these reasons are the contents of the perceiver's beliefs, then it would seem that there is an evidential dependence between the perceiver's belief that a is F and his other beliefs. Another possibility (not put forward by Brewer) is that the reasons

[10] See Brewer 2001: 461.

are all perceptual, in that they are the contents of the perceiver's experiences. This possibility, though, leads into a second problem: in order to explain why one has a reason to believe a content expressible by 'Marco is black', Brewer cannot appeal to the absolutely minimal character of the content expressible by 'That is thus' (which character, according to Brewer, provides the possibility of a priori knowledge of such contents). His account must somehow allow one to go beyond that minimal content. Now as noted, it is not true that entertaining an experienced content expressible by 'Marco is black' is sufficient for one to recognize its truth, for the content may be false. The same remarks apply to the contents expressible by 'That is Marco' and 'Thus is a way of being black'. It is just not clear in the end how Brewer's account is going to vindicate the non-inferential knowledge that Marco is black.

Finally, Brewer suggests yet another explication of (R). He says, in summarizing his notions of perceptual openness and objectivity, 'The epistemically crucial role for perceptual experiences [is] . . . their ineliminable role in making reference to the mind-independent world possible in thought at all' (1999: 207). We might reconstruct this remark as a compressed argument:

(A) There is a mind-independent world to which we make reference.
(B) Such reference is possible only if we have experiences with demonstrative content.

So:

(C) Such experiences provide reasons for perceptual beliefs.

There are several things to note about this argument, First, no connection is made between externalist content-determination of perceptual belief and the epistemic status of such belief. So the argument has little to do with the thrust of Brewer's reflections concerning (R). Second, there is a gaping hole in the argument as it stands. All that follows from (A) and (B) is

(C′) We have experiences with demonstrative content.

A general point about transcendental arguments: if P and true, and the truth of Q is a condition for the possibility of the truth of P, nothing follows regarding the epistemic status of Q (or of the phenomena described by Q).

4. CONCLUSION

In the end, both Burge's and Brewer's attempts to connect externalist content-determination with a claim about the epistemic status of perceptual belief are not convincingly successful. It remains to be seen whether such a connection exists.

REFERENCES

Brewer, B. 1999: *Perception and Reason* (Oxford: Clarendon Press).

——2001: 'Replies'. *Philosophy and Phenomenological Research* 63: 449–64.

Burge, T. 1986: 'Cartesian Error and the Objectivity of Perception.' In P. Pettit and J. McDowell (eds.), *Subject, Thought and Content* (Oxford: Oxford University Press).

——2003: 'Perceptual Entitlement.' *Philosophy and Phenomenological Research* 67: 503–48.

Martin, M. F. 2001: 'Epistemic Openness and Perceptual Defeasibility.' *Philosophy and Phenomenological Research* 63: 441–8.

Strawson, P. F. 1959: *Individuals: An Essay in Descriptive Metaphysics* (London: Methuen).

——1966: *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason* (London: Methuen).

8

What's Wrong with McKinsey-Style Reasoning?

James Pryor

I

It's widely accepted nowadays that some of our thoughts are externalist. What does that mean? I understand a **thought's content** to be those of its representational aspects that one *experiences as* representational and are essential to its being *that thought*. And we can understand a thought to be **externalist** when subjects who are internally the same can differ with respect to whether they're thinking the thought's content.[1]

Versions of this chapter have been circulating since 2000. I've benefited from useful comments at Rochester, Penn, Stanford, Brown, Barcelona, St Andrews, UNAM, CUNY, Indiana, and conferences in Tlaxcala, Utah, and Kentucky. Special thanks to Miguel Fernández, John Collins, Henry, Utah, and Kentucky. Jackson, and Skip Larkin for their detailed responses. I got further useful advice from the URG and Eminees reading groups in Cambridge; from my grad seminars at Princeton and NYU; and from Richard Heck, Alison Simmons, Mark Johnston, Scott Soames, Gideon Rosen, and Stephen Schiffer. I appreciate all their good advice, despite not being able to implement all of it.

[1] I leave it open whether a thought's content suffices to determine the thought's truth-conditions. On a view like Lewis 1979's, when I think to myself 'My pants are on fire,' and you think to yourself 'My pants are on fire,' we're thinking *the same* content though our thoughts have different truth-conditions. Segal 1989 proposes a similar view about demonstrative thoughts. I don't count extra-contentual semantic differences of these sorts as externalist. Neither do I count as externalist views that say that Earthling thoughts are about H_2O, but only contingently so: those very same thoughts could have been about XYZ instead. An externalist has to say there are *essential* differences in the contents of some duplicates' thoughts.

My gloss on 'externalist' employs the notion of subjects being 'internally the same.' It isn't straightforward what that amounts to. It'd be nice if our definition of externalism didn't presuppose any commitments about materialism. But it's not obvious that we can characterize 'how you are internally' in terms of phenomenology either. Some philosophers (Dretske 1995: ch. 5; Tye 1995: s. 5.4, and 2000: s. 3.5) say that sensory phenomenology can itself be externalist. Our definitions ought to permit such a view. Additionally, I think it can also be part of your phenomenology that you are now occurrently thinking such-and-such—where that content may intuitively be an externalist one. For these reasons, I find it difficult to specify what it is for subjects to be 'internally

Our thoughts about water are widely thought to be externalist. So too are our demonstrative thoughts.

If we accept any kind of externalism, then we confront a puzzle that arises about our ability to tell what we're thinking. This comes from what's commonly known as **McKinsey's Argument**.[2] It rests on two ideas.

First, we think you can tell the contents of your thoughts just by introspection. For example, you can tell just by introspection that you're thinking that water puts out fires:

McK-1 You're thinking a thought with the content *`Water puts out fires`*.

As we proceed, it will be handy to have a way of identifying thoughts, without prejudging anything about how they're individuated, or indeed, whether they're even externalist. One way to identify a melody is to *produce* it: you can say 'Hey, you know that jingle [here you hum: *la-la-lalalaa* . . .]?' Analogously, I think you can identify a thought content by thinking or entertaining it. You can say: 'the thought content [here you *think* the relevant content, perhaps by rehearsing to yourself the sentence "`Water puts out fires`"].' I'm using the notation:

the content *`Water puts out fires`*

to express this way of demonstrating a thought content, by thinking it.[3]

The second idea driving McKinsey's Puzzle is that you can *also* tell that some of your thoughts are externalist, purely by armchair philosophical reflection. And it would seem that if a given thought is externalist, then it's only available to be had by subjects in certain sorts of environments (more on this in a moment). Hence, it looks like you can establish a priori something of the form:

McK-2 If you're thinking a thought with the content *`Water puts out fires`*, then . . . [here some claim about your environment, e.g. it actually does or did contain samples of water].

Putting the two together, it looks like you can conclude, purely on the basis of introspection and a priori philosophy:

McK-3 Your environment is the relevant way [e.g. it actually does or did contain samples of water].

And while few doubt that we *do know* things like McK-3, it's extraordinary that we should be able to know them *purely on the basis of introspection and*

the same,' in a way that preserves the familiar classifications. I'll just ask you to use whatever rough and ready understanding you have of this.

[2] After McKinsey 1991; see also Brueckner 1986 and Brown 1995.

[3] Notice that this is not a *description*; it doesn't just mean 'whatever content I think by saying these words to myself.' It's a way of genuinely apprehending the content, and thereby enabling yourself to refer to it. This will be important later.

a priori philosophy. That's the puzzle. It's counter-intuitive that you should be able to tell what your environment is like just on the basis of this kind of reasoning.[4]

In discussing this puzzle, some authors refer to our supposed introspective knowledge of McK-1 and our supposed philosophical knowledge of McK-2 as all being 'a priori.'[5] The name doesn't matter much. But there are important epistemic differences between introspection and the usual paradigms of the a priori. Calling them all 'a priori' can encourage confusions; some of which we'll be disentangling later. Hence, I will reserve the name '`a priori`' for justification that comes from logical understanding, philosophical reasoning, and so on. Things like '`I am now thinking about a prime number`' I'll say instead are justified through your introspective experience or awareness of your occurrent mental life.[6] I'll use the umbrella term '`by reflection`' to cover the lot. So in my terminology, the surprising result posed by McKinsey's Puzzle is that you can tell what your environment is like *purely by reflection*.

One response to the McKinsey Puzzle *endorses* the result that you can establish things like McK-3 by reflection alone. That is one way of construing **Putnam's Argument** in Chapter 1 of *Reason, Truth and History*. Putnam argues that people who have always been brains in vats can't refer to or think about those vats. (At least, not *by* using the word '`vat`.') So, in order for us to have thoughts about vats (using '`vats`'), our environment has to be a certain way: it has to be such that we haven't always been brains in vats. As it happens, we can tell by introspection that *we do* have thoughts about vats. (In fact, such thoughts are *necessary*, to be entertaining the skeptical hypotheses we are entertaining.) So it follows that we haven't always been brains in vats.[7]

Other philosophers resist that result. They say we can only establish things like McK-3 through empirical investigation. So the puzzling reasoning has to be blocked somehow. One way to block it is to go **Incompatibilist** about externalism and your ability to know the contents of your thoughts by reflection alone. The Incompatibilist says it can't be true *both* that a thought has an externalist content *and* that you're able to tell your thought has that content, just on the basis

[4] As we'll discuss later, the puzzle isn't confined to your *knowing* or having some special *authority* about the presence of water. It's already counter-intuitive that you should be able in this way to acquire any justification at all to believe there really is water in your environment.

[5] See, for example, McKinsey 1991; Boghossian 1997; McLaughlin and Tye 1998; Brown 2004: 23–4. See also Kitcher 1980: s. V.

[6] I'll avoid any substantial assumptions about what that amounts to. Bonjour 1998: 7 ff. also denies that introspective justification is a priori.

[7] Sawyer 1998; Warfield 1995 and 1998; and Tymoczko 1989 defend such approaches to the McKinsey Puzzle.

The line I described is only one interpretation of Putnam's argument. Some interpretations employ further premises about what brains in a vat *do* refer to and think about, when they use the word '`vat`'; or what *I* would refer to with '`vat`' if I were a brain in a vat. Other interpretations focus on disquotational knowledge about one's language, rather than introspective knowledge of what one is thinking.

of introspection.[8] Some Incompatibilists take the incompatibility to discredit externalism. Others take it to discredit your ability to tell what externalist thoughts you're having by reflection alone.

I think these are all overreactions to the puzzle. We can find a more sober response that steers between them. We *can* tell what we're thinking by reflection alone, even when what we're thinking is externalist. However, this *doesn't* give us a route to reflective knowledge—or even reflective justified belief—about what our environment is like. So I will argue.

As it happens, I'm *not* sure that reflective knowledge of our environment is to be avoided at all costs. After all, some epistemologists maintain that thoughts like `My senses are reliable` are justified a priori (albeit defeasibly). Here's another possible route to reflective knowledge of your environment: (i) notice, on the basis of introspection, that you're having an experience as of hands, and that you have no evidence that your senses are misleading you; (ii) apply your favorite a priori epistemology of perception to get the conclusion `So I am justified in believing that I have hands`; (iii) help yourself to a defeasible but rational ampliative inference from `I am justified in believing P` to `P`. Voilà: now you've got a purely reflective (albeit roundabout) justification to believe you have hands. It may even suffice for knowledge.[9] This merits careful discussion—especially step (iii)—but it's not obvious that the reasoning is illegitimate. Neither does it trade on any assumptions peculiar to externalism. So my examination of the McKinsey argument won't take it for granted that reflective knowledge of our environment is flat-out impossible. I just want to get the details straight. As these turn out, I think the McKinsey argument does *not* give us a route to such knowledge.[10]

II

Let's attend to the bits we left unspecified at the end of McK-2. We supposed that, if a given thought is externalist, then it'd only be available to be had by subjects in certain sorts of environments. Is that supposition right?

It'll be useful to think about an example drawn from the history of chemistry.[11]

Mendeleev presented his first periodic table in 1869. At that time, and then again more carefully in 1871, he postulated the existence of four missing elements. He called one of them '`ekaboron`.' In 1879, Lars Fredrick Nilson, unaware of Mendeleev's predictions, spectrographically identified a new element

[8] This is McKinsey's, Brueckner's, and Brown's own response to the puzzle (Brown takes it back in her 2004). See also Boghossian 1997. There are also other arguments for Incompatibilism, besides McKinsey's; two notable sources are Woodfield 1982 and Boghossian 1989.

[9] Compare 'The Explainer' in Hawthorne 2002. [10] Compare Davies 2000. fn. 4.

[11] For background, see http://web.lemoyne.edu/~giunta/ea/contents.html and http://homepage.mac.com/dtrapp/periodic.f/periodicity.html (of those sites disappear, archives of them can be found at **http://www.archive.org/**).

in some Scandinavian minerals. He managed to chemically isolate an oxide of this element; and he dubbed the element '`scandium`.' This turned out to be Mendeleev's ekaboron. At least, it occupied ekaboron's position in the periodic table; and Mendeleev had closely predicted many of its properties, such as its atomic weight, its valence, and the density of its oxide.

Now, in reality, Mendeleev thought elements were individuated by their atomic weight rather than by what we call their 'atomic number.' Protons weren't discovered until 1918. But it will simplify our discussion to pretend that Mendeleev did stipulate how many protons ekaboron has (namely, 21) in advance. Additionally, I will assume that ekaboron *was* scandium, that is, that Mendeleev was able to think and talk about this element before it was discovered in nature and chemically isolated. That assumption may well be challenged; and I'm not certain it's true. But for this discussion we'll assume it. Finally, let's imagine that Mendeleev engaged in correspondence with a friend in 1871, discoursing at length about ekaboron but *not* telling her what the element's fundamental individuating properties were.

We now have three different subjects thinking about scandium. First, there's Mendeleev, who stipulates that with '`ekaboron`' he's referring to the element with 21 protons. Second, there's Mendeleev's friend, who acquires competence with the name '`ekaboron`,' but who doesn't know how the substance is chemically individuated. Third, there's the discoverer Nilson, who independently encounters the element in nature. Each of these subjects says to himself, '`Ekaboron [scandium] is a silvery metal`.' How many different thought contents will we have?

Different externalists will answer that question differently. Some will say that since all the thoughts concern a single substance, there is only a single content. Some will say that since the substance is cognitively presented to the subjects in three different ways, there are three contents. Some might argue that Mendeleev and Nilson think different contents, but Mendeleev's friend acquires the ability to think the same contents that Mendeleev thinks; so altogether there are only two contents. I'd *like* our enquiry to apply to all these views, so I hope to avoid taking sides here as much as possible. That forces us to proceed carefully in formulating the McKinsey reasoning.

Consider next a Sort-of-Twin Earth. The Mendeleev there is internally different: he's introduced '`ekaboron`' to refer to the chemically similar element with atomic number 39. But he says all the same things about it in his correspondence; and his friend is an internal duplicate of the Earthly friend. There's also an internal duplicate of Nilson there, who's happened upon a different metal, yttrium, that turns out to have atomic number 39 and to correlate with the Twin Mendeleev's predictions.

Now how many thought contents do we have? Everyone will say Mendeleev and his twin are thinking different thoughts; after all, they're not even internal duplicates. Externalists will want to say that Nilson and his twin are thinking

different thoughts, too, despite being internal duplicates. Some externalists will want to say that Mendeleev's and Twin Mendeleev's friends are also thinking different thoughts.

Are the different thoughts had by Nilson and his twin *only available* to subjects whose environments contain samples of the respective elements? That depends on whether their thoughts are the same as Mendeleev's and Twin Mendeleev's thoughts. For plausibly Mendeleev's thoughts *don't* constitutively depend on the environmental presence of the element he's postulating. Mendeleev would have had the same postulative thoughts even if Earth turned out to contain no traces of scandium. So if Nilson's thought has the same content as Mendeleev's thought, then it's not true that that content is only thinkable by subjects whose environments really contain scandium. What may be true is that the content is not thinkable *in the way Nilson thinks it* unless one has genuinely encountered scandium. But it's tricky to say what these 'ways of thinking' amount to, without taking sides about issues that externalists disagree about.

And what about Mendeleev's friend and her twin? Even if we agree that their thoughts differ, and so are externalist, it's plausible that, like Mendeleev, they too would have had the same (deferential) thoughts even if their environment had contained no traces of scandium.[12] What their thoughts seem constitutively to depend on is not samples of scandium but rather communication with Mendeleev or a twin. But here too matters are sticky, since we don't want to take sides on whether their thoughts do or don't have the same contents as Mendeleev's and Nilson's thoughts. The other guys' thoughts don't require the existence of any communication.

Here's my attempt to finesse these issues. Let's focus on Nilson and his twin. They each reason in the following way:

(1) I'm now thinking the content `Scandium is a silvery metal`.

(2) I'm understanding this concept `scandium` to be governed by my ostensive introduction, rather than by deference to any authority; and I'm not making any definite assumptions about how it's chemically individuated.

(3) Anyone who's able to think the content `Scandium is a silvery metal` in the way described in (2) must inhabit an environment that does or did contain samples of scandium.

(4) So my environment does or did contain samples of scandium.

Note that (1)–(4) specify *a form* of argument. When Nilson thinks it through, he thinks premisses about scandium. When his twin thinks it through, he thinks premisses about yttrium. It's prima facie plausible that these subjects should know—or at least be defeasibly justified in believing—premisses (1) and (2) just on the basis of introspective reflection. And it's prima facie plausible that they should also know—or at least be defeasibly justified in believing—premiss

[12] Compare Burge 1982: 116.

(3) on the basis of a priori, Putnam- and Kripke-style philosophical reasoning. The apparent result is that they acquire some wholly reflective justification to believe (4). That's surprising. (It remains surprising even if it's allowed that their justification to believe the premisses, and hence the conclusion, is empirically defeasible.)

Mendeleev's friend would instead reason to a different conclusion: perhaps, 'So my environment does or did contain other subjects.'[13]

III

The McKinsey-style reasoning is a form of *modus ponens*. So one way to block it would be to articulate and defend constraints on when *modus ponens* reasoning is legitimate. Some responses to McKinsey's Puzzle take that form.

One move is to deny Closure. Perhaps Nilson is able to rule out the relevant alternatives to each of (1)–(3), but *more* possibilities are relevant alternatives to (4), and he's not in a position to rule them out. Maybe that means he's able to know (1) through (3), without being able to know (4). Maybe. But denying Closure is not so popular these days. Even among epistemologists who employ the framework of 'relevant alternatives,' most would rather *keep* some form of Closure.[14] In any event, I want to restrict our attention to responses to McKinsey's Puzzle that *don't* require us to deny Closure.

Crispin Wright and Martin Davies have formulated a different sort of constraint, in their discussions of 'transmission-failure.'[15] Their analyses of the McKinsey argument reward careful consideration; and they interact in interesting ways with the diagnosis I'll be giving. But I'll have to reserve discussion of them for another occasion.

In order to avail ourselves of the McKinsey-style reasoning, we need to have reflective justification to believe all its premisses *simultaneously*. Sometimes *modus ponens* reasoning fails because in acquiring justification to believe one of the premisses, one loses justification to believe another. For example, I may start out thinking, quite reasonably, that you are not a rap star. The only time I ever heard *you* rap was when we did karaoke together. So I justifiably think `If you're a rap star, then so am I`. But then you confront me with evidence of your covert musical career. I'm not now justified in affirming the antecedent and concluding that I'm a rap star too. This is because in acquiring justification to believe `You are a rap star`, I *lost* my grounds for believing the conditional `If you're a rap star, then so am I`.[16]

[13] For this version of the argument, see Burge 1982; Brown 1995; McLaughlin and Tye 1998: 312 ff. Falvey 2000; and Brown 2001, 2004: 294 ff.

[14] For discussion, see Stine 1976; Brueckner 1985; Cohen 1988; Vogel 1990; DeRose 1995: esp. S. 10; Klein 1995; Hawthorne 2005.

[15] See especially Wright 2000 and 2003; and Davies 1998, 2000, 2003a, and 2003b.

[16] Compare Harman 1973: ch. 9; Ginet 1980; Sorenson 1988.

We should take care that no such funny business is going on with the McKinsey-style reasoning. At the moment, there is no special reason to think it *is*. In the McKinsey-style reasoning, your justification to believe conditionals like:

McK-2 If you're thinking a thought with the content `Water puts out fires`, then your environment is so-and-so.

comes from reflecting on a priori thought-experiments. Why should learning that the antecedents of these conditionals are true—that you are thinking thoughts with the indicated contents—defeat or undermine your grounds for believing the conditional? At the moment, we have no ground for saying it would. But we will return to this possibility later.

For the time being, then, the way looks clear for you to legitimately combine your justification to believe premisses (1) through (3), and thereby acquire justification to believe the surprising conclusion (4).

IV

But is your justification for believing all the premisses really wholly reflective? We need to think about this more carefully.

So far we've been thinking about Nilson and his twin. They inhabitant different environments—one where scandium is distributed in certain Scandinavian minerals, the other where yttrium is so distributed—and as a result they end up thinking different thoughts. But their environments have it in common that there's *some* substance of the appropriate sort that they're interacting with. Let's call any environment of that sort a **hospitable environment**. Nilson will also have duplicates whose environments are **inhospitable**. These will be places like Boghossian's Dry Earth.[17] Nilson's unlucky duplicate there will just be *hallucinating* handling some minerals, isolating a metallic oxide, and so on. From the inside, everything will seem to him just as it seems to the real Nilson.[18] But outside, there's no substance there for his thoughts to latch on to. What *will* the Nilson in that inhospitable environment be thinking, when he says to himself, '`Scandium is a silvery metal`'?

There are a variety of answers one might give here.

Perhaps he's thinking superficial descriptive thoughts: thoughts true in case some new metallic element he's just identified in such-and-such minerals is a silvery metal. Now, no doubt *he will* be thinking such descriptive thoughts; and so too will the real Nilson. What's controversial is whether they're the thoughts that unlucky Nilson has *in* saying to himself '`Scandium is a silvery metal`.' It's reasonable that they might be. On this proposal, Nilson would be thinking descriptions that nothing in his environment *manages to satisfy*; but the

[17] See Burge 1982: 114 ff.; Boghossian 1997.
[18] Though see n. 1 for some difficulty with this.

descriptive thoughts would be available to him, despite the inhospitality of his environment.[19]

A different proposal is that, in saying to himself '`Scandium is a silvery metal`,' Nilson is unwittingly having thoughts about his own experiences. This is what projectivists say about our color concepts.[20] As it turns out, they say, our environment is inhospitable to those concepts. There are no qualities 'out there' of the right sort for our color concepts to latch onto. Instead, when we think about colors, we're really thinking about our own experiences; we just wrongly *project* qualities of those experiences onto the outside world.

Another proposal is that Nilson is unwittingly thinking about a *fictional* metal—in the same way that our thoughts about Tolkien's *mithril* are thoughts about a fictional metal. Or perhaps he's thinking about a necessarily uninstantiated substance.[21] Or perhaps his thoughts have a special, 'gappy' content that is incapable of ever being true.[22]

Or perhaps, following McDowell and Evans on demonstrative thoughts, Nilson doesn't manage to have *any* contentful thought in saying to himself '`Scandium is a silvery metal`.'[23] He may well be having various thoughts at the same time: perhaps some of the descriptive thoughts we've already identified. But none of those will be what he's thinking *in* saying those words to himself. Nothing will be. It will just falsely seem to him that he's thereby thinking anything contentful.

As I said, there are a variety of things one can say here. Keep in mind that these are all proposals about what thoughts *unlucky* Nilson is having. We shouldn't have doubts about the extension of *our own* thoughts, or the real Nilson's thoughts, with respect to the inhospitable environment. Our thoughts are about scandium, a substance that happens not to be present there. Nilson's unlucky counterpart isn't in a position to be thinking *those* thoughts. We're trying to figure out what *he is* thinking, in their place.

Some philosophers will maintain that the contents unlucky Nilson thinks are *also* contents the real Nilson thinks in saying to himself '`Scandium is a silvery metal`.' Internalists say *that's all* that the real Nilson thinks. But

[19] Boghossian 1997 argues that the Dry Earthers' *`water`* concept cannot be a 'compound, decompositional concept.' I agree. Our complex description '`the new metallic element I've just identified in these minerals`' shouldn't be understood as sharing its *logical form* with the Dry Earthers' term '`water`.' I'm just using it to specify their term's *intension*. That may be the best we can do; we may not *have* a term with the same intension *and* logical form. Boghossian seems to argue further that, if we don't ourselves have a way *to say* what concept is expressed by the Dry Earther's term '`water`,' then there is no fact of the matter what its content and intension are. Here I'm unconvinced. See McLaughlin and Tye 1998: s. VI for discussion.

[20] See, e.g., Boghossian and Velleman 1989.

[21] Consider Shoemaker's 'NI-intentional properties' in his 1990; see also Stoneham 1999.

[22] On these 'gappy' contents, see Adams, Fuller, and Stecker 1993; Adams and Stecker 1994; Braun 1993, 2005; Salmon 1998; Reimer 2001a, 2001b.

[23] See McDowell 1977, 1984, 1986; Evans 1981, 1982. Consider also McGinn 1989's 'strong externalism.'

an externalist can say the real Nilson thinks these contents too—*in addition to* thoughts about the particular natural kind he's interacting with. Externalists disagree about whether the real Nilson's thoughts do have any internally common factors of this sort. I will not here take a stand on the question.

If the unlucky Nilson *is* thinking things that the real Nilson isn't *also* thinking in saying '`Scandium...`' to himself, then by our definition of 'externalist,' the unlucky Nilson's thoughts are externalist too. For he has internal duplicates (the real Nilson and his duplicate on Twin Earth) who fail to think those same contents.

However, there's an interesting question nearby that is still unsettled: Are the contents unlucky Nilson is thinking *available* to the real Nilson? Perhaps the real Nilson needs to use different words to formulate those thoughts—e.g. perhaps he wouldn't be thinking them in saying to himself '`Scandium...`' though he would be thinking them in saying to himself '`The new metallic element I've just identified in these minerals...`'[24]

On some of the proposals about what unlucky Nilson is thinking, it's plausible that his contents *would* be available in this way to the real Nilson. On other proposals, this may be less plausible. To take one example, consider the proposal that unlucky Nilson is thinking about a fictional substance. Perhaps the relevant fiction is not cognitively available to the real Nilson. Kripke argued that the thoughts we express with '`Sherlock Holmes`' are about an essentially fictional detective. A world where some real person lives at 221B Baker Street, is called '`Holmes`,' and does such-and-such wouldn't be a world where *our* Holmes really exists. That real detective may be beyond *our* referential reach.[25] Conversely, one might argue that *our* fictional Holmes is beyond the referential reach of the real detective and his neighbors. They're not appropriately acquainted with our fiction. So too, if unlucky Nilson is thinking about a fictional element when he thinks to himself '`Scandium...`,' he may be thinking a content that we're not ourselves in a position to think. At any rate, that's a view that could be argued.

Similarly, if unlucky Nilson is thinking a *gappy* content, and gappy contents are sufficiently fine grained, then this too may be a content that's unavailable to us using any words. That's another view that could be argued.

So, collecting these possibilities together: perhaps unlucky Nilson is thinking contents that the real Nilson is *already* also thinking; or perhaps he's thinking contents that the real Nilson *could* think, using other words; or perhaps he's thinking contents that are strictly unavailable to the real Nilson. We also mooted a different possibility: on a McDowell/Evans-inspired view, unlucky Nilson isn't thinking any contentful thought at all—at least not *in* saying to himself '`Scandium is a silvery metal`.' We'll consider this last possibility carefully in later sections; set it aside for now.

[24] See fn. 19 for some difficulties here.
[25] See Kripke 1980: 157–8.

We need to keep this variety of options in mind when judging the epistemic status of the conditional McKinsey premiss:

(3) Anyone who's able to think the content *Scandium is a silvery metal* in the way I do must inhabit an environment that does or did contain samples of scandium.

Unlucky Nilson will be thinking a premiss of this form to himself. What he's thinking will probably be false. The content *he's* demonstrating by thinking to himself '`Scandium is a silvery metal`' *is* available to subjects (like himself) whose external environments never did contain any new metallic element, any of his own experiences, any essentially fictional or necessarily uninstantiated metal, or whatever it is he refers to with '`scandium`.'

How does that bear on the epistemic status of the lucky Nilsons: the one in our environment and the one on Twin Earth? Presumably, they're not in any position to know a priori that they're the lucky ones. For all they know a priori, *they* may be in inhospitable environments. If they *are* in such environments, the premisses of form (3) that *they're* thinking would be false too. So it's unclear whether they *can* know these premisses a priori. Perhaps they're only in a position to know a priori premisses of *this* form:

(3*) *If* my environment is hospitable, *then* anyone who's able to think the content *Scandium is a silvery metal* in the way I do inhabits an environment that does or did contain samples of scandium.

But that only enables them to draw conclusions about what their environment is like if it's hospitable. And it's not that surprising that one should be able to know *things like that* purely by reflection. Alternatively, perhaps they're in a position to know a priori:

(3**) The content *Scandium is a silvery metal* seems or *purports* to be a content that, when thought in the way I think it, is only available to subjects in environments that do or did contain samples of scandium.

But that only enables them to draw conclusions about what their environment purports to be like. And that too may well be intuitively allowed to be knowable by reflection. The McKinsey result only manages to be *puzzling* when it stays close to the conclusion we originally formulated.

We've just rehearsed one popular way of defusing McKinsey's Puzzle.[26] I don't think it can be a *complete* solution—as we'll see, we haven't yet covered all the bases—but it's right as far as it goes. It's often an a priori open possibility for subjects that they're in inhospitable environments, and that *if they are* in such environments, the thoughts they're having *aren't* restricted to subjects in environments containing the relevant stuffs (or expert subjects). So they won't be in a position to know the conditional McKinsey premiss a priori, after all.

[26] See Gallois and Hawthorne 1996; McLaughlin and Tye 1998: esp. s. IV; Brown 2004: ch. 8.

Is this a form of Incompatibilism? Are we saying that the premisses of McKinsey's argument can't all be true?

No. As I understand Incompatibilism, it's the thesis that it can't be true *both* that your thought has an externalist content *and* that you're able to know your thought has that content, just by reflection. We're not saying that. On the current account, for instance, the real Nilson's thought—the thought he's reflectively aware of having—*does have* an externalist content. The conditional McKinsey premiss really is true of him. He's just not in a position to know a priori that that's so.

But isn't this still Incompatibilist in a sense? Aren't we granting that if externalism is true, you *can't* after all know by reflection alone *which* thought you're having—is it the thought about scandium? the thought about yttrium? or an unlucky thought about a merely fictional stuff? You can't tell.[27]

It is part of this account that you can't know some philosophically interesting things about your thoughts—namely, whether premiss (3) is true of them—by reflection alone. I think it's plausible that whatever `scandium` concept you have can survive *learning* whether (3) is true (or false) of it.[28] We are here giving up on having complete reflective knowledge of our thoughts' nature. But that's something an internalist might give up too. (For example, perhaps it's part of the nature of even internalist thoughts that they be materially realized in such-and-such ways; but that isn't knowable just by reflection.) Even when you fail to know something's complete nature, you might still know *what* that thing is. For example, even if you fail to know the complete nature of the metal that composes your ring, you might still know what metal it is. You might *know that* it's gold your ring is made of. McKinsey's argument threatens your having even this kind of epistemic achievement with respect to your thoughts, on the basis of reflection. It threatens your being able to tell by reflection *that* you're thinking `Water puts out fires`, or *that* you're thinking `Scandium is a silvery metal`. The present account defuses this threat. It says: Yes, you can know you're thinking those contents. You just don't know a priori what constraints the thinking of them places on a subject's environment.

Consider this analogy. You're a newcomer to Metropolis. On Monday, you meet newspaperman Jimmy Olson. On Tuesday you meet his colleague Clark Kent. On Wednesday you witness crimes being stopped by a red-caped superhero the locals call '`Superman`.' You don't realize you already met him in disguise the day before. Now are you in a position to know a priori that Clark Kent=Superman? Most philosophers will say no. A few will say that the proposition that Clark Kent=Superman is identical to the proposition that Superman=Superman, and that *you are* in a position to know that a priori.

[27] See here Boghossian 1994.

[28] Compare: before learning the facts about jade, we believed `Jade is a single natural substance`; after learning the facts, we denied that very same belief.

However, *all* theorists should agree that you're not yet in a position to know that *Jimmy Olson*≠Superman. That's just not something you could know without doing some investigation and gathering certain kinds of evidence. However, it's no obstacle to your knowing who you're talking to, when you're talking to Jimmy Olson. You can know you're talking to Jimmy without knowing everything interesting about him: his passport number, that he's distinct from Superman, and so on. (*On Monday*, you were doubly unable to know that Jimmy Olson≠Superman: in the first place because you hadn't done the necessary investigation, and in the second place because you hadn't yet encountered Superman, even in disguise, and no one had mentioned him to you. So you weren't yet capable of thinking *anything* about him.)

Your situation with respect to `water` and `scandium` thoughts is much the same. You can know by reflection that you're thinking this content, `Scandium is a silvery metal`, without knowing everything interesting about it—that it's about a substance with 21 protons, that it's not about yttrium, that it's about a real substance rather than a fictional one, and so on. (Here, too, you're *doubly* unable to know your thoughts don't concern yttrium: in the first place because you haven't done the necessary investigation, and in the second place because you're not yet in a position to think anything about yttrium.)[29]

V

Let's return to the McDowell/Evans view that we set aside. On that view, the unlucky Nilson isn't thinking any thought when he says to himself '`Scandium is a silvery metal`.' What if the real Nilson knew a priori that *that's* the correct account of the cognitive life of subjects in inhospitable environments? Then he'd know a priori that anyone who *does* manage to think contentfully *is* in a hospitable environment. Hence, he apparently *would* be able to know a priori that:

(3) Anyone who's able to (genuinely) think the content `Scandium is a silvery metal` in the way I do (or seem to) must inhabit an environment that does or did contain samples of scandium.

So the surprising McKinsey result is still with us.[30] And we wouldn't get off the hook just because Nilson *failed to know* that the McDowell/Evans account is

[29] Compare Falvey and Owens 1994's contrast between introspective knowledge of content and introspective knowledge of comparative content.

These issues are why I was so careful about how we're identifying contents in our discussion. If all you were in a position to know was `Whatever content I think by saying such-and-such is so-and-so`, then the complaint that you don't know what you're thinking might have some teeth. As it is, though, I think you're able to think these contents, apprehend and demonstrate them in doing so, and know by reflection *that* you're thinking them, despite not knowing whether premiss (3) is true of them.

[30] Gallois and Hawthorne 1996 and Boghossian 1997 stresses this. Wright and Davies also doubt that we can count on the strategy from section IV, above, always being available. They characterize themselves as giving responses to the 'worst-case scenarios,' where (3) *is* knowable a priori.

correct. If he merely has some undefeated a priori *justification to believe* it's the right account, that's bad enough. For then he'd have some a priori justification to believe (3) is true, and as a result, some reflective justification to believe the McKinsey result. That can seem just as counter-intuitive as saying he could reflectively *know* the McKinsey result. It's counter-intuitive that Nilson should be able to get any justification at all about his environment *in this reflective way*. And that prospect is still on the table, so long as he has any undefeated a priori justification to think the McDowell/Evans view might be correct.

We need to examine this possibility closely. In handling it, I'm going to make use of apparatus that I've argued for in other papers.[31] So I'll begin by summarizing the points I want to import.

In those other papers, I argued that just because a certain type of thought is true whenever it's entertained, it does not follow that subjects thereby have a priori justification to believe it's true. On a McDowell/Evans-style view, for instance, thoughts of the form:

(5) Jack exists.

and

(6) I perceive *this wall* [here you perceptually demonstrate a wall].

will be true whenever they're successfully entertained.[32] If Jack didn't exist, or you weren't perceiving a wall, then your referential attempts would fail and on this view you wouldn't be thinking contentful thoughts of the specified forms. But plausibly (5) and (6) *aren't* the sorts of thing you can know a priori. (In our present discussion, it's under dispute whether you can know them *by reflection*: that is, relying both on a priori sources and your introspective awareness. But in those papers, and for the moment here, I'm just talking about whether these claims are knowable strictly a priori.)

On the other hand, I argued that 'hedged' thoughts like:

(7h) If Jack exists, then he is self-identical.

may be knowable a priori. Not because they're true whenever they're entertained; but rather through your understanding of the logical relations they involve. Here matters get somewhat subtle. On the McDowell/Evans view, you won't be able to entertain (7h) unless Jack exists, too. But I argued for a distinction between *what has to be the case for a thought to have a given content*, and *what are the proper*

[31] Pryor 2006a, 2006b.

[32] Here some issues we tried to side-step in s. II re-arise. Millians will argue that thoughts with the same *content* as (6) can be entertained falsely: e.g. if you refer to the wall using a name acquired from other speakers, and aren't then perceiving it. Many Millians will grant that that would amount to entertaining (6) in a different 'way' than I'm envisaging in the text. If so, my talk of 'thoughts of such-and-such a form' should be read to include not just the thought's content, but also a relevant way of thinking it.

logical entailments of the thought. I think the epistemically perspicious logic for these thoughts is a free logic on which (7h) *doesn't* logically entail (5). The virtue of such a view is that it can do justice to our pre-theoretic impression that:

- experience plays some role in justifying (5); and
- it doesn't seem rational to believe (5) to degree 1.

Whereas it seems that:

- there *should* be some purely logical truth, along the lines of (7h), that we could rationally believe just on the basis of our logical understanding; and
- that we could have more rational confidence in than we have in (5).

So on the view I proposed, (7h) did not logically entail (5). For a McDowell/Evans-inspired theorist, you wouldn't be able to *contentfully think* (7h) unless (5) were true. Still, I argued, if (5) *were* true, and (7h) thinkable, then your logical understanding of (7h) should give you a priori justification to believe what you are thinking, without giving you a priori justification to believe that *you're thinking it*; nor that (5) is true. You can be in a position to think the thought, and have a priori justification for what you are thinking, without being in a position to tell a priori that that is so. It takes *more* to justify you in believing you *really have* the thought than it takes to justify the thought itself.

When it's an open question for you whether Jack exists, it may seem peculiar to try to think thoughts like (7h), without yet knowing that you'll succeed. (It's somewhat like prefacing an email with 'If you're still reading email at this address . . . ' That insures you no better against not being read.) However, you don't really have better alternatives. Your ability to *withhold belief* in (7h), or *be agnostic* towards (7h), is just as threatened by the prospect of Jack's not existing as full belief towards (7h) would be. There is such a condition as *having no degree of belief at all* towards a thought: e.g. if you've never considered the thought, or are incapable of considering it. But when you *are* considering (7h)—as it happens, successfully—then you'll *have to* have some cognitive attitude towards it that wouldn't be available were Jack not to exist. You just won't be in a position to know a priori that you do.

Now, experiences will be necessary for you to be able to think either of (5) or (7h). But sometimes the experiences that enable you to think a thought are part of what justifies you in believing it; and other times they merely enable the thinking. Compare: in order to have the concept `chromatic`, and believe:

(8) Chromatic colors are colors.

perhaps you need to have had visual experiences. But plausibly visual experiences play no role in justifying your belief. On the other hand, in order to be able to think the demonstrative thought:

(6) I perceive *this wall* [here you perceptually demonstrate a wall].

you also need to have perceptual experiences. And here the experiences *are* plausibly part of what justify your belief. I maintain that `Jack exists` is a posteriori, like (6), whereas `If Jack exists, then he's self-identical` is a priori, like (8). From the premiss *that* you're thinking this latter thought, you may be able to infer a priori that Jack exists. But from the thought itself, you cannot.

I've been talking so far about a priority, strictly construed. What I've said leaves it open whether we can know `Jack exists` and `I perceive this wall` by *reflection*. One view will say you need ordinary *perceptual* justification to believe the wall exists, *in order to* be able to understandingly think the demonstrative thought (6). Similarly, you'll need ordinary perceptual evidence about Jack to even understand the hypothesis that he exists. A different view will say that you can be justified in believing these thoughts solely from your introspective awareness of seeming to successfully think them, and your a priori knowledge that thoughts of these types need referents. That is, you could have McKinsey-style, wholly reflective, justification to believe them. My sympathies lie with the former view. But at this stage, we still need to regard both alternatives as open.

VI

Armed with that background, we're now in a position to complete our diagnosis of the McKinsey-style reasoning.

In some ways, the apparatus I've introduced *aids* the McKinsey-style reasoning. For we suggested that on a McDowell/Evans-style view, the premiss:

(3) Anyone who's able to (genuinely) think the content `Scandium is a silvery metal` in the way I do (or seem to) must inhabit an environment that does or did contain samples of scandium.

will be knowable a priori. And you might balk at that, thinking that if for all you know a priori, you're in an inhospitable environment and thinking nothing contentful by saying those words to yourself, then you *can't* know this premiss a priori—even if, as it happens, you're lucky and *are* contentfully thinking the premiss. The apparatus I've introduced makes it possible to finesse that worry. We'll replace (3) with the 'hedged' claim:

(3h) If the content `Scandium is a silvery metal` [which I hereby seem to be thinking] really exists, then anyone who is able to genuinely think it in the way I seem to must inhabit an environment that does or did contain samples of scandium.

If the McDowell/Evans-style view is correct, there's still no contentful thought of this form for unlucky Nilson to have. But if the lessons I summarized are right, then such a hedge will give us a form of premiss that subjects in hospitable

environments will be able to think *and* may well have a priori justification for. Unlike the hedges (3*) and (3**) we considered earlier, this hedge doesn't yet obviously handicap the McKinsey Argument.

Premiss (2) can be hedged in a similar way. What about premiss (1)?

There are various things Nilson knows by reflection that fall short of premiss (1), and are *not* enough, given what else he knows by reflection, to infer that his environment contains samples of scandium. For instance, he can know by reflection *I seem to be thinking a thought by rehearsing to myself the sentence* 'Scandium is a silvery metal.' He can know by reflection *I seem to be having a thought that attributes being a silvery metal to some substance*. These are general, non-externalist thoughts. Their contents don't even purport to be constitutively dependent on scandium; so they are thoughts that even the unlucky Nilson in his inhospitable environment can be in a position to think, and to know by reflection. As I said, they're too weak to provide knowledge by reflection of a McKinsey-style conclusion.

However, if they're *all* we can know about our thoughts by reflection, we should be disappointed. On a McDowell/Evans-style view, I think *they are* all that unlucky Nilson can know. But the real Nilson, and his duplicates in hospitable environments, can do better.

Let's consider a hedged version of premiss (1):

(1h) If the content *Scandium is a silvery metal* really exists, then I am now thinking it.

The original premiss (1) would amount to the combination of (1h) and:

(1∃) The content *Scandium is a silvery metal* really does exist.

On a McDowell/Evans-style view, neither (1h) nor (1∃) will be contentfully thinkable unless you're in a hospitable environment. But if you *are* lucky enough to be able to think them, then I think it's plausible that (1h) should be justified for you just on the basis of your introspective awareness of your occurrent mental life. You don't need first to *establish that* your environment is hospitable. And it's (1h) *itself* you have justification for: not just a metalinguistic claim, like *If (1h) expresses a thought, then what it expresses is true*. (1h) itself constitutes a more intimate and satisfying reflective knowledge of what you're thinking than what we considered before.

What about (1∃)? Is it *also* knowable by reflection? That's not so clear. We're supposing it to be an open epistemic possibility that you're having a McDowell/Evans-style illusion of contentful thought. That suggests that *you won't* be able to know (1∃), not until you acquire some evidence that you're really one of the lucky subjects in a hospitable environment. I'm not saying you need justification to believe your environment is hospitable in order to *think*

Scandium is a silvery metal. Neither do you need it in order to be *justified in believing* every *scandium* thought: the thought *If scandium exists, it's self-identical* may be justified a priori. What I'm saying is that—to the extent that McDowell/Evans-style illusions are an open possibility—you need antecedent justification to believe your environment is hospitable to be justified in believing that *you're genuinely thinking* *scandium* thoughts. That sacrifices *some* reflective knowledge about your own mind, but it's a sacrifice I think we can be comfortable with. We can leave your justification to believe (1h) in place. Like the claim *If Jack exists, then he's self-identical*, I think (1h) is something you can have justification to think *without* yet having justification to believe you *can think* it. The source of your justification to believe (1h), and the degree to which you're justified in believing it, may be different from the source and degree to which you're justified in believing you do successfully think (1h).

For the moment, then, here's what I propose. Lucky Nilson is (perhaps without knowing it) in a position to think premisses (1h), (1∃), and so on. His introspective awareness of his thoughts justifies him in believing (1h), that he's *purporting* to think a certain content: namely this one [here he mentally produces the content *Scandium is a silvery metal*]. Nilson doesn't know by reflection *whether* he's succeeded in thinking anything. But in fact *he has* succeeded, and introspection does justify him in believing that *this* is what he's thinking, if he's thinking anything. Because Nilson can't tell by reflection alone that he is successfully thinking, he's not in a position to infer via McKinsey reasoning that his environment is a hospitable one. That takes further, empirical evidence.

The relevant empirical evidence will be readily had: the same experiences that enabled Nilson to think about scandium, and apply the name '*scandium*' to it, will also justify him in believing it really is present in his environment.[33] So his thought *My environment contains scandium* will be like the thought we considered earlier, *I perceive this wall*. Assuming you have no defeating or undermining evidence,[34] no *more* experiences may be needed, to justify you in believing the thoughts, than it takes to be able to think them. Nonetheless, your experiences in these cases are playing an essential justifying role; and so your justification will be perceptual rather than reflective.

[33] Brewer 2000: 428–9 says this too. He seems to think that's the full story: our justification for believing (1) is wholly and univocally empirical. As you'll see in a moment, I think the story is less straightforward.

[34] What if you *do* have defeating or undermining evidence? Consider a case where scandium does exist, and you see it, but you have overwhelming evidence that you're hallucinating. In such a case, you might think, falsely, that *That stuff* [ostending what you take to be a hallucination, but is in fact some scandium] *doesn't exist*. If you believed the McDowell/Evans-style view to be correct, you might also think, falsely, that you just then failed to think any contentful demonstrative thought. You may even have all things considered *justification* to believe those things: to believe of scandium that it doesn't exist, *and* that your attempts to think so are unsuccessful. You may know, too, that *if* your attempts to think so *were* successful, they'd have to be false. This is delightfully perverse, and all merits closer discussion. But I see no paradox.

Now, the account I've just sketched is not mandatory. Though we may be comfortable saying you can't *know* (1∃) by reflection alone, we may feel uneasy saying you don't even have reflective justification to believe (1∃). As I suggested at the end of section V, a different view is that your introspective awareness of seeming to successfully think a thought does give you some justification to believe *there is* a content that you're thinking. On that view, you *would* have some reflective justification to believe premiss (1∃). And if you continued to have reflective justification to believe (2) and (3h), then it looks like you might achieve reflective knowledge, or reflective justification to believe, that your environment contains scandium, after all.[35]

But let's proceed carefully. There's plausibility to the idea that introspection gives you some justification to believe (1∃). There's also plausibility to the idea that you can have reflective justification to believe that the McDowell/Evans view, and hence premiss (3h), are correct. But in order for the McKinsey-style reasoning to carry through, you need reflective justification to believe *the conjunction* of these premisses. As we saw in section III, sometimes justification to believe one premiss in a modus ponens argument defeats or undermines your justification to believe another. I think something like that is happening here.

If having a McDowell/Evans-style illusion of contentful thought is an open possibility for you, then presumably any justification you have to believe (1∃) will be *defeasible*. One kind of evidence that would contribute towards its defeat would be evidence that the McDowell/Evans view is correct and you *really are in* an inhospitable environment. The more confident you became that that was so, the less confident you could rationally be in (1∃). That's one kind of defeating evidence. The analog in the perceptual case would be getting evidence that you're hallucinating things other than as they are. In the perceptual case, there are also other kinds of defeating evidence. For example, there's evidence merely that you're dreaming, or evidence merely that your eyes don't work. These undermine your entitlement to think *you're perceiving* your environment, without (directly) justifying you in believing that things are other than they appear. Now, consider evidence that your eyes are so configured that it would look to you as though you have hands regardless of whether you really do have hands. This isn't (directly) evidence that you lack hands. Neither is it (directly) evidence that you *are* right now failing to perceive. But plausibly, it too will to some degree undermine your perceptual justification to believe you have hands.

I think the McDowell/Evans view undermines your introspective justification to believe (1∃) in the same way. The more confident you become that the McDowell/Evans view is correct, and that inhospitable environments induce illusions of contentful thought, the less entitled you are to think that introspection

[35] This is more or less the position Wright thinks we're in with respect to *some* versions of McKinsey's Argument: see Wright 2003: s. V. The theorists I cited in n. 7 think it's our position with respect to *every* propery-formulated version of the argument.

tells you that an episode of seeming-to-think is an episode of really-thinking.[36] This is so even *in advance* of your getting evidence to think *you are* in an inhospitable environment.

So I will agree that your introspective awareness of seeming to successfully think a thought gives you *some* justification to believe *there is* a content that you're thinking. That justification is ordinarily seamless: you don't have to take any inferential step from `This is what I seem to be thinking` to `This is what I really am thinking`. No more than you have to take any inferential step from `This is how things look to me` to `This is how things are`. In both cases, though, additional evidence can undermine your justification to believe the stronger claim, and leave you merely with justification to believe the weaker.

This undermining will usually be a matter of degree. Usually your situation will be one of having *some* degree of rational confidence that you are genuinely aware of real thoughts, and a real perceptual environment, and *some* degree of rational confidence that introspection alone, or your unaided senses, lack the authority to tell you that's so. As the balance of evidence changes, you will seesaw between these alternatives.[37]

Applying this to our present discussion means you will seesaw between having reflective justification to believe not just (1h) but (1∃) too, on the one hand, and having reflective justification to believe (3h), and that it's beyond the authority of introspection *to tell you* that you're having a genuine thought, on the other. You'll usually have some degree of reflective justification on each side. But that just means you're superposed between two epistemic situations, neither of which by itself justifies you in believing all the premisses the McKinsey-style reasoning requires. I don't see any persuasive reason to think you ever have reflective justification to believe *the conjunction* of (1∃) and (3h). Until you do, you're in no position to draw inferences about your environment, by reflection alone.

[36] Wright 2000 s. VI also entertains (but in the end rejects) the possibility that your theoretical justification to believe the externalist premise might 'compromise' your justification to believe the introspection premise.

[37] This is not precisely the same as the scenario we discussed in s. III. There you started with justification to believe a conditional, that was undermined by the evidence you got for its antecedent. Here I expect you start out with reflective justification to believe the antecedent—(1∃) —and that gets undermined by the reasons you get to believe the conditional—(3) or (3h). The cases are interestingly similar.

Notice that it's not (3) or (3h) *itself* that's doing the undermining. *Empirical* evidence that your environment is hospitable, and hence that your `scandium` thoughts are only available to other subjects whose environments contain scandium, too, needn't undermine your ability to tell by reflection that you're having genuine thoughts. Rather, it's *the particular philosophical ground the McDowell/Evans view gives you* for believing (3h) that's the villain.

VII

Let me close with one final wrinkle. Consider a mathematician who carefully reasons through a proof that has some subtle undetected flaw. Is the mathematician justified in believing the conclusion? Well, perhaps he has some inductive justification: he knows that proofs in that journal, and proofs he finds compelling, have a good track-record. But set that aside. Does he have any a priori or at least *reflective* justification to believe the conclusion? Does his experience of seeming to deduce the conclusion give him any justification for it?

I don't know what's the best thing to say about that. I can be persuaded either way. Suppose we decide to say that reasoning through a flawed proof *can* sometimes give one reflective justification to believe its conclusion. We might say that one has reflective justification, but it's not *impeccable*. It has some flaw the discovery of which would undermine the justification. But we may want to allow that, if you're reasonably ignorant of the flaw, you do have reflective justification.

If that's what we decide to say, then we should probably say the same about McKinsey's argument, too. I've argued that the McKinsey-style reasoning *is* flawed. However, subjects may not have seen these flaws. They may find the reasoning pretty compelling, and perhaps, given their unenlightened state, they're even reasonably entitled to do so. If we allow the mathematician to have reflective justification to believe his conclusion, then we ought also to allow these unenlightened subjects to have reflective justification to believe McKinsey-style conclusions about their environments. What they'll lack is *impeccable* reflective justification. That's what I've intended to be discussing throughout.[38]

REFERENCES

Adams, F., Fuller, G., and Stecker, R. 1993: 'Thoughts without Objects.' *Mind and Language* 8: 90–104.

_____ and Stecker, R. 1994: 'Vacuous Singular Terms.' *Mind and Language* 9: 387–401.

Boghossian, P. 1989: 'Content and Self-Knowledge.' *Philosophical Topics* 17: 5–26. Reprinted in Ludlow and Martin 1998: 149–73.

_____ 1994: 'The Transparency of Mental Content.' *Philosophical Perspectives* 8: 33–50.

_____ 1997: 'What the Externalist Can Know A Priori.' *Proceedings of the Aristotelian Society* 97: 161–75. Also appears in C. Wright, M. Smith, and C. Macdonald (eds.), *Knowing Our Own Minds: Essays in Self-Knowledge* (Oxford: Oxford University Press) 1998: 271–84.

[38] Compare s. 6 of Pryor 2004. I was aided in coming to this view by discussions with Stephen Schiffer of his 2005.

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

——— and Velleman, D. 1989: 'Color as a Secondary Quality.' *Mind* 98: 81–103.

Bonjour, L., 1998: *In Defense of Pure Reason* (Cambridge: Cambridge University Press).

Braun, D. 1993: 'Empty Names.' *Noûs* 27: 449–69.

——— 2005: 'Empty Names, Fictional Names, Mythical Names.' *Noûs* 39: 596–631.

Brewer, B. 2000: 'Externalism and A Priori Knowledge of Empirical Facts.' In P. Boghossian and C. Peacocke (eds.), *New Essays on the A Priori* (Oxford: Oxford University Press), 415–32.

Brown, J. 1995: 'The Incompatibility of Anti-Individualism and Privileged Access.' *Analysis* 55: 149–56. Reprinted in Ludlow and Martin 1998: 185–93.

——— 2001: 'Anti-Individualism and Agnosticism.' *Analysis* 61: 213–24.

——— 2004: *Anti-Individualism and Knowledge* (Cambridge, Mass.: MIT Press).

Brueckner, A. 1985: 'Skepticism and Epistemic Closure.' *Philosophical Topics* 13: 89–117.

——— 1986: 'Brains in a Vat.' *Journal of Philosophy* 83: 148–67.

——— 1992: 'What an Anti-Individualist Knows A Priori.' *Analysis* 52: 111–18. Reprinted in Ludlow and Martin 1998: 197–205.

Burge, T. 1982: 'Other Bodies.' In A. Woodfield (ed.), *Thought and Object* (New York: Oxford University Press), 97–120.

Cohen, S. 1988: 'How To Be a Fallibilist.' *Philosophical Perspectives* 2: 91–123.

Davies, M. 1998: 'Externalism, Architecturalism, and Epistemic Warrant.' In C. Wright, M. Smith, and C. Macdonald (eds.), *Knowing Our Own Minds: Essays in Self-Knowledge* (Oxford: Oxford University Press), 321–61.

——— 2000: 'Externalism and Armchair Knowledge.' In P. Boghossian and C. Peacocke (eds.), *New Essays on the A Priori* (Oxford: Oxford University Press), 384–414.

——— 2003a: 'The Problem of Armchair Knowledge.' In S. Nuccetelli (ed.), *New Essays on Semantic Externalism and Self-Knowledge* (Cambridge, Mass.: MIT Press), 23–55.

——— 2003b: 'Externalism, Self-Knowledge, and Transmission of Warrant.' In M. Frápolli and E. Romero (eds.), *Meaning, Basic Self-Knowledge, and Mind: Essays on Tyler Burge* (Stanford, Calif.: CSLI Publications), 105–30.

DeRose, K. 1995: 'Solving the Skeptical Problem.' *Philosophical Review* 104: 1–52.

Dretske, F. 1995: *Naturalizing the Mind* (Cambridge, Mass.: MIT Press).

Evans, G. 1981: 'Understanding Demonstratives.' In H. Parret and J. Bouveresse (eds.), *Meaning and Understanding* (New York: Walter de Gruyter). Reprinted in G. Evans, *Collected Papers* (Oxford: Oxford University Press, 1985), 291–321.

——— 1982: *The Varieties of Reference*, ed. J. McDowell, (Oxford: Oxford University Press).

Falvey, K. 2000: 'The Compatibility of Anti-Individualism and Privileged Access.' *Analysis* 60: 137–42.

Falvey, K. and Owens, J. 1994: 'Externalism, Self-Knowledge and Skepticism.' *Philosophical Review* 103: 107–37.

Gallois, A. and Hawthorne, J. 1996: 'Externalism and Scepticism.' *Philosophical Studies* 81: 1–26.

Ginet, C. 1980: 'Knowing Less By Knowing More.' *Midwest Studies in Philosophy* 5: 151–61.

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

Hawthorne, J. 2002: 'Deeply Contingent A Priori Knowledge.' *Philosophy and Phenomenological Research* 65: 247–69.

——— 2005: 'The Case for Closure.' In M. Steup and E. Sosa (eds.), *Contemporary Debates in Epistemology* (Malden, Mass.: Blackwell), 26–43.

Kitcher, P. 1980: 'A Priori Knowledge.' *Philosophical Review* 89: 3–23.

Klein, P. 1995: 'Skepticism and Closure: Why the Evil Genius Argument Fails.' *Philosophical Topics* 23: 213–36.

Kripke, S. 1980: *Naming and Necessity* (Cambridge, Mass.: Harvard University Press).

Lewis, D. 1979: 'Attitudes De Dicto and De Se.' *Philosophical Review* 88: 513–43. Reprinted in D. Lewis, *Philosophical Papers*, vol. I (New York: Oxford University Press, 1983), 133–56.

Ludlow, P. and Martin, N. (eds.) 1998: *Externalism and Self-Knowledge*. Stanford, Calif.: CSLI Publications).

McDowell, J. 1977: 'On the Sense and Reference of a Proper Name.' *Mind* 86: 159–85. Reprinted in J. McDowell, *Meaning, Knowledge and Reality* (Cambridge, Mass.: Harvard University Press, 1998), 171–98.

——— 1984: 'De Re Senses.' In C. Wright (ed.), *Frege: Tradition and Influence* (Oxford: Blackwell), 98–109. Reprinted in J. McDowell, *Meaning, Knowledge and Reality* (Cambridge, Mass.: Harvard University Press, 1998), 214–27.

——— 1986: 'Singular Thought and the Extent of Inner Space.' In P. Pettit and J. McDowell (eds.), *Subject, Thought and Context* (Oxford: Oxford University Press), 137–68. Reprinted in J. McDowell, *Meaning, Knowledge and Reality* (Cambridge, Mass.: Harvard University Press, 1998), 228–59.

McGinn, C. 1989: *Mental Content* (Oxford: Blackwell).

McKinsey, M. 1991: 'Anti-Individualism and Privileged Access.' *Analysis* 51: 9–16. Reprinted in Ludlow and Martin 1998: 175–84.

McLaughlin, B. and Tye, M. 1998: 'Externalism, Twin Earth, and Self-Knowledge.' In C. Wright, M. Smith, and C. Macdonald (eds.), *Knowing Our Own Minds: Essays in Self-Knowledge* (Oxford: Oxford University Press), 285–320.

Nuccetelli, S. (ed.) 2003: *New Essays on Semantic Externalism and Self-Knowledge* (Cambridge, Mass.: MIT Press).

Pryor, J. 2004: 'What's Wrong with Moore's Argument?' *Philosophical Issues* 14: 349–78.

——— 2006a: 'Hyper-Reliability and A Priority.' *Proceedings of the Aristotelian Society* 106: 327–44.

——— 2006b: 'More on Hyper-Reliability and A Priority,' at www.jimpryor.net/research/papers/More-Hyper.pdf

Putnam, H. 1981: *Reason, Truth and History* (Cambridge: Cambridge University Press).

Raffman, D. 1998 'First-Person Authority and the Internal Reality of Beliefs.' In C. Wright, M. Smith, and C. Macdonald (eds.), *Knowing Our Own Minds: Essays in Self-Knowledge* (Oxford: Oxford University Press), 363–9.

Reimer, M. 2001a: 'A "Meinongian" Solution to a Millian Problem.' *American Philosophical Quarterly* 38: 233–48.

——— 2001b: 'The Problem of Empty Names.' *Australasian Journal of Philosophy* 79: 491–506.

Salmon, N. 1998: 'Nonexistence.' *Noûs* 32: 277–319.
Sawyer, S. 1998: 'Privileged Access to the World.' *Australasian Journal of Philosophy* 76: 523–33.
Schiffer, S. 2005: 'Paradox and the A Priori.' In T. Szabó Gendler and J. Hawthorne (eds.), *Oxford Studies in Epistemology*, vol. I.
Segal, G. 1989: 'The Return of the Individual.' *Mind* 98: 39–57.
Shoemaker, S. 1990: 'Qualities and Qualia: What's in the Mind?' *Philosophy and Phenomenological Research* suppl. vol. 50: 109–31. Reprinted in S. Shoemaker, *The First-Person Perspective and Other Essays* (Cambridge: Cambridge University Press, 1996), 97–120.
Sorensen, R. 1988: 'Dogmatism, Junk Knowledge, and Conditionals.' *Philosophical Quarterly* 38: 433–54.
Stine, G. C. 1976: 'Skepticism, Relevant Alternatives, and Deductive Closure.' *Philosophical Studies* 29: 249–61.
Stoneham, T. 1999: 'Boghossian on Empty Natural Kind Concepts.' *Proceedings of the Aristotelian Society* 99: 119–22.
Tye, M. 1995: *Ten Problems of Consciousness* (Cambridge, Mass.: MIT Press).
_____ 2000: *Consciousness, Color, and Content* (Cambridge, Mass.: MIT Press).
Tymoczko, T. 1989: 'In Defense of Putnam's Brains.' *Philosophical Studies* 57: 281–97.
Vogel, J. 1990: 'Are there Counter-Examples to the Closure Principle?' In M. D. Roth and G. Ross (eds.), *Doubting: Contemporary Perspectives on Skepticism* (Dordrecht: Kluwer), 13–27.
Warfield, T. 1995: 'Knowing the World and Knowing Our Minds.' *Philosophy and Phenomenological Research* 55: 525–45.
_____ 1998: 'A Priori Knowledge of the World: Knowing the World By Knowing Our Minds.' *Philosophical Studies* 92: 127–47.
Woodfield, A. 1982: 'Introduction.' In A. Woodfield (ed.), *Thought and Object* (New York: Oxford University Press).
Wright, C. 2000: 'Cogency and Question-Begging: Some Reflections on McKinsey's Paradox and Putnam's Proof.' *Philosophical Issues* 10: 140–63.
_____ 2003: 'Some Reflections on the Acquisition of Warrant By Inference.' In S. Nuccetelli (ed.), *New Essays on Semantic Externalism and Self-Knowledge* (Cambridge, Mass.: MIT Press), 57–77.
_____ Smith, M., and Macdonald, C. (eds.) 1998: *Knowing Our Own Minds: Essays in Self-Knowledge* (Oxford: Oxford University Press).

9

A Priority and Externalism

John Hawthorne

1. INTRODUCTION

If an epistemological distinction fails to carve at the epistemological joints, then it is not worthy of serious and protracted discussion. The residual issue of whether the putative distinction is incoherent or merely gerrymandered ought not to strike anyone as especially important. My own externalist commitments—epistemological and semantic—lead me to think that the a priori–a posteriori distinction is not a particularly natural one, and hence that its importance to epistemology has been grossly overestimated.[1] I shall be defending that perspective in what follows. Discussions of a priority typically assume that there is a distinguished subclass of beliefs whose epistemic status does not depend upon experiential encounter with the world. Yet the project of delimiting such a class in a non-gerrymandered fashion turns out to be surprisingly difficult. As an organizing theme, I shall examine the bearing of a safety-based account of knowledge on the traditional conception of a priori knowledge.

2. ENVIRONMENT INDEPENDENCE

Epistemologists divide according to whether they take knowledge or, instead, some kind of justification (or 'warrant' or 'rationality' or 'entitlement') as the starting point for foundational enquiry. In line with my preferred orientation,

I am grateful to Elizabeth Camp, Maya Eddon, and Sanford Goldberg for comments on an earlier draft, and to Timothy Williamson and an audience at Oxford for helpful discussion. I am especially grateful to David Manley for discussions that helped to develop and sharpen the material in this chapter. We hope to develop some of these themes further in *Thoughts and their Objects*, forthcoming

[1] A similar conclusion is argued for in Sawyer 2001.

much of the discussion that follows will concern a priori knowledge. In a final section, I shall speak to the topic of a priori justification.

Let us begin with a simple externalist picture of knowledge, a version of the safety account:[2]

S knows p iff there is no close world where S makes a mistake that is relevantly similar to his actual belief that p.[3]

That there is a close world where S is mistaken about p does not show that he does not actually know p, since he may use a relevantly different method at that close world. (I could easily have asked an unreliable informant the score of the game rather than checked the scoreboard, and at a close world where the score is different the informant provides me with the score that obtains at the actual world.) And that there is no close world where S is mistaken about p does not show that he actually does know that p. For there may be a close world where S is not mistaken about P but does make a mistake that is sufficiently similar to the actual case in relevant respects. (Suppose I form a demonstrative belief of Jones that he is Jones but could easily have formed a demonstrative belief of Jones's Twin—who was also in the room—that he is Jones. The possible mistake concerns a different proposition but still undermines actual knowledge.)[4] This point is relevant to evaluating beliefs in mathematical necessities too. Suppose I use an unreliable method to form a belief in a mathematical necessity, P. Even though I could not easily have believed P falsely, there are relevantly similar cases in which I am in error.

My interest here is not in whether the picture allows us to make sense of mathematical knowledge: certainly it does. It is rather to ask what a priority comes to when knowledge is conceived along these lines. It is often thought that in a case of a priori knowledge, the status of a belief as knowledge does not constitutively depend on the external environment (this being one natural take on the idea that a priori knowledge is independent of perceptual experience). Let us pursue this idea.[5]

[2] Versions of the safety idea can be found, inter alia, in Sosa 1999 and Williamson 2000. The formulation here is different.

[3] The discussion of empty thoughts below indicates that we need to be liberal about the notion of 'mistake'.

[4] The picture is not intended as a way of adjudicating cases where intuitions vary about whether someone knows. At best it offers a kind of diagnosis of such disputes: intuitions will vary as to which kinds of possible mistakes are relevantly similar, and coordinately, which kind of individuation of methods is relevant to the case as hand. Insofar as knowledge is conceptually basic, no specification of relevant similarity conceptually prior to knowledge will be available.

[5] One finds something like this idea in Philip Kitcher's well-known attempt (1980) to capture the traditional conception of the a priori. Here is his account: (i) X knows p a priori iff X knows p and X's belief was produced by a process which is an a priori warrant for it. (ii) α is an a priori warrant for X's belief that p if and only if α is a process such that, given any life e sufficient for X for p, then (a) some process of the same type could produce in X a belief that p, (b) if a process of the same type were to produce in X a belief that p then it would warrant X in believing

Here is a crude first pass:

x knows p a priori after duration d of x's existence iff any possible intrinsic duplicate y of x knows p after duration d of y's existence.[6]

The truth of semantic externalism prohibits us from this way of articulating the idea. My semantic life does not supervene on my inner life. As a result, there may be cases where an intrinsic duplicate does not think p at all, let alone know p. Suppose, for example, that the semantic value of 'thin' in my mouth constitutively depends on the dispositions of my linguistic community. Suppose at t I know that thin people are thin. Some possible duplicate of me does not think about that proposition at all, but instead, believes a proposition about thin+ things (where thin+ is the semantic value of 'thin' in his mouth). Even more obviously, while I may be reckoned to know a priori that if I am very happy then I am happy, a possible intrinsic duplicate who is numerically distinct from me will not be thinking *that* proposition, since it is about me, not the duplicate. Here is a more promising way of capturing the intuition of environment independence:

Independence$_1$: A case of x's believing p is a case of a priori knowledge iff for any possible intrinsic duplicate y,[7] the counterpart in y of x's belief that p is a case of knowledge.[8]

The notion of a counterpart is intuitive enough here. Crucially, a counterpart of my knowledge that thin people are thin might be knowledge of a different proposition. Further qualifications will be needed if one is the kind of semantic externalist who thinks that there are possible intrinsic duplicates of oneself who do not have any beliefs at all, or who thinks that there are at least some beliefs such that there is a possible intrinsic duplicate with no counterpart belief. Such a theorist might consider the following, alternative proposal:

that p, (c) if a process of the same type were to produce in X a belief that p then p (1980: 9–10). Note that 'a life sufficient for X for p' means ' X could have had that life and gained sufficient understanding to believe that p' (5–6). There is plenty to worry about here. It does not handle cases where experience provides knowledge destroying counterevidence (the experience includes experts who all say 'Not-P'). And it does not take stock of the fact that a life sufficient for the conclusion of a paradigmatically a priori proof might not be a life sufficient for one or more of the premisses. Standing back from these matters of detail, though, it is quite clear that Kitcher's basic idea is that a process warrants a belief a priori iff, no matter how the environment is, that process is a warrant provider. The considerations that follow will make trouble for that vision, no matter what the details of its implementation.

[6] We require here that the possible individual duplicate the entire intrinsic history of x.

[7] Some philosophers will think that my intrinsic nature radically undetermines the laws governing me. On such a view, some possible intrinsic duplicates might have radically different dispositions, and enjoy radically different causal connections between even the components of their own inner life. Such a philosopher had better restrict the thesis to intrinsic and nomic twins.

[8] Here I take intrinsic duplication to include both my physical and conscious life beneath the skin. Versions of the principle that merely include conscious life will be even more easily susceptible to troubling examples.

Independence$_2$: A case of x's believing p is a case of a priori knowledge iff for any possible intrinsic duplicate y such that the counterpart in y of x's belief that p is a case of believing that p, the counterpart in y of x's belief is a case of knowledge.[9]

(This is already a kind of retreat. The idea behind Independence$_1$ is that the environment makes no difference to *whether* an episode is a case of knowledge—it only makes a difference to *which* proposition one knows. Once we take seriously the possibility of a duplicate's having empty thoughts, the contribution of the environment has to be recognized as more substantial.)

Let us try to embed these ideas within a safety-theoretic conception of knowledge. In a posteriori knowledge, the environment plays the role of providing a *safe haven*. Take that part of an a posteriori knowledge-delivering process that is intrinsic to the subject. That intrinsic profile can be embedded in environments where the subject's counterpart belief is not knowledge at all. But according to the gloss encoded by Independence$_1$, a priori knowledge is distinctive in that it stands in no special need of a safe haven from the environment, since the environment can't but provide a safe haven: the intrinsic character of the process guarantees knowledge, whatever the vicissitudes of the environment. Meanwhile, Independence$_2$ is motivated by the following idea. In a case of a posteriori knowledge that P, the environment plays a dual role: it affords the connections to the world that allow the thinker to have the propositional object P as the object of belief in that case, and it provides a safe haven for that belief. In a case of a priori knowledge that P, by contrast, the first role is sufficient to ensure that certain inner lives are cases of knowledge.

How does the environment independence picture stand up? The trouble is that, with a bit of imagination, it is not hard to contrive ways that even paradigmatically a priori beliefs stand in need of a safe haven. Let me offer a few illustrative examples, the second and third of which turn on semantic externalism.

2.1. The Danger of Bad Influence

We are all familiar with fake barn cases.[10] A person sees a barn in an environment where there are lots of non-barns that look just like barns. Even though he

[9] It will not do to say 'counterpart who believes that p'. Suppose x believes P on good grounds and P′ on bad grounds. Owing to externalist considerations, there may be cases where a duplicate is such that the counterpart of x's belief that p is a belief that p′ and the counterpart of x's belief that p′ is a belief that p. In that case the duplicate would believe that p but not know p. But it doesn't seem that this should all by itself vitiate a claim of a priority for the actual belief that p.

I don't mean to suggest that the two proposals are the only reasonable candidates in the vicinity. One gets a conception intermediate between the two in the text by substituting 'a case of believing some proposition' for 'a case of believing that p' in the second conception. The considerations that follow will apply to this conception as well.

[10] See, notably, Goldman 1976. For discussion of the delicate nature of our intuitions in these cases, see Gendler and Hawthorne 2005.

forms a true perceptual belief of the barn that it is a barn, many of us are inclined not to count the belief as knowledge. The safety-based approach that I have sketched offers a natural diagnosis of that judgement. Supposing barns are essentially barns, the proposition believed is a necessary truth. Still the belief is not knowledge, since mistakes that are relevantly similar populate close worlds. The same judgement would no doubt be elicited if the real barn was in the vicinity of barn gas that induces the hallucination of barns. But the last case can easily be adapted to allegedly a priori beliefs. Suppose there exists a priori gas that induces the phenomenology of blatant obviousness for false propositions. Consider a person who believes a proposition not for any empirical reason but because the phenomenology of obviousness causes him to do so. Suppose the claim in question is that all bachelors are men. Consider a duplicate of that person who is embedded in an environment riddled with a priori gas. As a matter of luck he does not stumble into the gas. He in fact forms the belief that all bachelors are men. But he could very easily have stumbled into the gas and believed—due to felt obviousness—that all bachelors are women.[11] Insofar as one judges that the person does not know in fake barn cases, it is natural enough to judge that the person does not know in a priori gas cases. But this means that if we cling to the environment dependence idea, very few of our beliefs will count as a priori. Even so-called a priori beliefs need a safe haven.

The example turns on close worlds where one comes under a bad influence. Simple examples abound. The bad influence could be a priori gas. It could be a neuron-disturbing magnetic field or some belief-disturbing quantum entanglement. More mundanely, it could be a group of persuasive but sophistical intellectuals that could very easily have induced different habits and tastes in one's abstract belief formation (suppose that whether one believed that arithmetic was true or a useful fiction depended on which of a pair of adjacent school districts one is sent to.) Judgement in particular cases will not always be straightforward. Some will say that the protagonist is lucky enough to know. Others will say the danger of error is close enough to destroy knowledge in the actual case. But the situation is not crucially different from fake barns cases. There too intuitions will differ in various versions of the case but nonetheless, there will be a range of cases where most of us will be inclined to think that one is not safe enough to know.

2.2. The Danger of Empty Thoughts

Let us assume, with orthodoxy, that claims containing empty singular terms fail to express propositions. (I set to one side those free logics that validate identity claims even in the absence of a referent.) Suppose I accept the claim 'Muhammed Ali is Muhammed Ali' and thereby judge that Muhammed Ali is Muhammed Ali.

[11] I leave it an open question whether this person should be described as someone prone to spells of madness.

By the Independence$_1$ test, the claim is not a priori, since I have an intrinsic duplicate who does not know (or even think) anything at all by the counterpart episode. It is also easy to contrive an argument that this is not a case of a priori knowledge even by the Independence$_2$ test. For consider a world where I think the singular thought that Muhammed Ali is Muhammed Ali but where I could very easily have had an empty thought instead. In that case, my singular thought will not be safe (assuming that we should craft the notion of safety so that the risk of a failed attempt at belief is as damning as the risk of false belief).

Now this conclusion may be welcomed by some. Strictly speaking, I do not know a priori that Muhammed Ali is Muhammed Ali. All that I know a priori is that everything is identical with itself. But this conclusion can be welcomed only if it can be contained. For it is at least arguable that if the world is not compliant, very many of our *predicates* will be semantically empty as well. Certainly, the mere fact that an expression is a predicate does not guarantee its having a semantic value. (If 'Arabia' has no semantic value, then surely 'Arabian' will not either.) I lack the space here to explore whether the danger of emptiness is confined to a very limited class of predicates. Let me, for now, offer a conditional conclusion: If the possibility of emptiness is pervasive, then the environment dependence test will, if insisted upon, reckon pretty much all of our knowledge a posteriori.

2.3. The Social Nature of Meaning

One lesson of semantic externalism is that one's beliefs are constitutively dependent on the practices of one's community. This constitutive dependence (as I am conceiving it) is not an artefact of some deliberate, or even tacit, act of deference on the part of the thinker. Rather it is the inevitable consequence of the thinker's participation in a linguistic community.[12]

Once this lesson is absorbed, it is easy enough to make trouble for the environment independence idea. Consider my belief that, necessarily, all unmarried adult male humans are bachelors. Consider a series of cases, each pairwise slightly different, at one end of which is my actual belief, at the other end of which, owing to modal semantic drift induced by modal variance in community practice, the counterpart of my belief that necessarily all unmarried males are bachelors is a false belief. Let us stipulate further that there is no intrinsic variation among my counterparts across the possible situations. We could imagine, for example, that at the actual world there is very little tendency for the population to say that Adam (of the Adam and Eve story) is not a bachelor, but at the other end there is a very strong tendency among the population to make the counterpart of this speech (reasoning that 'bachelor' is not true of Adam since there is no institution

[12] Put picturesquely: a speaker simply lacks the power to lay down a law that his words and thoughts are not going to be constitutively determined by those around him. He simply hasn't got that kind of semantic authority.

of marriage).[13] At that world my inclinations with regard to the case are swamped by my community's in such a way that 'bachelor' in my mouth is not true of possible individuals like Adam. At that world, the counterpart of my belief is false and hence, obviously, not knowledge. If I assume the first conception of environment independence then I have to conclude that my actual belief is not a priori knowledge. Meanwhile, we can show that it is not a priori by the second conception by considering a world in the series of cases where I do believe the same as at the actual world but where I am situated so that very small changes in the community would have meant that I believed a false proposition. At that world my belief that, necessarily, all unmarried males are bachelors is not safe enough to count as knowledge.

Given a natural version of semantic externalism, there is a further related kind of threat, having to do with diachronic drift. If the contents of my thoughts are constitutively dependent upon community practice, then my thoughts can undergo semantic drift over time without my noticing, due to shifting practices in my community. If standards change without my noticing it, this can in turn be disruptive to my intuitions and reasoning in all sorts of ways. Sentences stored away in lexical memory might at one time express a true proposition and later express a false one. Predicates might shift semantic value during the diachronic process of some long chain of reasoning. And so on. It is in turn not hard to see various ways in which external semantic stability provides a safe haven for various kinds of reasoning that would typically be counted as a priori. Once again we are faced with a dilemma: abandon the environmental independence idea or else admit that hardly any of our knowledge is a priori.

3. EXPERIENCE INDEPENDENCE

We have been exploring the idea that a priority is constituted by the irrelevance of the environment to one's knowledge. The idea has not proven promising. All human knowledge—even knowledge that philosophers naturally classify as a priori—requires that the environment be so configured as to provide a safe haven for the knower. Perhaps, though, one can make something of the idea that knowledge is independent of experience in a way that does not challenge the material just presented. Think of some paradigmatically a priori proof. Grant that the environment plays the epistemological role of providing a safe haven for the knower. Still, it seems that experience of the environment is not an important part of the *process* whereby the knower achieves knowledge of the conclusion of the proof. And that might appear to constitute the basis of a natural division between certain kinds of human knowledge and others, perhaps the most eligible

[13] Thanks to Jose Benardete for this example. I note in passing that actual informants tend to vary in their judgments as to who is a bachelor in a community where the legal age for getting married is, say, forty.

joint in the vicinity of traditional discussions of a priority. Here is a recent and representative proposal in the vicinity:

> To say that something can be known without any justification from the character of the subject's experience is to say that there is a way of coming to know it which does not rely on any such justification. . . . On this approach, an a priori proposition is one such that there is a way of coming to know it under which the thinker's entitlement to accept the proposition does not involve the character of the experience. (Peacocke and Boghossian 2000: 1–2)

Some will find it natural to develop this idea by invoking the notion of *grounds*: A case of knowing is a priori iff experience does not form part of the grounds of the belief. I find the notion of ground a bit unclear, or at least philosophically tendentious. When I form a perceptual belief about the world, it is not obvious that I need do this on the basis of beliefs or knowledge *about* my experience. Relatedly, then, it is not clear that propositions about my experience form part of my evidence. What commitments are taken on board, exactly, when one says that experience is part of the grounds of belief in this case? Consider also a memory-based belief about the past configuration of the external world, one that is not accompanied by any present experience. What exactly does it mean to say that experience is the grounds of the belief?

We can bypass the concept of grounds by appealing instead to methods, where a method of belief-formation is a process that delivers or sustains a belief. Now some methods will be experience involving—they will be processes, one step or element of which will involve perceptual engagement with the outside world. Others will not. This provides us with a method-based conception of a priority.

> Experience Independence: A case of knowing is a priori if it is sustained by a method that is not experience involving.

On the face of it, this gives us pretty good results. Mathematical knowledge achieved by pure ratiocination counts as a priori, paradigmatically a posteriori knowledge does not. Laxer, derivative notions of a priority can be contrived to handle cases that do not strictly count as a priori by the above test. If I use pen and paper to do long division then, by the current test, my knowledge is not a priori. But the method I use does bear a close resemblance to methods that are not experience involving. In an extended sense—one according to which methods that sufficiently resemble those that are strictly a priori themselves count as a priori—the pen and paper case will count as a priori knowledge.[14]

[14] It is less helpful to say that the knowledge does not *essentially* depend on the pen and paper. If that just means that it is possible to know that proposition without pen and paper, that will not distinguish the pen and paper method from a paradigmatically a posteriori method. If it means that that very method could have occurred without the pen or paper, the thesis takes on tendentious commitments about the individuation of methods that can be bypassed by falling back on the true and less tendentious observation that a relevantly similar method could have occurred with the pen and paper.

It is worth noticing that if we drop the requirement of environment independence and tie a priority to experience independence, then we will thereby make room for the possibility of all sorts of knowledge that is a priori and contingent. Of course, we have been habituated to allow for certain cases of contingent a priori knowledge. If I believe the proposition expressed by 'Actually there is an odd number of red lego blocks iff there is an odd number of red lego blocks' on the basis of an appreciation of the semantic workings of 'Actually', then it is natural to count that as contingent a priori knowledge.[15] But with the requirement of environment dependence dropped, and the experience independent notion instituted, we will be encouraged to cast the net of contingent a priority much wider.[16] Suppose, for example, that someone is born with an innate mechanism that is a wonderfully reliable source of beliefs about a certain contingent subject matter—phonetics, syntax, the behavior of falling objects, or the psychological patterns of humans . . . On the safety-theoretic conception, such an individual will have not merely innate beliefs, but innate knowledge. And on the experience-dependent conception, such beliefs would count as a priori. That a duplicate might not have knowledge will be no grounds for complaint here—that would be to fall back on the discarded environment dependence idea. And that such an individual would have no *reasons* for his belief would be no grounds for complaint either—on the safety-theoretic conception, there is no obstacle to foundational knowledge that has no further propositional basis. However, I don't take any of this as a decisive reason to reject the experience-dependent strategy for discerning a natural a priori/a posteriori divide. The main obstacle to a workable conception of the a priori that is built upon the experience dependence idea has instead to do with the individuation of methods.

Consider a particular case of believing that p. There are all sorts of channels of cause and effect leading from the near and distant past to that belief. Which are to count as part of the method of forming the belief? Let us take a closer look at some of the pertinent decision points.

(i) It is often noted that even in paradigmatic a priori cases, empirical evidence can destroy knowledge. Even though I have carefully worked through a mathematical proof that p, I will not know p if I get empirical evidence that I am mad, or that human or mechanized experts have agreed that not-p, or that there is a priori gas in the area, or that I have made lots of mistakes using a very similar proof technique in the past, or that lots of smart people are inclined to laugh when they hear my proof. Were such experiences part of my history, then certain episodes in which I make a mistake would then (arguably) count as relevantly

[15] Of course, as the preceding discussion shows, even this may not count as a priori by the environment dependence test since (i) the relevant semantic workings may not be internally determined and (ii) in any case, one might need a safe haven in order to *know* the semantic workings, a fact that tends to be obscured by the use of verbs like 'appreciate'.

[16] For relevant discussion, see Hawthorne 2002.

similar,[17] destroying knowledge in the case at hand. Call knowledge-destroying experiences Bad experiences. Call the remainder Good experiences. My stream of experience is not irrelevant to my knowledge that p. That my proof counts as knowledge appears to depend crucially on its being accompanied by Good experiences. But if the process of arriving at putatively a priori knowledge is individuated so as to include Good experiences, then it will count as a posteriori by the experience-dependent criterion. (There is a general problem of *width* here. All sorts of facts at one time have some causal bearing on belief at a slightly later time. Which of them are to count as *part of* the belief-forming process?)[18]

A natural reaction to this concern is to try to make good on the intuitive distinction between cases in which the presence of a certain kind of experience is epistemologically important versus the cases in which the absence of a certain kind of experience is epistemologically important.[19] There are a variety of tricky questions in the vicinity here. Can omissions as well as positive events count as part of a process? If so, should the presence of an experiential omission in a process count as experience involving in the relevant sense? I shall not pursue these matters further here. For now, we should recognize that without some acceptable resolution of the difficulty, the experience-independent conception will exclude pretty much all of our putatively a priori beliefs.

(ii) Suppose a teacher instructs someone in the laws of nature when he is young and then he remembers those laws later. There is a store in the memory bank that calls forth the laws on demand, applies them to various possible cases, thereby extracting conditional predictions, derived generalizations, and so on. (We can think of two versions of the case. On one version, preservative memory of the laws is accompanied by memory of the teacher giving instruction in the laws. On another version, the laws are stored away without accompanying episodic memories.) Let us suppose that the person's nomic beliefs, conditional predictions, and so on are highly reliable. Just fill out the details of the case in such a way that at least many of the relevant beliefs are safe enough to count as knowledge. (I assume, here of course, that if the mechanism of belief production

[17] There is more to be said here concerning when evidence that one does not know destroys knowledge. I shall not pursue the topic here. For now, I shall make the simplifying assumption that, in general, powerful evidence that one does not know destroys knowledge. I am not, however, altogether averse to the idea that in a case where one competently proves a result and sticks to one's guns despite (uncontested) testimonial evidence from acknowledged experts to the contrary, one knows despite having evidence that one does not. (It might destroy knowledge that one knows in such a case without destroying the knowledge of the result.) That view would certainly help with the current challenge.

[18] This obviously relates to the generality problem for reliabilism that is widely discussed in the literature, though there the emphasis is often on the plurality of types that a single process instantiates, rather than the plurality of processes that deliver a belief. (There are various process types instantiated by a process, some more abstract than others, leaving out more or less detail. Construed abstractly, my opening a can with an electric opener will be the same kind of process as my doing so with a manual opener.)

[19] Relatedly: Bad experiences lower evidential probability but Good experiences don't raise it.

is safe enough, then the believer will know its deliverances even if he cannot produce any empirical reasons whatsoever for the laws.)

Recall Frege's test for a priority in the *Grundlagen*: 'If on the contrary, its proof can be derived exclusively from general laws, which themselves neither need nor admit proof, then the truth is a priori' (Frege 1980: 4). If we apply this to the case at hand we will get an affirmative verdict on a priority. Suppose, for example, that the laws are retrieved from memory and applied to deliver a particular conditional prediction. It does not seem that belief in the laws *needs* proof to count as knowledge, at least from the safety-theoretic perspective. Also, it does not seem the laws *admit of* proof from more banal empirical premisses. Now of course it would have been utterly anathema to Frege to count the relevant case as one of knowing a priori. And that is because it never occurred to him that one could have foundational knowledge of contingent propositions by virtue of safe mechanism rather than propositional evidence or experience. All of this merely underscores the need to carefully rethink traditional discussions of the a priori in the light of the safety-based picture.

Let us think about the law of nature case afresh. Should we count the believer as knowing the relevant propositions a priori? One will naturally react—on either version of the case—by insisting that the laws were not known a priori. And one might naturally invoke something like the experience-dependent conception in justifying such a claim: the process that led to the fixation of belief included experiential exposure to the teacher. The knowledge is a posteriori knowledge, achieved via testimony, not a priori knowledge.

It is important to realize here that when one considers a stream of events, various processes can be distinguished, some longer in temporal extent than others. Suppose a cake is baked. There is a process that begins with shopping for the ingredients and culminates in the taking the cake out of the oven. But there is a shorter process that begins with putting the unbaked cake in the oven and ends with taking it out. So with judgement. Suppose someone extracts a conditional prediction about the course of events. There is a process that begins with the teacher telling him the laws and ends with applying some laws to derive a conditional prediction. But there is a shorter process that begins with retrieving the laws from the relevant internal information bank and ends with producing the conditional prediction. One of the processes is experience dependent. One is not. Which shall we use to test whether the belief is a priori? Let us call the process beginning with the retrieval Short and the process beginning with the interaction with the teacher Long. Is there any deep mistake in taking Short to be the relevant safe method? (We earlier encountered the problem of width; we are now confronted with the problem of *length*.)

It no use complaining that Short would not be safe were it embedded in a world where the laws were different. That is true of Long as well and, given the failure of environment independence, is in any case not a good test for a priority.

One might argue that it is very artificial to treat Short as the relevant process, since a process like Short brings knowledge with it only in the presence of Long. The form of argument is not immediately persuasive. But in any case, its premiss is faulty. Someone who was born with an innate store of the laws of nature would exemplify something like Short in contriving conditional predictions, but not Long. And as we have seen, with environment independence discarded, there is no good reason to disallow some such cases as knowledge. Even in the version of the case where retrieval is accompanied by episodic memory of teaching, it is not clear that Short can deliver knowledge only in the presence of Long. Consider the following thought experiment: Suppose someone is born with an innate storehouse of true mathematical beliefs and reliable computational techniques. Those beliefs are accompanied by pseudo-memories of being taught mathematics in an earlier life. Those pseudo-memories trick the believer into thinking that Plato's doctrine of recollection—or something like it—is true. It seems to me that we may well be inclined to count such a person as knowing mathematics, despite the misleading pseudo-memories. Suppose now someone is born with a storehouse of beliefs about the laws of nature—accompanied with relevant pseudo-memories of nomic training from a sage. Here we would have Short—in the version from the episodic memory case—without Long. Once again, it is far from clear that the person fails to know.

(iii) Consider a case of mathematical knowledge. Here again, there are a variety of candidate processes to use as a test for experience independence. Let us focus on two: (a) Math-Short, which begins with the retrieval of mathematical information from preservative memory and ends with the application of computational techniques to answer the problem at hand. (b) Math-Long, which begins with the training in the relevant techniques and provision of the relevant mathematical information at school, home, or college, and ends with the application of computational techniques to answer the problem at hand. Using Math-Short as our benchmark, the relevant belief is experience independent. Using Math-Long as our benchmark, it is not. Now what is it that justifies using Long in the law of nature case but Math-Short in the mathematical case? A certain kind of answer will be natural to many philosophers at this point. They will insist that the training in the mathematical case 'doesn't count' because that training merely played the role of allowing the person in question to grasp the relevant propositions/acquire the relevant concepts. In this connection they will likely invoke some standard platitude about the distinction between conditions of acquisition and conditions of entitlement. Here is a recent affirmation:

> In the case of a priori propositions, much experience, perhaps of a specific character, may be required to grasp the concepts implicated in the proposition or to access the entitlement to believe it; but conditions of grasp and of access remain distinct from the nature of entitlement. (Peacocke and Boghossian 2000: 2)

Various concerns might be raised about this reaction to the question under consideration. First, the original training may have gone well beyond that of allowing the person to understand the relevant propositions. If a mathematical belief is reckoned a posteriori in cases where the training goes beyond introduction to the concepts to provide testimony about the objects of the concepts, many of the beliefs in question will turn out a posteriori. Suppose, for example, the training includes various sophisticated proof techniques that are quite obviously not preconditions of understanding anything in the domain. Moreover, once we take the social character of meaning seriously, it is natural to think of the conditions of meaningful participation in the discourse as quite minimal. Someone who was unsure whether '1–2' had an answer or was undefined counts as meaning what everyone else does by '1–2' on account of social facts. (Meanwhile, any suggestion of some notion of 'real understanding' according to which such a person means but does not *really grasp* the proposition in question posits a natural kind of real understanding that may very well be chimerical.) The more one thinks in this social externalist way, the more one has to recognize the training in question as going beyond providing the conditions for understanding. One should also wonder whether belief in any of the basic propositions of arithmetic is really a condition for understanding. Would a community of naysayers that were trained to regard arithmetic as false (on the grounds that Hartry Field is right about numbers)[20] not count as grasping arithmetical propositions? The attempt to discount the training on the grounds that it merely introduces the person to the relevant concepts is not compelling. And of course it is no good to fall back on the thought that once the training is over one does not need to remember the experience of the training in order to know the math (since one now has what it takes to know the math just by relying on inner resources). For of course the same could be said in the law of nature case.

(iv) As Timothy Williamson has emphasized in some recent work, much of what actually goes on when people describe themselves as having intuitions about a subject matter is that they form judgements about contingent counterfactual propositions.[21] Let me briefly summarize a few of the points that he has pressed using a vivid example. Suppose I reflect and judge that were someone to have less than two hairs he would be pretty bald. Now there are worlds where the counterfactual is false: If a person had a single hair that was fifty feet long and wrapped into a bun, he would not be pretty bald, nor would a person with a single hair that was eight inches thick. Now of course we might think that associated with the contingently true counterfactual is some necessarily true proposition. But it is a delicate and challenging task to try to 'clean up' the counterfactual so as to turn it into a necessary truth. What we do in general is make do with the contingent verdict.

[20] See Field 1980. [21] See Williamson 2005.

Consider someone who makes various contingent counterfactual judgements about bald people, pretty bald people, and very bald people. There is no actual perceptual engagement with the world during the course of the judgement. He merely reflects, as it were, and produces the counterfactual judgement. This pervasive phenomenon forms an excellent kind of test case for the current discussion. If we adopt the experience-dependent test of a priority, our verdict will once again turn on the individuation of methods. To simplify, let us suppose the person has a little module in his head that is responsible for the baldness verdicts, one that has been crafted by exposure to the world, but which is now eminently reliably with regard to such verdicts, and can, indeed, be relied upon without resorting to further perceptual consultations with the environment. We can individuate the belief-forming process in a short-term way—as involving deployment of this inner module. So construed the belief will be counted as a priori. Or we can individuate it in such a way that the past experiential engagements that were formative of the module were part of the process. So construed we will count the belief as a posteriori. Do we then think there is is some deep fact about which process is constitutive of the knowledge? Either way the person counts as safe. So we cannot ground a choice of method on the question whether the person knows the relevant counterfactual. Moreover, it hardly seems that ordinary folk ever think about such questions, even tacitly. It doesn't seem, then, that we then can ground a preference between the short and long versions of the process by appealing to folk practice. Why then be confident at all that there is some natural joint that is being tracked by philosophers' talk about 'relying on experience', one that can settle questions of a priority in a (far from unusual) case like this one?[22]

4. JUSTIFICATION AND THE INTELLECT

Some will react to the discussion so far by claiming that the notion of a priori justification is primary. On one natural version of this view, my discussion has not touched on one of the deepest epistemological facts in the vicinity, namely that intellectual intuition is a source of justification. The picture I have in mind reckons there to be some deep analogy between the way that perceptual experience offers justification for perceptual beliefs and the way that intellectual intuition—'intellectual seemings' as George Bealer puts it[23]—justifies various mathematical, logical, and metaphysical beliefs.

[22] Williamson's own perspective is fully in accord with the line taken here: 'Both crude rationalists and crude empiricists in effect assume that evaluations of counterfactuals can somehow be divided according to their sensitivity or otherwise to experience into two drastically contrasted classes. That assumption is neither intrinsically plausible nor adequately supported by evidence' (2005: 18).

[23] See, for example, Bealer 1996.

Begin with the case of perceptual experience. One natural idea is that it is an intrinsic fact about perceptual experience that it lends support to beliefs in the propositions represented by that experience. On such a view, this justificatory role isn't a contingent fact about the experience; it is of the essence of a perceptual experience that represents p that it lends support to a belief that p. Of course if there is other countervailing evidence, the belief that p may, all things considered, not be justified. But, ceteris paribus, the experience justifies the belief. Put another way, the experience offers prima facie justification for the belief. If we use the language of evidential probability to encode how well supported a belief is, we can articulate one consequence of this picture: of necessity, a perceptual experience that p raises the evidential probability that p unless the evidential probability that p is already 1 or unless the experience is accompanied by countervailing evidence.[24] Now on the picture that I am currently considering, much of the same structure carries over to intellectual seemings: by their very nature, they offer prima facie justification for beliefs and, coordinately, tend to be evidential probability raisers.

Some might see some of the earlier examples through the lens of this kind of picture. Consider a priori gas. One might argue that the risk of exposure defeats knowledge but not the justification delivered by intellectual seemings. (One might go further, arguing that even actual exposure does not destroy justification unless there is evidence that exposure is going on.)[25] This kind of reaction is most naturally motivated by the idea that a priori knowledge decomposes into an 'internalist' component that is accessible to the subject—some kind of intellectual seeming, and which forms the bedrock of justification, and an 'externalist component' that includes various reliability conditions.

For clarity's sake, let me address the idea in its most brutish form. (Perhaps there are more nuanced, more compelling versions of the idea. I leave it to others to contrive them.) Suppose there is a propositional attitude—intellectually seeming that p (henceforth Seeming) that manifests itself to inner consciousness by a special kind of intellectual phenomenology (call it Glow). (Let me be neutral on whether Seeming brings Glow with it necessarily or contingently.) The proponent of the view then claims that Seeming is by its nature an evidential probability raiser, at least ceteris paribus, and that we are clued into

[24] The picture says more, of course: that in the absence of counterevidence experience raises the evidential probability to a level such that the belief then counts as 'justified'; but I don't wish to focus on that aspect of the view—and its difficulties—here.

[25] One might instead think while intellectual seeming always bring prima facie justification, justification is defeated by the mere presence of a priori gas. (In his 1998, Bonjour argues for a view of this kind when discussing the justificatory role of apparent rational insight when it occurs in a being whose judgement is 'irreparably clouded' (137).) For my part, I find it quite hard to keep track of the terms of art that figure in such discussions. Rather than ponder directly whether the justification in such cases is merely 'prima facie', it seems better to ask what the range of the subject's evidence is in such cases, and in particular whether intellectual seeming still counts as evidence for the proposition presented.

Seemings that occur in ourselves by Glows.[26] And on the package that I am envisaging, he claims that such Seemings are at least a core component of a priori justification.[27]

Here are some salient pressure points (ones that will have already occurred to many readers).

First, what should one say about possible cases (if there are such) where Seeming that P occurs without Glow? Is the evidential probability of P raised in that case, even when the intellectual act is not manifest from within? The picture we are discussing is, quite evidently, driven by an internalist mandate to allow only that which is 'accessible' from the inside to count as relevant to justification. To allow Seeming to be a probability raiser in this kind of case is to fail on the mandate, which in turn will call into question the intellectual underpinnings of the picture in the first place. Suppose instead one insists that Seeming can justify only when accompanied by Glow (those who posit a mysterious necessary connection will get this result for free). The shape of the worry does not go away. For consider cases in which one has Glow but is not in a position to know one has Glow.[28] Is Seeming plus Glow still an evidential probability raiser in that case? Why should unnoticed facts about one's life beneath the skin be able to raise evidential probabilities if unnoticed facts about the outside world cannot?

The proponent of the view has a natural retreat: Known Seemings are probability raisers. When one knows that it Seems that p, this tends to raise the probability that p. Henceforth I will assume this version of the view.

Second, it is worth pressing on the modal commitments of the view that known Seemings are, by their very nature, probability raisers. Consider a race of creatures of the following sort. They are so constituted that Seemings come easily. All sorts of wild and wonderful propositions Seem to be the case, and have an accompanying Glow. Ancestors that believed what Seemed to be so didn't last long. Evolution selected for species members that took a more questioning attitude to their Seemings and Glows. (We might, if we wish, imagine that there

[26] Of course if we wish to develop a serious theory of evidential probability for mathematical knowledge, we shall have to drop the idealization of logical omniscience common to standard decision theories, as well as (relatedly) the picture of sets of worlds as the objects of knowledge that is also orthodoxy within decision theory.

[27] Some versions of the idea proceed in terms of a more inflated ideology. Laurence Bonjour posits a relation of rational insight wherein one is 'able to simply see or grasp or apprehend that the proposition is necessary' (1998: 106). If 'see that' and 'apprehend that' merely mean 'know that' then rational insight could not play the desired role of explaining a priori knowledge. Bonjour presumably intends it to be understood as a sui generis fundamental, quasi-perceptual, relation holding between a subject and necessary propositions. Supposing there were a fundamental relation of this sort, it could certainly be used to identify a kind of knowledge. But I doubt very much that there is. (I note in passing that Bonjour assumes that a proposition must be necessary in order to be known a priori. I find his condensed discussion and rejection of the contingent a priori in 1998: 12–13 altogether unpersuasive, but here is not the place to pursue the matter.)

[28] For relevant discussion see Williamson 2000: ch. 4.

was a special kind of phenomenology—Superglow (perhaps Glow accompanied by an invigorating tingle)—that was in fact reliable and that the creatures did take Seemings accompanied by Superglow at face value. Imagine some such creature is born into the world, and has his very first Glow. He has a Seeming that, say, $5 + 12 = 18$, accompanied by Glow, and takes little doxastic notice of it. (We might imagine that, instinctively, he has the habit of responding to a Seeming that P by filing the question whether P as one to be enquired about later. Perhaps Glows are good indications of questions worth asking for these creatures, even if they are not good indications of the answer.) It's not that the creature knows that Glows are unreliable—he hasn't achieved that kind of self-awareness quite yet. It's just that the creature is so constituted as to be generally unaffected doxastically by Glows. The picture that I have been considering maintains that despite all this Known Glow plus Seeming raises the evidential probability that p, for some p, when the creature Glows for the first time. I do not find this position especially plausible.

Third, one wonders why my current Seemings are given such a privileged role vis-à-vis a priori justification. Let me note two issues in this connection. First, is there some special epistemic status to my current, here and now Seemings? Suppose I remember a Seeming that p from the past but do not have one at present? Won't that do just as well? Or suppose I become aware that you have a Seeming? Second, and more importantly, what should the proponent of the view think about, say, mathematical knowledge that is unaccompanied by Seemings? After all, it seems hopelessly naive to suppose that all bona fide mathematical knowledge has to be preceded by the intellectual act of Seeming or the phenomenological fanfare of Glow. I suppose that the proponent of the package I have in mind would think that if there was no evidence from perception (testimony) or memory (memories of testimony or earlier Glows), then these would be cases of knowledge without justification. But this only serves to highlight the burden placed on the notion of 'accessible' in standard internalist ideology. As we noted above, Seemings make their epistemological presence felt by being known. Once known, facts about Seemings justify other beliefs, including ones about the necessary structure of reality (at least according to the picture we are entertaining). Yet parity of reasoning ought to tell us that mathematical facts make their epistemological presence felt by being known, and that they can then in turn justify other beliefs as well. But once this is conceded, it becomes unclear what *special* role Seemings (as opposed to other known facts) play with regard to justification. If 'accessible' means 'known', the internalist restriction to the 'accessible' as delimiting 'what was has to go on' will not in any way force one to typical internalist conclusions as to range of justificatory resources. But if 'accessible' means something else, I do not know what it is.[29]

[29] For relevant discussion of this point, see Williamson forthcoming.

5. CONCLUSION

Shifty vocabulary that fails to mark any natural joints may yet prove to be an informative vehicle in context. Suppose there is no natural divide among human beings between the sinners and the saints. Still, in context, someone can convey useful information by describing someone in one of these ways. Similarly, when a philosopher tells me 'I figured it out a priori', that will often provide me with information. In context there will be a range of processes that, for all I know, constitute the etiology of his belief. Some will be comparatively more heavily or more recently experience reliant than others. From his speech, I will be able to rule out certain of the processes. Of course, none of this shows that the category of a priority will prove a fertile one from the point of view of serious epistemological theorizing, or that it marks a natural joint in our epistemological lives. Indeed, its prospects on that score seem rather dim.

REFERENCES

Bealer, G. 1996: 'A Priori Knowledge and the Scope of Philosophy.' *Philosophical Studies:* 121–42.

Bonjour, L. 1998: *In Defence of Pure Reason* (Cambridge: Cambridge University Press).

Field, H. 1980: *Science without Numbers* (Oxford: Blackwell).

Frege, G. 1980: *The Foundations of Arithmetic* (2nd rev. edn. Evanston, Ill.: Northwestern Press).

Gendler, T. and Hawthorne, J. 2005: 'The Real Guide to Fake Barns: A Catalogue of Gifts for your Epistemic Enemies.' *Philosophical Studies* June: 331–52.

Goldman, A. 1976: 'Discrimination and Perceptual Knowledge.' *Journal of Philosophy* Nov.: 771–9.

Hawthorne, J. 2002: 'Deeply Contingent A Priori Knowledge.' *Philosophy and Phenomenological Research* 65: 247–65.

Kitcher, P. 1980: 'A Priori Knowledge.' *Philosophical Review* 89.1: 3–23.

Peacocke, C. and Boghossian, P. 2000: 'Introduction.' *To New Essays on the A Priori* (Oxford: Clarendon Press).

Sawyer, S. 2001: 'The Epistemic Divide.' *Southern Journal of Philosophy:* 385–401.

Sosa, E. 1999: 'How is Knowledge Modally Related to What is Known?' *Philosophical Topics* 26: 373–84.

Williamson, T. 2000: *Knowledge and its Limits* (Oxford: Oxford University Press).

____ 2005: 'Armchair Philosophy, Metaphysical Modality and Counterfactual Thinking.' *Proceedings of the Aristotelian Society:* 1–23.

____ forthcoming: 'On Being Justified in one's Head.' Festschrift for Robert Audi.

10

The Inference that Leaves Something to Chance

David Sosa

> Imagine someone saying:
> 'But I know how tall I am!'
> and laying his hand on top of his head to prove it.
>
> (Wittgenstein 1953: entry 279)

Here are two theses about inference:

1. *Ignorance is insufficient for incoherence*: inferring subjects are in principle in a position to avoid invalidity, no matter what their state of knowledge (indeed, no matter what the truth of their premiss beliefs).
2. *Internalism about inference*: if S_1 and S_2 are internal duplicates during a period of time, then S_1 draws an inference during that time if and only if S_2 draws an inference then.

Externalism about mental content is incompatible with the conjunction of these two theses.[1] Because I think that, suitably interpreted, both theses are true, I take externalism to be challenged by this incompatibility. But my purpose here is primarily to exhibit the nature of the challenge posed by the theses, not to adjudicate the results.

1. EXTERNALISM AND PRIVILEGE

Notoriously, Externalism has been alleged to be incompatible with plausible views about the nature of self-knowledge (its privilege or immediacy or directness, for

[1] Externalism about the content of linguistic utterances is an independent view, compatible with the theses bruited here.

Incidentally, differences between externalisms on which causal relations to physical kinds are central and those on which the relation of *membership in a linguistic community* is central will not be at issue below.

example).[2] I think there is no question that externalism should *not* be thought incompatible with privileged self-knowledge as such. To see this, we need consider only the case of what we might call 'indexical knowledge.'

We should all be externalist about such things as spatio-temporal location: an individual's location is not determined by her intrinsic physical properties. Two molecule-for-molecule duplicates might be differently located. Spatio-temporal location is an external property. It does not supervene on the internal—so much should be uncontroversial.

Is this externalism incompatible with immediate, direct, or privileged access to our spatio-temporal location? If our internal state is insufficient to determine our location, then it would seem no amount of looking within will suffice to disclose it. We have to perform empirical investigation in order to discover our location. That's what follows from the externalism of spatio-temporality, or so it just might seem.

But appearances are deceiving, in a way. Plausibly, there need be no conflict between externalism about spatio-temporal location and privileged access thereto. We *can*, it turns out, know our spatio-temporal location in a privileged way without empirical investigation. It is a familiar point that no matter where you are, if you think (as it were) *I am here now*, then what you think will be true. We can always locate ourselves spatio-temporally. If you were elsewhere and thought *I am here now* you would be placing yourself in that other location. It's not that you always place yourself in the same location—*here and now* (so that you have to keep very still in order to avoid going wrong!). There's an infallible way to keep track of your location, without need for empirical investigation. All this in spite of the fact that spatio-temporal location is something duplicates could fail to share.

There is, nevertheless, something unsatisfactory—inadequate—about the sort of knowledge we can have, without empirical investigation, of our spatio-temporal location. In a minimal sense, for any given time and place, there is a way one can know one's location just by believing that one is there then. But some privileged self-knowledge is in any case very different from this. Spatio-temporal location is not supposed to be available to us in the way that our own thoughts must be. That any immediate and authoritative knowledge of spatio-temporal location will be of what we might call an 'essentially-indexical' variety is not philosophically troubling.

An analogous result in the case of self-knowledge is unsatisfactory. Immediate, authoritative, *and substantial* self-knowledge is a deep presupposition of many of our most significant practices. Without it, the legitimacy of normative evaluation, for example, would be called into question. In particular, I will argue that externalism undermines a sort of privileged self-knowledge that is a presupposition of our view of inferential status.

[2] See e.g. McKinsey (1991). I will not be concerned here with relations between privilege, immediacy, and directness.

So the intuition about the incompatibility of externalism and self-knowledge has to be relocated and rearticulated. It's not that externalism is incompatible with self-knowledge: externalism is incompatible with a specific sort of self-knowledge that we do or must enjoy. This latter point can itself be developed in several ways. I will argue that externalism is compatible only with varieties of self-knowledge that cannot sustain jointly the two highlighted theses: *internalism about inference* and *ignorance is insufficient for incoherence*.

2. BOGHOSSIAN ON EXTERNALISM AND INFERENCE

Before turning to inference, Boghossian argues that Burge's position on the compatibility of externalism and privileged self-knowedge is inadequate to plausible considerations about memory. According to Boghossian, we think of it as 'central to our capacity for self-knowledge' to be able to know 'what one has thought immediately after one has thought it' (170). And the mechanism to which Burge appeals in accounting for the compatibility of self-knowledge and externalism is said not to account for this phenomenon. Boghossian thinks Burge has not given us even '*one* case in which a thought is known directly despite the relational nature of its individuation conditions' (171).

The argument involves an example of slow switching and its support of an inconsistent triad that can be summarized[3] as follows:

i. S knows at t that S is thinking at t that p.
ii. If at $t + n$ S remembers everything S knew at t, then S knows at $t + n$ that S was thinking at t that p.
iii. At $t + n$, even if S has forgotten nothing, S does not know that S was thinking at t that p.

I think the indexical case is again instructive. Suppose that at t S knows where he is by thinking: *I am here now*. By the time of $t + n$, S has been surreptitiously moved. He may in an important sense no longer know where he was at t. He might, for example, think that he was where he'd now be placing himself if he were *now* to think: *I am here now*. Has he *forgotten* where he was? We 'ought to be able to exclude memory failure by stipulation' (172) says Boghossian. Perhaps; but we should take care about what that exclusion entails. How could he think that he'd be placing himself in the same location again, when in fact he would *not* be (since he earlier placed himself *there* and would now place himself *here*, a different place), unless he never really knew where he was in the first place?

So imagine the following analog of Boghossian's inconsistent triad:

(a) S knows at t that S is in location l at t.

[3] Cf. Ludlow's (1995) summary.

(b) If at $t + n$ S remembers everything S knew at t, then S knows at $t + n$ that S was in location l at t.

(c) At $t + n$, even if S has forgotten nothing, S does not know that S was in location l at t.

The above considerations force the admission that if at $t + n$ S is to know that he was in location l at t (which we are to 'stipulate'), that knowledge will not be constituted by his thinking, at $t + n$: *I am here now*. But this consideration suggests rather a line of *resistance* to Boghossian here: what he has perhaps made plausible is that at the later time, after the switch, S will not know *directly* (in the way he did at t) what he was thinking at t—S will not know directly (or immediately, or in a privileged way) that he was in location l at t.[4] But it does not follow that S will not or cannot know what he was thinking or where he was located at t—the examples make plausible only a weaker version of (c) here: At $t + n$, even if S has forgotten nothing, *S* does not know directly that *S* was in location l at t. That weaker (c) is not incompatible with the conjunction of (a) and (b).

For the inability, any longer, to know *directly* what you were thinking, or where you were, to be a problem, it has to be plausible that if at a certain time you know directly that p, then if at a later time you've not forgotten, then at that later time you still know directly that you were thinking at the earlier time that p: we would need a stronger (b). But whatever we may think about Boghossian's platitude—(ii) above—it is a much more committed position to claim that the directness of one's knowledge is maintained under the operation of memory. It is not implausible to suppose rather that although memory will preserve the knowledge we once had, it will not preserve its directness. The idea, accordingly, will be that although S knows directly at t what he is thinking then, and although at $t + n$ S remembers everything he knew (and thus, at least, continues to know what he knew at t) still that knowledge may no longer be direct. The lesson of the case then—let's call it the 'case of premature memory loss'—is that so long as there is a way for the relevant knowledge to be preserved, even if there is alteration in the nature of the state or of our access to it, the difficulty alleged for externalism can be deflected.

Setting aside now for the moment the instructive issue of the compatibility of externalism and memory, Boghossian elsewhere raises an issue about the compatibility of externalism and *inference*; he urges that externalism will allow for

(B) A situation in which two tokens of Mentalese belonging to the *same* syntactic type have *different* meanings, but the subject is in principle not in a position to notice that they do.

[4] It is not immediately plausible that duplicates who both know something should be such that one does so *directly* if and only if the other does. You and I could both know that I am here, but even if we were duplicates, only I could know it in the special way I would if I were to think *I am here*.

And he uses the existence of such cases to argue that externalism is inconsistent with the *a priority* of our logical abilities.[5]

Schiffer responds with an example in which one reasons as follows:

Pavarotti once swam in Lake Taupo
Whoever once swam in Lake Taupo got wet.
So, Pavarotti got wet.

The intuitive thing to say about this type of example, according to Schiffer, 'in which one token gives rise to another of the same type wholly by an inferential process and with no further operative external causes of the second token' (1992: 35), is that the second token has the same meaning as the first. The dialectical situation is subtle here: recall that Boghossian is, with such examples, intending to object to externalism. Boghossian might agree with Schiffer that the claims in the example concerning to whom the various uses of 'Pavarotti' refer are not 'intuitive.' But Boghossian's argument turns on *externalism's apparent commitment to* so characterizing those uses. He need not accept Schiffer's limiting interpretative gloss according to which, in the example, 'one token gives rise to another of the same type wholly by an inferential process and with no further operative external causes.' It is not generally the case that in an inference, the first premiss belief itself causes the second. What is plausible is only that the two premiss beliefs should, in an inference, cause the conclusion.[6]

Let's turn again to an indexical case for clarity:[7]

He once swam in Lake Taupo.
Whoever once swam in Lake Taupo got wet.
So, he got wet.

Externalism about the content of uses of indexicals holds that it is insufficient, for the second token of 'he' to have the same content as the first, that the two uses involve the same 'descriptive content,' or 'linguistic meaning,' or any other internalistically individuated feature. Whether a pair of tokens of a type like 'He' have the same content is determined in part by the (external, physical) context of use (who else is around, for example). But then it is not so clear that, *given externalism*, the second token of 'he' has the same content as the first. If the context has changed (especially, without one's noticing), then the token's content

[5] Some have argued (Falvey and Owens 1994; Brown 2004) that internalism faces the same problem—that the issue does not *differentially* affect externalism. Although the relevant argument proceeds from unacceptable (and unacceptable independently of internalism) presuppositions (*inter alia*, that proficient users of a language cannot fail to know the meanings of terms in that language), considerations of space recommend discussing it elsewhere.

[6] I should emphasize that the relevant tokens are not those of 'Pavarotti': we needn't object to a demand that the second tokening of that type—in the conclusion—be caused (in part) by the first—in the first premiss. The relevant issue is whether the two occurrences of 'Lake Taupo' need be linked in the alleged way. I'm grateful to Tim Pickavance for discussion here.

[7] See Sorensen 1998 for related discussion.

may have changed without one's noticing. And so the second token might have a new content. The appropriate demand that the premiss concept be caused to reappear in the conclusion is put at risk by externalism.

3. BURGE ON EXTERNALISM AND INFERENCE

The analogy with the indexical case raises the issue of whether we are already subject to the phenomenon to which Boghossian wants to point, independently of externalism about, for example, natural kind terms. In fact I think it is clear that indexical thoughts have a kind of content determined independently of external physical environment.[8] Burge himself holds, indeed, that unlike in the indexical cases, it is not in general 'possible to separate rather neatly what aspects of propositional attitudes depend on the person holding the attitudes and what aspects derive from matters external' (1982: 98). The phenomena grounding externalism are said not to be indexical: on the contrary, they 'involve the constant context-free interpretation of the terms' (107). Whatever accommodation we find for the indexical case should thus *not* be assumed to carry over unproblematically to the externalism developed by Burge.[9] If indeed we are already subject to the phenomenon Boghossian points to, some kinds of externalism will be subject to it in a distinctive form.

Responding to Boghossian, Burge claims to bring out 'a generalization of Schiffer's points . . . in actual reasoning, we *typically* tie key terms in premises together through preservative memory' (1998: 366 n. 16, emphasis added); and though he is 'doubtful that there are any clear cases of invalid equivocation deriving from switching cases,' Burge holds that if there are any, 'they are *marginal*' (368, emphasis added). I find Burge's approach here troubling: one might have thought that whether such cases are typical or marginal—how often they occur, that kind of thing—is not for philosophers to judge. Our question is whether such cases are possible: does externalism distinctively and counterintuitively allow that an agent could draw an invalid inference without being in a position, in principle, to detect that invalidity *a priori*? We will consider also whether, even

[8] See Owens 2003 for a dissenting view.

[9] This point could be developed: As, in effect, observed earlier, McKinsey's argument, alleged to demonstrate the incompatibility of externalism and self-knowledge, is smoothly accommodated on externalisms that do not aspire (as, for example, Putnam originally did not) to more than what is true about indexicals.

The accommodation does require, however, resolution of the following McKinsey-style apparent problem: Possibly, there is a subject that (i) knows in a privileged way that she is at spatio-temporal location *l*, (ii) knows *a priori* that if she is at spatio-temporal location *l*, then there is an external world, but (iii) cannot know in any privileged way that there is an external world. I suspect the apparent need for an 'egocentric' space is related to an issue about externalist conceptions of *internality*, to be mentioned briefly in closing below. So when all is said and done, I myself do not believe the accommodation is perfectly smooth.

if externalism does not have that consequence, it will entail that duplicates can differ with respect to whether they are drawing an inference.

Two distinctions are relevant: first, 'Disjoint Type' cases versus 'Amalgam Type' cases; and second, 'substantive memory' versus 'preservative memory.' In disjoint type cases, the subject, after a slow switch, simply takes on a new concept, one that does not apply to the stuff for which he had the corresponding concept before the switch. In amalgam type cases, the subject's concept broadens to refer also to the new stuff. So long as disjoint type cases are possible, the externalist is properly challenged on that basis.

And by contrast to 'substantive' memory, which works by discrimination, 'preservative' memory is said normally to retain 'the content and attitude of commitments of earlier thinkings, through causal connections to the past thinkings' (Burge 1998: 357). Such a memory is not '*about* a past event or content at all. It can simply link the past thought to the present, by preserving it' (357).[10] Despite a 'broad but qualified analogy between preservative memory and pronominal back-reference,' and although preservative memory operates through 'causal chains' which connect later thoughts to earlier ones, there is an important difference: 'Anaphora is a syntactic device, whose semantic interpretations may vary' while '[p]reservative memory is not primarily a syntactic matter. . . . In the memory case, the content and referent of the remembered material is not distinct from that of the antecedent thought content' (358–9). This, in effect, is how we resolved Boghossian's argument about memory earlier. Even after a slow switch, subjects might maintain concepts acquired earlier, and continue to deploy them in beliefs maintained (as knowledge) in memory. Accordingly, on analogy with the unacceptability of the indexical version (c), Boghossian's (iii) was unacceptable.

But we should '[c]onsider the role of preservative memory in deductive reasoning,' (363) which might 'undermine another of Boghossian's criticisms of anti-individualism' (364). Accordingly, consider Burge's discussion of Alice, who remembers 'the event of picking up, and feeling the light weight of, some aluminum on earth, before she was switched' and then, 'remembering a sample of twaluminum on twin earth that she saw yesterday, [] might think that yesterday there was some twaluminum beside her. . . . She might reason from these premises, fallaciously, to the conclusion that she once picked up the same sort of thing that was beside her yesterday' (366).

Does the attempt to use this sort of example in objection to externalism overlook the centrality of preservative memory in reasoning? Although he finds it 'natural' to interpret Alice's first premiss as the belief that she once picked up some aluminum, Burge insists that

> [i]nsofar as we think intuitively that Alice is not making a mistake in reasoning, it is natural, and in most cases [] correct, to take her to be holding constant, through

[10] See Falvey 2003 for a challenge to Burge's conception of one's entitlement to preserved memories.

preservative memory within the argument, the concept used in the first premise in her thinking in the second premise. . . . Although free-standing memories normally evoke the concepts utilized in or appropriate to the remembered context, the exigencies of reasoning will often take precedence. (367)

Again, I find Burge's various hedges ('in most cases,' 'normally,' and 'will often') puzzling: the issue is not about whether the problematic cases are extraordinary or rare. The question is whether externalism distinctively permits an unacceptable consequence about inference.

Burge's characterization of this sort of case is most plausible where one *rehearses* premisses *en route* to a conclusion—cases where it is not implausible that the first premiss belief is, in effect, itself causing the other premiss belief to participate in the inference. But that is only one sort of case in point; it is different from a case in which one relies on thoughts previously acquired, incorporates and activates them through *substantive* memory (remembering the thought and identifying its content), and then simply draws a consequence from them.

There should be nothing wrong with inferentially activating each of them—making them such that their having been held then causes an inference—and then drawing a conclusion, caused to do so by those premiss thoughts.[11] As one's beliefs become causally significant, through the role of substantive memory, one will be unable to avoid the implicit equivocation: the beliefs activated may involve a difference in content that is not transparent to the subject. So the subject will be *prone* to this sort of invalid reasoning, as a result of the lack of transparency of the contents involved. Old beliefs figure in new inferences, but the relevant contents are not preserved across the inference because the premiss beliefs are preserved with original contents intact (this possibility is guaranteed, remember, by the externalist response to the case of premature memory loss); and those original contents (relevantly) vary, as a result of the environments in which they were acquired. The status of the inference depends on those external contingencies—it is an inference that leaves something to chance.

4. INFERENCE

When we *draw* an inference, we draw it from premisses. The status of the inference depends in part on relations between the conclusion drawn and the premisses from which it is drawn. Significant among such relations will be (a) *causal* relations among those elements and (b) relations among their *contents*.

[11] Indeed, even the later 'activating,' through substantive memory, of the beliefs previously held is unnecessary: the beliefs might already be operating causally in the relevant way. The holding of the premiss beliefs at the earlier times may simply be in the process of causing the acquisition of the conclusion belief. Whether such a causal process involves the same concept in the premisses in the relevant way (involves the relevant 'linking' or 'redeploying' of the same concept) will be something with respect to which internal duplicates could vary.

The elements of an inference, by contrast to the elements of an *argument*, are beliefs.[12] In a standard case, we form a belief on the basis of others in a distinctive, *inferential* way. (Perhaps there is no other way to form a belief that is properly said to be formed on the *basis* of others.) Such inferences can be more or less self-conscious: we can be more or less aware of our inferring a conclusion from other beliefs. And even when one is aware that an inference occurs, one can be more or less fallible about the details: we can believe our conclusion is based on beliefs that are not, in fact, its basis. In part, this is because beliefs themselves can be more or less tacit and because we are more or less fallible about the contents of our beliefs. But even when the beliefs involved in an inference are all occurrent, the fact that one has been inferred from the others may not be apparent to the subject. Even when we are correct about the contents of our beliefs, we may be mistaken about the causal relations among them.

Inference, it might be thought, takes time. I myself am not prepared to judge this idea here. I will simply grant that inference takes time.[13] Still, it is not trivial to make sense of how long an inference takes. Does any inference one makes from a given belief in some sense begin when one first acquires that belief? That seems wrong. So one can have a belief before any inference that takes it as a premiss has begun. Let us say that when an inference involving one's belief as a premiss begins, the belief becomes *inferentially active*. It would be good to have a clearer idea about what it is for a belief to become inferentially active as a premiss in an inference. But we will take for granted that beliefs one has can at a given time become inferentially active only later, when they are serving as premisses in an inference.

A number of complications arise immediately. Imagine a simple inferential pattern, with beliefs whose contents are related as propositions symbolized in a *modus ponens* argument:

(i) p
(ii) if p, then q
(iii) q.

In other words, imagine that (i), (ii), and (iii) here represent the contents of beliefs that constitute an inference one makes. Accordingly, one's belief that p became inferentially active at some time; similarly for one's belief that if p, then q. And then there is the time of one's coming to believe that q.

Now, perhaps we should suppose that there must be at least some time at which both one's belief that p and one's belief that (if p, then q) are inferentially active. But need we suppose further that the period of time during which they're

[12] The elements of an argument, on one view, are strings of symbols. On another view, arguments consist of interpreted sentences or propositions.

[13] I say 'grant' instead of 'presuppose,' even though the assumption is useful in production of the objection discussed, because externalists, such as Burge 1998: 363, insist that reasoning takes time.

both active should include also the time at which one comes to believe *q*? I don't see why that should be necessary: so long as the premiss beliefs are inferentially active at the same time, and their being held at that same time is what causes one to believe the conclusion, perhaps some time later, at a time when their being held is not causally relevant to the conclusion being drawn, then the beliefs would seem to fit a recognizably inferential pattern.[14]

What is contemplated is *not* that one can complete an inference—coming to believe a conclusion—on the basis of premiss beliefs that one does not then hold (though that is not *excluded* either). What is contemplated rather is that the beliefs on the basis of which one is drawing the conclusion need no longer be inferentially active at the time the conclusion is drawn.[15] It could be because the premiss beliefs were held *earlier* that one is drawing the conclusion now: whether or not one holds those premiss beliefs now may be causally irrelevant.

We needn't resolve all of these complications here. (Some are the result of more general issues about causation by states.) But if one were to require for a set of beliefs to constitute an inference that the holding of the premiss beliefs at the time the conclusion is drawn be causally relevant to the drawing of that conclusion, then it would be less clear why inference should take time: an inferring subject could have the premiss beliefs for a period, those beliefs could become inferentially active for just a moment—the moment at which the conclusion is drawn—and then all of the resulting beliefs could be again inferentially inactive. The idea that there is a temporally extended transition in the inference is coherent rather with insisting that there be a non-instantaneous *period* of time comprising the inferential activity of the premiss beliefs and the drawing of the conclusion: and then the prospect of the premiss beliefs' inferential activity terminating before the drawing of the conclusion (whether or not the premiss beliefs themselves persist) is a live option.

I have said that causal relations will be significant in determining the status of an inference. *Prima facie*, it would seem that an inference consists in certain beliefs' (the premisses) *causing* another belief (the conclusion). Perhaps this is not yet adequate; but the thought that even conscious inference has anything like a *phenomenology* (a phenomenology that distinguishes the causal relations that qualify as inferential), for example, is likely not a popular philosophical position. I will not assume it; I will not assume that anything needs to be added to premiss beliefs' causing a conclusion belief in order to capture the phenomenon of inference. This enables additional elucidation of the notion of inferential

[14] I ignore the intermediate case in which premiss beliefs are lost after having been inferentially active, only to be reacquired (though not inferentially reactivated) in time for the drawing of the conclusion.

[15] Another alternative, neither assumed nor excluded, would be that although not all of the premiss beliefs need be active at the time the conclusion is drawn, not only must there have been some time at which they were all active together, but at least one of them must continue to be active at the time the conclusion is drawn—to serve as an inferential link, as it were.

activation: beliefs are inferentially active at just those times when their being held at that time causes the acquisition of the conclusion belief. Inferences last exactly as long as a certain causal process—the process in which certain beliefs cause another.[16] But the period during which the holding of the premiss beliefs causes the drawing of the conclusion may end before the inference does.

The picture that emerges so far is something like this: inference consists in certain beliefs' becoming inferentially active and in that activity causing the acquisition of another belief. Indeed, the inferential activity of the premiss beliefs just *is* their causing the acquisition of that other belief. Inference is a temporally extended, causal, psychological process.[17]

5. EXTERNALISM AND TWO THESES ABOUT INFERENCE

Burge is certainly correct that instead of inferring, fallaciously, from the belief that she once picked up some aluminum and the belief that there was some twaluminum beside her yesterday, Alice *might* infer a conclusion validly instead. No doubt, if she did, that would be thanks to her redeploying the same (relevant) concept in the two premisses. We need have no objection to using the term 'preservative memory' for whatever faculty might produce that result. But given that the subject's aluminum concept and twaluminum concept are internalistically indistinguishable, and given that preservative memory is a causal process, there is an issue about how preservative memory might operate this way within the inference: it would be akin to the body's systematically reacting differentially to sunburn and to an internalistically indistinguishable burn from a tanning lamp. Moreover, there is in any case an additional issue about any requirement that preservative memory operate in the inference at all.[18]

Sometimes, although one remembers the premiss beliefs, it is *substantive* memory that enables those preserved premiss beliefs to have their causal effect on the subject's inferential apparatus. The beliefs acquired earlier are preserved in memory, with original contents intact. But because the inferential activation is

[16] It may be that other psychological phenomena, perhaps forming and considering hypotheses, can also be constitutive of particular inferential processes.

[17] We might note as well, here, the existence of a different but similar process (I don't think it deserves the name 'inference'). Sometimes, we change the beliefs on which another of our beliefs is based. One believes that *q* on the basis of a belief—which perhaps warrants it only weakly—that *r*; but noting (or perhaps feeling more confident) that it follows from other beliefs that one has, one 'refers' one's belief that *q* to those others. As a result, one's belief that *q* is no longer based on one's belief that *r*. Again, I don't think this is a case of inference, quite. But that may be a terminological matter: the phenomenon is epistemically quite similar to inference.

[18] Goldberg 1999 raises issues, in many ways complementary to those to be raised here (though his are raised from an externalist perspective), for the responses by Schiffer and Burge to the sort of example Boghossian introduced.

the result of the use of substantive memory, given externalism, the subject might not be in a position successfully to perform the 'linking' required. And so the subject is at terminal risk of drawing the inference invalidly.[19]

This puts pressure on an issue about 'identifying' one's thoughts as containing the concepts they do. Burge of course admits (even 'begins' with that claim) that we cannot in one way distinguish the thought preserved in memory from the corresponding thought one has after the switch. The issue is whether that admission has the counterintuitive consequence that ordinary inferences are at terminal risk of invalidity.

Elsewhere Burge seems to want to characterize the flaw in the agent's reasoning to which Boghossian is trying to point as a case simply of mistaken presupposition. Admittedly, it is not intuitive that any error in presupposition is one that we're in principle in a position to correct *a priori*. So this would be a way of blunting the point of the objection. But though we don't think the truth of any presupposition is available to *a priori* inspection, we also do not think you should be *a priori* driven to a false one. Here the relevant mistake would be thinking that aluminum is twaluminum. This is a distinctive sort of erroneous presupposition: remember, the subject cannot distinguish these concepts *a priori*. The subject could not distinguish it from the belief that aluminum is aluminum. Assuming he has a belief on the matter, it would be irrational for the agent *not* to make the identification (however erroneous it might be).

Contrast a case in which a subject erroneously presupposes that Hesperus is not identical to Phosphorus. In a standard version of this case, the subject has two concepts that he, in effect (though perhaps no more than that), believes to be instantiated by different objects. That presupposition could lead to invalid inferences that the subject, having made that presupposition, would not avoid. But the subject would still be in principle in a position to avoid the presupposition involved. Whether or not Hesperus is Phosphorus is not, on the internalist view, supposed to be knowable *a priori*. In this case, it would *not* be irrational for the agent not to make the identification.[20] If the concepts are internalistically indistinguishable, the agent cannot be expected to deny the corresponding identity claim. If the concepts are internalistically different, on the other hand, neither affirming nor denying the associated identity claim would be irrational.

A different way of posing the challenge is now as something like this: given externalism, one can have erroneous beliefs that it would be irrational for you to reject, and that ignorance can undermine the validity of reasoning in a way that

[19] That the concepts be linked is not, itself, objectionable. What is objectionable is the demand that the concepts be so linked that the very same concept appear and reappear in the two premisses: in effect, that the linkage be successful. The point is that (the attempt) to link the concepts is not guaranteed to produce that result: the 'impostor' concept could erroneously be substituted. But that error would be the result simply of the lack of *a posteriori* knowledge, from which, the claim here is, the validity of our inferences should be immune.

[20] Thanks to Derek Ball for discussion here.

the agent is not in principle in a position to detect *a priori*. If we can resist Burge's appeal to preservative memory in response here, then these considerations lead to externalism's violating the *ignorance is insufficient for incoherence* thesis.

The ensuing dilemma is I think best appreciated abstractly: if at this point it were simply stipulated that preservative memory successfully ties key terms together in such a way that the sorts of invalid inferences threatening externalism are avoided—if we grant, for example, that unless preservative memory is so operating amidst the inference, the subject is simply not drawing an inference—then the corresponding issue is that since duplicates may be such that preservative memory is operating appropriately in one and not so operating in the other, duplicates may fail to be such that one is drawing an inference just in case the other is. And that consequence violates the *internalism about inference* thesis above.[21]

Let me put the challenge that remains for externalism as bluntly as possible: after a slow switch, after enough time has passed (and the relevant external conditions obtain), a subject will come to have a new concept that is internalistically indistinguishable from another concept previously possessed. Suppose we allow—because of the pressure of the case of premature memory loss—that in such cases, the agent may continue to possess that earlier concept. The subject's beliefs involving that earlier concept will be internalistically indistinguishable from beliefs involving the new concept. Accordingly, the agent will, no matter how internally coherent, be disposed to draw inferences on the basis of belief premisses with contents that are not such that they cannot both be true without the conclusion drawn being true. Thus there can be an individual who invalidly draws an inference from beliefs that do not sustain that conclusion and who is not in a position to discover that invalidity *a priori*. And, seeking to circumvent this result, demanding that, for an *inference* to be in place at all, the subject must be employing preservative memory in order to link the relevant concepts across the premiss beliefs will lead externalism to violate the *internalism about inference* thesis above.

Two agents are internalistically indistinguishable throughout a period of time. According to externalism, however, one may be drawing a valid inference while the other, because of ignorance due to external circumstance, draws an invalid inference; or, worse, one may be drawing an inference while the other is drawing no inference at all.[22]

[21] Another defense would involve supposing that successful preservative memory is something that duplicates must share. As in effect noted earlier, the supposition seems implausible: take duplicates one of whom has been switched, the other not. When they perform the inferences discussed here, it could be that one uses preservative memory successfully to make a valid inference, the other not.

[22] This is not the criticism that one of a pair of duplicates might be drawing a valid (or not) inference while the other draws a *different* valid (or not, respectively) inference. Such a criticism would not draw on our distinctively internalist intuitions about *inference*—rhetorically, it would not go far enough beyond the intuitions that are challenged, however successfully, already in the Twin Earth example.

Inference is a causal process in which premiss beliefs cause the acquisition of a conclusion belief. Beliefs sometimes become inferentially active only after they're acquired. And a previously acquired belief, which has been maintained with original content intact through the operation of preservative memory, can become inferentially active in virtue of the operation of substantive memory. If externalism is true, however, because the subject does not 'identify' the relevant beliefs appropriately—because the self-knowledge delivered is insufficiently robust—substantive memory may inferentially activate beliefs that do not stand in appropriate inferential relations. The subject may be, throughout, intuitively, perfectly rational. But the subject will now not be in principle in a position to avoid the invalid inference.

This chapter has been concerned to understand how externalism is incompatible with plausible theses about the nature of inference. Externalism is not incompatible with self-knowledge *per se*. Even properties that duplicates needn't share are properties one can know oneself to have in a distinctive way. And even if externalism is held not to be grounded on indexical phenomena, it can explain how subjects can have even infallible privileged knowledge of the contents of their thoughts. But the variety of self-knowledge that results appears to be insubstantial in an important way; and the associated inadequacy can be exhibited in connection with the phenomenon of inference. If preservative memory enables past contents to remain as the contents of current beliefs, even through changes in circumstance that produce substantial conceptual change (addition, in the disjoint type case), then, given externalism, mere false belief (that aluminum is twaluminum, for example) will suffice to make rationally unimpeachable subjects in principle unable to maintain the validity of their inferential processes: ignorance will suffice for incoherence.

Externalism can attempt to block this by denying that, in the sort of case at issue, the relevant subject is drawing an inference: one might insist that unless preservative memory is operating in the way Burge considers, so that the relevant premiss concepts are 'linked together' in a way that will prevent invalidity, the agent is not so much as performing an inference. Granted, if the premisses are so linked, the inference is valid. And although it may be that no rational agent could draw an inference without drawing it validly, the externalist now faces the second horn of the dilemma: duplicates may be such that one is drawing an inference (validly) and the other is not so much as drawing an inference. If whether one is so much as drawing an inference (never mind its validity) is a matter determined by one's internal states, then externalism is rendered problematic.

One interesting upshot of our discussion, then, is that it dramatizes a distinction between the sort of externalism deriving from consideration of Davidson's Swampman example and the sorts of externalism deriving immediately from consideration of Twin Earth (and Dry Earth) examples. On one sort of externalism, although they cannot vary with respect to whether they are so much as having a belief, or even with respect to whether their beliefs are contentful,

internal duplicates can vary with respect to *what* they believe. On a slightly more committed externalism, internal duplicates can vary not only with respect to *what* they believe, but also even with respect to whether their beliefs have any content at all (although they cannot vary with respect to whether they are so much as having a belief (or a belief-like contentless state, if we are prepared to contemplate beliefs with null content—all such beliefs being the same belief)).[23] Finally, on the most committed sort of externalism, the sort that I think is better supported (if perhaps still insufficiently) by Swampman than by Twin Earth examples, internal duplicates can vary even with respect to whether they are having beliefs at all, with or without any particular content, or any content at all.

Rejection of the *internalism about inference* thesis is likely implicit in this strongest sort of externalism. The argument of this chapter can be read as the claim that any sort of externalism will eventually be driven to be of the strongest sort: intermediate externalisms, according to which *which* content one has is a matter determined externalistically but *whether* one has a belief at all is determined internalistically, are unstable. A question that will have to be left pending here is whether even the strongest sort of externalism—once the relevant notion of internality has been yielded—should ever have been at all surprising: since internal duplicates (for the externalist, internal *physical* duplicates) needn't be the same subject, it is unclear why they need even be subjects alike. Room for the intermediate position now seems no greater than for the unstable corresponding position above. And if duplicates needn't be subjects alike—if *being a subject* is an external property—it should not have been at all surprising that duplicates needn't be believers alike, much less believe alike. Internalism should not yield the relevant notion of *internality* to externalism.

REFERENCES

Boghossian, P. 1992: 'Externalism and Inference.' In E. Villanueva (ed.), *Philosophical Issues* 2: *Rationality in Epistemology*, 11–28.

Brown, J. 2004: *Anti-Individualism and Knowledge* (Cambridge, Mass.: MIT Press).

Burge, T. 1982: 'Other Bodies.' In A. Woodfield (ed.), *Thought and Object* (Oxford: Clarendon Press), 97–120.

——1998: 'Self-Knowledge and Memory.' In P. Ludlow and N. Martin (eds.), *Externalism and Self-Knowledge* (Stanford, Calif.: CSLI Publications).

Falvey, K. 2003: 'Memory and Knowledge of Content.' In S. Nuccetelli (ed.), *New Essays on Semantic Externalism and Self-Knowledge* (Cambridge, Mass.: MIT Press), 219–40.

——and Owens, J. 1994: 'Externalism, Self-Knowledge, and Scepticism.' *Philosophical Review* 103.1: 107–37.

[23] It is interesting that, although there is little if any room between them, the prospect of 'illusions of content' is more readily contemplated than the prospect of 'illusions of belief': how could one be mistaken about whether one is so much as in a state of belief? (It is not vagueness—belief *vs.* something just short of belief—that is pressing here.)

Goldberg, S. 1999: 'The Relevance of Discriminatory Knowledge of Content.' *Pacific Philosophical Quarterly* 80: 136–56.

Ludlow, P. 1995: 'Externalism, Self-Knowledge, and the Prevalence of Slow Switching.' *Analysis* 55: 45–9.

McKinsey, M. 1991: 'Anti-Individualism and Privileged Access.' *Analysis* 51: 9–16.

Owens, J. 2003: 'Exernalism, Davidson, and Knowledge of Comparative Content.' In S. Nuccetelli (ed.), *New Essays on Semantic Externalism and Self-Knowledge* (Cambridge, Mass.: MIT Press), 201–18.

Schiffer, S. 1992: 'Boghossian on Externalism and Inference.' In E. Villanueva (ed.), *Philosophical Issues* 2: *Rationality in Epistemology*, 29–38.

Sorensen, R. 1998: 'Logical Luck.' *Philosophical Quarterly* 48: 319–34.

Wittgenstein, L. 1953: *Philosophical Investigations* (New York: Macmillan).

11

Semantic Externalism and Epistemic Illusions

Sanford C. Goldberg

I

Potential sources of false belief is a topic that is of some interest in epistemology. The topic is relevant to the theory of justified belief since, whatever else it is, justified belief is belief that is epistemically reasonable—where the relevant notion of reasonableness will have to do with the adequacy of the belief's grounds (or the processes through which it was formed) in connection with the twin aims of acquiring (interesting) truths *and avoiding falsehoods*. And the topic is relevant to the theory of knowledge as well since, whatever else it is, knowledge involves true beliefs whose truth is not a matter of luck—where the relevant notion of luck will reflect the way(s) in which the subject might have gone wrong in coming to believe what she did in the manner that she did.[1]

In thinking about such matters, it is helpful to think in terms of belief-forming and -sustaining processes. Whether a belief is justified, or amounts to knowledge, depends on features of the processes through which it was formed and sustained: the kinds of processes involved, their tendency to generate (or sustain) true beliefs and their liability to produce false ones, and the functioning of these processes on the present occasion. There are substantive disagreements among epistemologists regarding how this sort of assessment should proceed, and from whose perspective it should take place. Thus, there are questions about the relation (if any) between various reliability properties and epistemic justification, and questions as well about whether

I would like to thank the members of the audience at the 'Internalism and Externalism in Semantics and Epistemology' conference (Lexington, Kentucky, April 2005), and the members of a seminar in the Philosophy Department at Hebrew University (Fall 2005), where I presented early versions of this chapter. Special thanks to James Beebe, Jessica Brown, Tony Brueckner, Earl Conee, Gary Ebbs, Richard Fumerton, Chris Gauker, Mikkel Gerken, Deborah Gherman, Duncan Pritchard, Jim Pryor, Gil Sagi, Assaf Sharon, David Sosa, Levi Spectre, and Jonathan Yaari, for helpful discussions. I would also like to thank two anonymous referees from OUP for their comments on an earlier version

[1] See Pritchard 2004, where the anti-luck idea is developed into a full theory of knowledge.

justifiers must be reflectively accessible to the subject herself. But the general claim, that belief-forming and -sustaining processes are relevant to epistemology, should be relatively uncontroversial no matter one's answers to these and other questions of background ideology.[2]

Once we are thinking in terms of the cognitive processes underlying belief-fixation, it is possible to group the sources of possible error into categories, according to the components of the processes that might be responsible for the introduction of false information.[3] Without any pretension to exhaustiveness, four categories will loom large. These categorize cases according to whether (A) false information enters a subject's cognitive system as a result of (the subject's reliance on) a process that, by its very nature, is not reliable (at least in most or all of the sorts of circumstances in which the subject typically finds herself); (B) false information enters as a result of (the subject's reliance on) a *cognitively immature* process; (C) false information enters as a result of (the subject's reliance on) an *improperly functioning* cognitive process; or (D) false information enters as a result of (the subject's reliance on) a cognitive process in situations in which 'external' conditions are not conducive to the reliable employment of that process.[4]

I want to make three initial comments about these categories.

First, if we restrict ourselves to mature subjects who rely only on faculties/processes that are sufficiently reliable in standard circumstances—normal adult humans, say, who typically rely in belief-fixation on e.g. perception and memory but not on guesses or wishful thinking—then we need not attend to categories (A) or (B).[5] In what follows I propose to do just this, and so for the most part I will be setting (A) and (B) aside.

Second, the categories are not mutually exclusive. Even restricting ourselves to (C) and (D), we can imagine a case involving a false belief that falls under both categories. Consider the case envisaged by John Hawthorne in his contribution to this volume. He imagines a situation involving '*a priori* gas', which induces

[2] Even if one conceives of justification in terms of having reasons, it should elicit no strong reaction to say that (for example) whether a subject *S*'s claim to have seen Smith at the party constitutes a good reason to believe that Smith *was* at the party will depend, in part, on whether conditions prevailing at the time (lighting, unobstructed view, etc.) were appropriate for *S*'s reliance on (the processes constituting) visual perception.

[3] If one prefers to use 'information' in such a way that the notion of 'false information' is a contradiction in terms, then simply replace my use of 'information' with 'representational content' or some such expression.

[4] I am disregarding the sorts of error-possibilities envisaged by radical skeptics: manipulation by an evil demon, for example. Arguably, such error-possibilities can be grouped under (C), improper functionality. However, if one denies this, another category could be introduced. I won't do so, however, since here I am more interested in realistic ways in which we can go wrong, and I am assuming that such radical error-possibilities are not realistic.

[5] One might wonder when (B) is *ever* relevant: when does a subject ever rely on an immature process? One possible example is the case involving the young child's credulity. Very young children tend to be quite credulous, accepting virtually all testimony; their cognitive immaturity is seen in their insufficiently-developed sense of credibility. See Goldberg forthcoming for a discussion.

'the phenomenology of blatant obviousness for false propositions'.[6] Suppose that the presence of *a priori* gas cannot be discerned through the senses. Then if one has the bad luck of being in a place where *a priori* gas is present, and if in such a place one accepts a false proposition in virtue of the corresponding (gas-induced) phenomenological obviousness of the proposition, the resulting false belief is one that might be accommodated under (C), improperly functioning cognitive system, but also under (D), environmental factors. For the purposes at hand, however, it will be helpful to read (C) and (D) so that they are mutually exclusive. So let us say that a case falls under (C) so long as it involves improper functionality (whatever its etiology), and let us say that a case falls under (D) only when the relevant cognitive processes are mature, standardly reliable, and properly functioning. On this construal, the *a priori* gas case falls under (C) alone.

Third, category (D) itself can be further subdivided, according to the two main ways that the environment can affect the output of a mature, standardly reliable, proper-functioning cognitive process. Present 'external' conditions might fail to be conducive to the reliable employment of a given cognitive process in virtue of either *generic* or *local* factors. Take beliefs formed through reliance on the (processes involved in the) visual system. The reliability of this system depends on certain 'external' conditions: sufficient lighting, an unobstructed view, relative clarity of the visual image, and so forth. If a subject forms a vision-based belief under conditions in which these 'external' conditions are not present, the subject runs a greater risk of acquiring a false belief. In this way 'external' conditions fail to be conducive to the reliable employment of a given cognitive process in virtue of *generic* factors. Designating it as '(D.1)', this is the sort of case in which the conditions needed for the reliable employment of the process, in the sorts of case in which the process is standardly employed, are not present. But 'external' conditions can affect reliability, and so expose a subject to an increased risk of acquiring a false belief, in another way. Even if relevant generic conditions do obtain, it might happen that *local* factors are such that one is likely to acquire a false belief. The classic illustration of this is Goldman's (1976) barn-façade case: a normally functioning visual cognizer who, in standard conditions, can easily tell barns from non-barns might easily misidentify a given object as a barn when in an environment teeming with excellently constructed fake barns. Other, less far-fetched examples are easy to come by: you come to believe that the man you see in the corner is Smith; but unbeknownst to you, there are a good many people around whom you would mistake for Smith, and (also unbeknownst to you) that man is one of them. In cases like these, we can say that 'external' conditions fail to be conducive to the reliable employment of a given cognitive process in virtue of *local* factors. These will be cases—we will designate them as '(D.2)'—in which, although generic conditions for the reliable use of a

[6] See John Hawthorne, 'A Priority and Externalism,' this volume.

given cognitive process are present, conditions *specific to the present case* render the process less reliable—more susceptible to generating false belief—than it standardly is.

It is easy to see how categories (C), (D.1), and (D.2) might apply in standard cases of a perception-based judgement or belief. Suppose *S* is a mature cognitive subject, and suppose vision to be among the reliable sources. *S* comes to believe, regarding a particular person whom *S* presently sees, that he is Luciano. If such a belief is false, the case might fall under any of (C), (D.1), or (D.2). (C): Perhaps *S*'s visual system is malfunctioning, with the result that *S* has a visual impression indistinguishable from those *S* enjoys in the presence of Luciano, even though Luciano is not in fact present. (D.1): Perhaps *S* caught only a quick glimpse, and at long distance, of the man in question, and hastily concluded that it was Luciano. (D.2): Perhaps it was the case that, while *S* saw him at close range and under good lighting, the man *S* saw was one of the area's many Luciano-lookalikes, any one of whom *S* would have mistaken for Luciano.

Perception is what we might call an information *source*: it introduces novel information into one's cognitive system. But we might also wonder about the applicability of our categories to information-*preserving* processes, such as memory or reasoning.[7] These processes do not introduce novel information into one's cognitive system; instead, they enable the system to *preserve* previously introduced information for subsequent use by the subject (in action and belief-fixation). How do our categories (C), (D.1), and (D.2) bear on preservative processes? Take the process(es) involved in recollection. Reflecting on whether Luciano has been by the house this past week, *S*, seeming to remember that she saw him there recently, comes to believe that he was. If such a belief is false, the source of the error might be found in one of two places. Perhaps the process(es) that produced the original belief were flawed: here memory succeeds in preserving what was (and remains) a false belief. Alternatively, the process of memory itself might be to blame for the falsity of the belief: as when one misremembers or otherwise takes oneself to remember something one did not in fact experience. How might this sort of situation come to pass? It is clear how error introduced by the process of recollection might fall under (C), improper functionality: perhaps, unbeknownst to *S*, the underlying cognitive processes are not properly functioning. More difficult is the question whether 'external' conditions can affect one's (otherwise mature and properly functioning) memory. We might speculate that such 'external' conditions affect only those processes that are information sources, not those that are merely information-preserving. After all, if a process is merely preservative in its aim, it purports to deal only with

[7] I borrow the idea of information-preserving processes from Burge 1993, where it is introduced under the label 'content preservation'. When I speak of processes being preservative, this is a comment about the *purport* or *function* of such processes; it is not a comment on their success in carrying out the function. (The fact that on occasion 'preservative' memory fails to preserve belief does not show that such memory does not have a preservative function.)

information already in the system (so to speak); so if such a process is mature and properly functioning, it is unclear how an 'external' condition might have the effect that such a process introduces novel falsity into the system. I will be returning to this matter below, in section II.

Let us finally move on to consider the error-possibilities that arise through the preservative process of reasoning. On the basis of having accepted an argument which subject *S* takes to be valid, *S* concludes, and so comes to believe, that Luciano ate the last cookie. As in the case with the preservative process of recollection, such a belief faces two main sources of potential risk. The first has nothing to do with the process of reasoning *per se*: her reasoning-based belief might be based on a valid argument alright, only one or more of the argument's premisses was/were false to begin with. (The falsehood might have entered the system on acquisition, or else in the process of recollection.) Alternatively, the inference-based belief faces a risk that is specific to the process of reasoning itself—as when *S* errs in taking the relevant piece of reasoning to be valid when in fact it was not. If we restrict ourselves to the latter sort of case, we might again speculate that, no matter how error entered into the system, it would not be in the manner of (D), the presence of some noise-introducing 'external' condition. For while it is easy to imagine a reasoning case falling under type (B)—as when *S* was not logically competent in the first place, say—and while a case of type (C) is easily imaginable as well—as when *S* suffered from some sort of cognitive malfunction, undetectable by reflection, that rendered her unreliable in her reasoning—it would appear that there is no analog, in the case of reasoning, for an error of type (D). For how *could* an 'external' condition affect one's deductive reasoning in such a way that a false piece of information is introduced into the system—where the falsity in question is not introduced by one of the premisses, but rather is introduced in the process of reasoning itself, and flows from something other than a cognitive malfunction of some sort? Any illusory impression of good reasoning, it would seem, must be a matter of the improper functioning of the processes subserving the reasoning itself.

This idea—that there is no external condition that, without affecting the subject's proper cognitive functionality, might nevertheless introduce a flaw *in her reasoning*, and in this way lead her to form a false reasoning-based belief—conforms to standard lore regarding the nature and epistemology of deductive reasoning. It is traditionally assumed that a properly functioning, logically competent subject can tell via reflection alone whether a simple argument is valid. What is more, reasoning is truth-preserving when valid. So as long as the truth of one or more of the premisses is not at issue, it would seem that engaging in a short piece of deductive reasoning itself will not expose one to risks in virtue of the presence of any further 'external' condition. Again, this is not to say that one who *takes* herself to be reasoning validly must *be* reasoning validly. The reasoning might be complicated, and so hard to follow. Or, alternatively, one's cognitive system might be deficient in a way that is both relevant to

one's reasoning, where the deficiency in question is not discernible through one's searching reflection. This sort of possibility, described in some detail by Burge (1993), has been acknowledge even by the arch-internalist Bonjour, who concedes that

> . . . it is possible to believe oneself to be justified *a priori* [through demonstrative reasoning] and still not be thus justified, where the reason for the failure of justification is something that is, at least at the time in question and perhaps in extremely rare cases even permanently, outside one's subjective grasp. This is a stronger concession to externalism than I have heretofore been willing to make, but one that seems required by the facts of the situation. (Bonjour 1998: 128)

But the Burge/Bonjour point just tells us that one's reliance on one's own deductive reasoning competence is (like reliance on other purportedly *a priori* sources, I suppose) empirically defeasible. What it does not tell us is that such empirical defeat might have as its source some *brutely external fact* having nothing to do with (no affect on) the proper functionality of one's cognitive system. The existence of such 'external' conditions is all but ruled out by the traditional view of the nature and epistemology of deductive reasoning.

The traditional picture, then, is this. In acquiring new information about the world we expose ourselves to certain risks of introducing falsehood into our belief corpus. What is more, 'preservative' processes such as reasoning and memory *increase* one's exposure to the risk of false belief through the possibility of a relevant cognitive malfunction. At the same time, *prima facie* it is plausible to suppose that these 'preservative' processes do not render one susceptible to error in virtue of the presence of 'brutely external' conditions. That is, while the proper functionality of preservative processes might be affected by the presence of some external condition (Hawthorne's *a priori* gas case), it would appear that this is the *only* way that such processes might fail owing to an external condition. There is simply no analog of the barn-façade case for memory or reasoning.

Or so says the traditional view. In what follows, my thesis will be that if Semantic Externalism is true, then there *can* be such brutely external conditions—conditions that, without affecting the subject's cognitive functionality in any way, introduce a flaw into her reasoning, and in this way further expose her to the risk of acquiring a false (reasoning-based) belief. After bringing this out in section II, I defend the assumptions that lead to this conclusion in section III. I will conclude in section IV, briefly, by noting how this result bears on the internalism/externalism dispute in epistemology.

II

Let Semantic Externalism (SE) be the thesis that 'the mental natures of many of an individual's mental states and events are dependent for their individuation

on the individual's social and physical environments' (Burge 1986: 697).[8] It is familiar by now to wonder about the epistemic implications of this doctrine. The vast majority of people who have pursued such a question have sought to trace SE's implications regarding first-person authority and external-world skepticism. Here however I am concerned with SE's implications for the epistemic character of beliefs acquired through reasoning.

The topic is not entirely novel, as others have pursued SE's implications for the epistemology of reasoning. Of particular interest here is the argument from Boghossian 1992. A prominent critic of SE, Boghossian argued that SE is incompatible with the '*a priority* of logical abilities' (1992: 22). Although others—Schiffer 1992, Burge 1996, and Brown 2004 (esp. Chapter 5)—have responded to this argument, these reactions have all focused on showing that Boghossian's argument does not succeed in establishing anything that should prompt the rejection of SE. I agree with this assessment (for which, see below). But at the same time I think that, even assuming that it does not warrant the rejection of SE, Boghossian's argument does succeed at bringing out results suggesting SE's implications for the epistemic character of inferential belief (and perhaps for memorial belief as well; more on this below).

I begin with Boghossian's argument, which is based on the following thought experiment. Peter, an opera fan who is hiking in New Zealand,

> comes across Lake Taupo and is startled to see the famous tenor Luciano Pavarotti floating on its pristine waters. . . . This experience of Peter's gives rise to many subsequent memories on his part, and to beliefs based upon them. . . . Some years go by and Peter moves to twin earth and becomes happily esconced there. Of course, he maintains his interest in opera, and so continues to read and hear about his favorite performers. Eventually, some of the tokens of his mental names come to refer to the twin counterparts of the familiar earthly performers: tokens of 'Domingo' refer to twin Domingo, tokens of 'Pavarotti' to twin Pavarotti, and so on. (Boghossian 1992: 22)

Boghossian then has us imagine a scenario in which 'true premises conspire, through a fallacy of equivocation that Peter is in principle not in a position to notice, to produce a false conclusion' (1992: 22). Thus, some time after 'the tokens of his mental names' have the distinct twins as their references, Peter engages in the following reasoning:

(1) Pavarotti once swam in Lake Taupo.
(2) The singer I heard yesterday was Pavarotti.

(Therefore)

(3) The singer I heard yesterday once swam in Lake Taupo.

[8] So formulated, this is an instance of social or anti-individualistic externalism. There are other versions of externalism (for a discussion of which see Brown 2004 ch. 1). What I have to say here carries over to any version on which content is not transparent (in the sense discussed in Brown 2003: ch. 5).

As Boghossian describes the situation, the problem is this: although each of the premisses is true and Peter's reasoning appears (introspectively, to Peter) to be deductively valid, (3) is in fact false (the two tokens of 'Pavarotti' used in the premisses refer to different individuals). The point Boghossian wanted to make with this example concerned the epistemic status of a subject's judgements regarding the formal properties of her own arguments. In particular, Boghossian concluded that, on the assumption of SE, a subject such as Peter fails to be in a position to determine *a priori* the formal validity (or not) of her own arguments: or, as Boghossian put the point, SE is inconsistent with 'the thesis of the *a priority* of logical abilities' (1992: 22).

Let the 'propositional contribution' of an expression *N* be the contribution *N* makes to the propositions expressed by assertoric utterances of *N*-involving sentences.[9] On the assumption that the two tokens of 'Pavarotti' in the premisses of Peter's argument refer to different individuals, SE implies that these tokens differ in their respective propositional contribution.[10] Of course, Peter is oblivious to this difference in their respective propositional contributions: to Peter, the argument appears to be valid. So given SE we encounter a situation in which, although a subject is reasoning with true premisses in a way that has all the introspectible trappings of validity, he arrives at a false conclusion. On the further assumption that there is nothing wrong with the functioning of Peter's cognitive system throughout the process of his reasoning, we reach the desired result: a *brutely external* condition—one that does not affect the proper functionality of Peter's cognitive system—has introduced a flaw in that reasoning, and in this way is the source of the falsity of his belief in (3).

In short: if SE is true, we have an error, introduced in the process of reasoning itself, that falls under type (D) after all. Not only can there be (à la Burge 1993 and Bonjour 1998) reasoning cases conforming to type (C), improper cognitive functionality; given SE there can also be cases of type (D.2), in which present external conditions are not favorable to the use of reasoning owing to conditions *specific to the present reasoning itself*. This will be so whenever we have what we might call a 'Boghossian Situation', in which (i) there is a word-form in a given subject's lexicon which is assigned two distinct meanings, (ii) the subject herself is unaware of this difference, and (iii) the subject employs the word-form more than once in the course of a piece of reasoning, in a way that (unbeknownst to her best searching reflection) exhibits the fallacy of equivocation. Boghossian Situations would thus appear be an analog, for reasoning, of what the barn-façade case is for perception.

[9] The existence of semantically context-sensitive expressions, whose propositional contribution varies from use to use, suggests that we should speak of a propositional contribution of an expression *as it is used on a given occasion*; but it will not confuse matters very much to neglect this relativization to use, and so I will ignore this complication save where illumination is served by bearing it in mind.

[10] Here I am assuming that two individuals named 'John' actually have different names—something that could be represented (in a more regimented language) by the use of subscripts. My argument could be reworked so as to avoid this assumption; but I won't bother doing so.

Now one might accept this result but deny that it is newsworthy.[11] In particular, one might deny that our result depends essentially on the assumption of SE. A subject who reasons with indexicals, for example, might suffer from the same sort of fate. The following example illustrates. *S* is told, regarding a particular chair that she sees at an exhibit of antique furniture, that it was once sat in by Velázquez. As *S* is talking with someone else, the relevant chair is replaced by another lookalike. After some time *S* is told, regarding the lookalike, that it belongs to Aunt Sue. (Both pieces of testimony are true and reliable, and come from a source *S* knows to be reliable.) Then some time later *S* reasons as follows:

(4) That chair was once sat in by Velázquez.
(5) That chair belongs to Aunt Sue.

(Therefore)

(6) Some chair both belongs to Aunt Sue and was once sat in by Velázquez.[12]

If *S* uses 'that chair' in each premiss as a memory-based demonstrative picking out what (unbeknownst to her) are the two distinct chairs she observed, then we have a case that parallels the case of Peter: the subject draws a false conclusion from true premisses, under conditions in which the invalidity of the inference is indiscernible to even searching reflection, under conditions in which there is no issue of cognitive malfunction of any kind. Only here this result is obtained without the assumption of SE—thereby calling into question whether the sort of illusion in question is a new one.

However, even granting that the case of (4)–(6) is otherwise like that of Peter, there is one important difference: the vacillation in the case of (4)–(6) turns on the use of a indexical expression (the demonstrative 'that chair'), whereas the vacillation in Peter's case does not. This is based on a point that proponents of SE themselves have long made: as Burge 1982 argued (and Putnam 1996 subsequently agreed), the expressions giving rise to 'Twin Earth' cases ought not to be regarded as implicitly indexical. The context-sensitivity of (for example) 'water' is nothing like that of 'that liquid': even granting that there is a loose sense in which 'water' is context-sensitive, this is the sort of context-sensitivity of an expression whose propositional contribution is *semantically stable*. For, unlike the propositional contribution of 'that liquid', which can vary from use to use (in a way that is familiar to all mature speakers), the propositional contribution of 'water' will typically be invariant across uses within a given community.[13]

[11] I thank both Chris Gauker and an anonymous referee from OUP for this point. A similar point is made, in connection with SE and memory-based belief, by Falvey 2003.

[12] More idiomatically: 'Aunt Sue owns a chair that Velázquez once sat in.'

[13] 'Typically': most proponents of SE allow for a context-induced difference in the semantic contribution of a given term in a case of so-called slow switching. (The case of Peter is such a case.) However, such differences obtain only after a good deal of time is spent in the 'twin' environment. This alone makes such cases very different from cases of indexicals and demonstratives. Although

In this way the context-sensitivity of 'water' is unlike the context-sensitivity of 'that chair'. It is for this reason that neither 'water' nor any of the other terms for which a Twin Earth argument can be run should be assimilated into the category of '*semantically* context-sensitive expressions', where this is the category of expressions whose propositional contribution(s) can vary from use to use (in a way that is familiar to all mature speakers).

The foregoing points suggest how we can characterize the sort of epistemic illusion at play in the case of Peter, so as to make clear that the envisaged possibility of the illusion is owed to the assumption of SE. The illusion in question is an illusion of validity of a sort that is neither traceable to any flaw in logical competence or cognitive functionality, nor to any vacillation in the use of one or more of what are properly designated as 'semantically context-sensitive expressions'. It is clear that the possibility of this sort of illusion is generated on the assumption of SE itself. SE's core doctrine is that differences in reference suffice for differences in propositional contribution. This doctrine, together with the possibility of differences in reference that have gone undetected by the speaking subject herself, put the subject in a position in which Boghossian Situations are possible. But this is just to acknowledge the existence of external conditions that, without affecting the mature subject's cognitive functionality in any way, inject into her reasoning a flaw which, though not attributable to the use of any semantically context-sensitive expressions, nevertheless exposes her to the risk of forming a false reasoning-based belief.[14]

III

The foregoing result depends on three assumptions. The first is that, in the course of Peter's reasoning, the two tokens of 'Pavarotti' in the premisses refer to different individuals. The second is that there is nothing wrong with the functioning of Peter's cognitive system throughout the process of reasoning. And the third is that differences of reference entail differences in propositional contribution (SE's core doctrine). In this section I want to consider whether the proper reaction might be to repudiate one of these assumptions. I will argue that we should not.

How might one challenge the assumption that the two tokens of 'Pavarotti' occurring in the premisses of Peter's argument refer to different individuals?

it is hard to spell out the difference with any precision, I take it that the intuitive point is clear enough.

[14] In his contribution to this volume, David Sosa describes these as 'inferences that leave something to chance'. I might propose to replace that description with 'infer*rings* that leave something to chance'. Though much clumsier than Sosa's, such a description would place the emphasis where I think it belongs: on the process of reasoning itself.

Interestingly, a line of thought of this sort is pursued by Schiffer 1992 and is endorsed in a qualified way by Burge 1996.[15] Their idea is that, in contexts in which a subject is engaged in explicit reasoning, she typically intends throughout her reasoning to be using non-indexical terms in a univocal way, where the reference of the various occurrences of a single word-form are determined in something like the way that references are fixed for anaphoric pronouns used in the course of an extended passage. The motivation for such a proposal is clear: the subject who reasons in this way will not be susceptible to the sort of undetected vacillation Boghossian describes in the case of Peter. The implications of this proposal, however, are substantial.

In bringing these out it will be helpful to have a schematic account of the story. Let t_1 be the time at which Peter originally acquires the belief he expresses with (1), and let p_1 be the propositional content of that belief; let t_2 ('some years' after t_1) be the time at which he originally acquires the belief he expresses with (2), and let p_2 be the propositional content of that belief; finally, let t_3 be the time at which the reasoning takes place, and let p_1^* and p_2^* be the propositional content of the premisses Peter expresses at t_3 with (1) and (2), respectively. Now, it must be conceded straight off that the token of 'Pavarotti' that occurs in Peter's assertion of (1) at t_1 refers to Pavarotti, whereas the token of 'Pavarotti' that occurs in Peter's assertion of (2) at t_2 refers, not to Pavarotti, but to Twin Pavarotti. So those who would deny that the tokens refer to different individuals *at* t_3 (the time of reasoning) are committed to saying that one of his tokens *fails to preserve the original reference*. That is, at least one of the following two inequalities holds:

$(\neq 1) \quad p_1 \neq p_1^*;$
$(\neq 2) \quad p_2 \neq p_2^*.$

In this sense a reference-shift—and so, given SE, a shift in *propositional contribution*—has occurred.[16]

[15] Burge's claim is that speakers 'typically' reason in this way, but there may be cases in which they do not. (Even in such cases, though, Burge argues that we have no reason to abandon SE itself.)

[16] Proponents of SE will differ over which of these inequalities holds. Thus, for example, Burge (1996: 367) has argued that the first occurrence of 'Pavarotti' in the argument will fix the content of the other occurrences of 'Pavarotti' in the remainder of the argument—in which case ($\neq 2$) holds. Heal (1998) has argued (in a world-switching case involving a natural kind term rather than a name) that after the switch an 'amalgam' concept is acquired—in which case, assuming the same analysis is applied to the case above involving names, at least ($\neq 1$), and possibly ($\neq 2$) as well, holds. (($\neq 2$) holds as well if Heal allows that at t_2 Peter can use 'Pavarotti' (on Twin Earth) as a name for Twin Pavarotti. It is not clear that Heal would allow this. If, on her view, once Peter starts to apply 'Pavarotti' regularly to Twin Pavarotti, its content becomes that of an 'amalgam,' then her use of 'Pavarotti' throughout the period on Twin Earth would express an amalgam concept. In that case ($\neq 2$) is false, since the content expressed by Peter's utterance of (2) at t_2 = the content expressed by Peter's utterance of (2) at t_3 = an 'amalgam' singular concept. Even so, ($\neq 1$) holds (since Peter's use of 'Pavarotti' t_1 did not express an amalgam singular concept).) We can equally imagine a position according to which the present environment determines the content expressed by both uses of the name-form (see Ludlow 1995, where such a view has been advanced on grounds pertaining to the

The postulation of such a reference-shift has untoward implications of its own. An initial one has to do with preservative memory: insofar as there has been a reference-shift, preservative memory has failed to preserve the content of the original belief. *Prima facie*, it is strange to suppose that a semantic doctrine should have such an effect on preservative memory.[17] Perhaps it will be thought that this result can be explained away (see for example Gibbons 1996). But even if it can, the epistemic costs of doing so will be great—more than any reasonable subject should be willing to pay. For consider that Peter's story can be told in such a way that at t_1 his belief that p_1 amounted to knowledge, and at t_2 his belief that p_2 amounted to knowledge. Now, as a rational subject, Peter should not want to do anything to risk the knowledge he has, and so should not want to do anything that would risk these pieces of knowledge.[18] But using (or aiming to use) 'Pavarotti' univocally will undermine at least one of these pieces of knowledge—for the simple reason that doing so will involve surrendering the belief content that is the content of that knowledge.[19]

Consider in this light the point made by Schiffer and Burge, that a reasoning subject typically reasons with the intention to be using her non-indexical terms univocally. I submit that the case of Peter should not be taken to be 'typical', and that in any case he ought not to reason—or to be construed as reasoning—in the manner described. Consider matters from the perspective of Peter himself. (The points I make will generalize to any reasoning subject.) Presumably, if he has in fact formed the intention to be using 'Pavarotti' univocally, he has done so out of the desire to be reasoning validly. But in ordinary life the aim of valid reasoning, far from being an end in itself, serves other epistemic aims—including the aim of using one's true beliefs and knowledge to acquire other true beliefs and knowledge. As a result, if Peter's intention to be using his terms univocally

semantics of memory)—in which case, since Peter is on Twin Earth at the time of his reasoning at t_3, ($\neq$1) holds. But the point is that, given SE, at least one of ($\neq$1) and ($\neq$2) hold. In this sense content has shifted.

It is worth noting that the present point concerns not just linguistic content, but also belief content. So for example if ($\neq$1) holds, then what Peter believed when he expressed a belief with (1) at $t_1\neq$ what Peter believes when he expresses a belief with (1) at t_3. Letting the former be p_1 and the latter p_1^*, the claim is that $p_1\neq p_1^*$ (so the belief that $p_1\neq$ the belief that p_1^*). Since the claims that I will want to make concern epistemology, and in particular the epistemic statuses of Peter's beliefs throughout the process of his reasoning, I will formulate my claims at the level of belief contents, rather than at the level of linguistic contents.

[17] At any rate this would be a clear case in which the preservative process of memory is affected by a brutely external condition.

[18] Here I am supposing that he correctly takes himself to have acquired such knowledge.

[19] In reply it might be thought that, in reasoning with the univocality intention, Peter does not *lose* his old belief that p_i (for $i = 1$ or 2); rather, through the reasoning he comes to *acquire a new belief* that p_i^*. This view mirrors the so-called supplement account of switching (on which a subject who world-switched does not lose her old concept, but instead supplements it with a new one), as against the replacement account (on which the world-switched subject loses her old concept). But below I will argue that such a construal is less charitable than the alternative on which the subject reasons with *the very belief contents* she acquired at t_1 and t_2. See also Goldberg 2005.

comes into conflict with his desire not to do anything that would undermine the knowledge he has, or his desire to avoid unnecessary risks of acquiring false beliefs, it stands to reason that the latter desires undermine the motive he has for forming the univocality intention in the first place.

Some quick arithmetic can bring the present point out in connection with false belief. Without the univocality intention, Peter's reasoning is invalid, but he has on his hands only one false belief, namely, his conclusion-belief; whereas *with* the univocality intention, Peter's reasoning is valid, but he has on his hands at least *two* false beliefs, namely, his conclusion-belief, and whichever of the premiss-beliefs incurred the reference-shift.[20] Is valid reasoning really worth all that? In response it might be said that Peter himself would avow the intention to be univocal in his usage of 'Pavarotti'. But to this it can be replied that he would *also* avow his intention to preserve his original 'Pavarotti'-beliefs. The case thus calls for philosophical adjudication. And my claim is that so long as philosophical adjudication here involves being charitable, charity supports construing him as doing nothing that would unnecessarily expose him to the risk of undermining his knowledge and/or acquiring false beliefs. (Better he should have only one false belief than two.) In that case, given that he has conflicting semantic intentions, we should construe his intention to preserve the references of his original beliefs as fundamental. And, as the references of those beliefs were to Pavarotti and Twin Pavarotti, this supports the original assumption: the two tokens of 'Pavarotti' that occur in the premisses of Peter's argument refer to different individuals.

Having said this, one point must be conceded. *No matter what Peter does* at t_3, and no matter how we construe the contents of the beliefs with which he is reasoning at t_3, it is arguable that at t_3 he will fail to retain the pieces of knowledge he acquired earlier. I say this is 'arguable': it can be argued by appeal to the Relevant Alternative analysis of knowledge. According to this analysis,

(RA) *S* knows that *p* at a time *t* iff *S*'s belief that *p* is based on grounds (or formed through a process) which rule(s) out all alternatives relevant at *t*.

Suppose for the sake of argument that at t_3 (the time of reasoning) Peter has retained the belief he acquired at t_1—a belief he expressed then, and continues to express at t_3, with

(1) Pavarotti once swam in Lake Taupo.

(That is, $p_1 = p_1^*$.) And suppose too that at t_3 that Peter retains the belief he acquired at t_2 —a belief he expressed then, and continues to express at t_3 (shortly after t_2), with

(2) The person I heard singing last night was Pavarotti.

[20] See Goldberg 1999, where I bring the point out in connection with the sort of justification acquired through reasoning.

(That is, $p_2 = p_2^*$.) Even so, Peter is now in a context in which he aims to derive a conclusion from premiss-beliefs regarding both Pavarotti and Twin Pavarotti. For this reason he is now in a context in which hypotheses regarding Twin Pavarotti *are relevant alternatives* to hypotheses regarding Pavarotti, and vice versa. Focusing on his Pavarotti-belief (the one he expresses with (1)), we can then say that the following is a relevant alternative (formulated, of course, as we would formulate it):

(1A) *Twin Pavarotti* once swam in Lake Taupo.

Given (RA), Peter counts at t_3 as knowing (1), *only if* at t_3 (1A) can be ruled out. But at t_3 (1A) cannot be ruled out, since by hypothesis Pavarotti and Twin Pavarotti are indistinguishable to Peter.[21] (Similar reasoning can be used to show that he no longer knows (2).) So it would seem that *no matter what Peter does* at t_3, at t_3 he will fail to retain the pieces of knowledge he acquired at t_1 and t_2—and this, *even if he preserves the content of the beliefs involved*.

I concede the point. This does not affect my contention that Peter is best (most charitably) construed at t_3 as preserving the references of the belief contents he acquired at t_1 and t_2. For one thing, the point about false belief remains: we minimize the number of false beliefs he acquires through the process of reasoning, if we construe as fundamental his intention to be preserving the references of his original beliefs. But another point can be made in defense of my contention. We are trying to decide what grounds there are for accepting or rejecting the assumption that the two tokens of 'Pavarotti' that occur in the premisses of Peter's argument refer to different individuals. My initial claim was that Peter should avoid doing anything that would jeopardize the knowledge he acquired at t_1 and t_2; and since construing his univocality intention at t_3 as fundamental would do just that, I argued, he ought not to be construed in this way. Now it is true, as the preceding paragraph argues, that even if it is allowed that he is reasoning with the very beliefs he acquired at t_1 and t_2, the knowledge he acquired at t_1 and t_2 is in jeopardy: it is put in jeopardy by the very act of reasoning itself (which makes some previously irrelevant alternatives relevant). But I submit that losing knowledge *owing to the introduction of a new relevant*

[21] Admittedly, given SE, Peter would not have formed a belief regarding Twin Pavarotti at t_1: at that time he had never been exposed to Twin Pavarroti, so his use of 'Pavarotti', conditioned by his causal interaction with Earth Pavarotti, would have referred to Earth Pavarotti, and so would have been part of a belief regarding Earth Pavarotti. In that case, so long as Peter's memory preserved the content of his belief, he would continue at t_3 to hold on to that same true belief. But our question here regards, not the truth, nor even the *reliability*, of his belief at t_3; it regards rather his ability at t_3 to discriminate the case in which hypothesis regarding Earth Pavarotti holds, from the case in which an hypothesis regarding Twin Pavarotti holds. And, as argued in Brown (2004: 45–59), the reliablity of one's belief is compatible with one's inability to discriminate the actual case from alternatives incompatible with one's belief. So the fact that Earth Pavarotti and Twin Pavarotti are indiscriminable to him can be used to conclude that he cannot perform the requisite discrimination—in which case he does not know at t_3 that Pavarotti once swam in Lake Taupo.

alternative is one thing; losing knowledge *owing to the failure to preserve the content of a piece of knowledge* is another. The former loss is less bad, epistemically speaking.

To see this, suppose that we construe Peter, as I have been suggesting we should, as reasoning with the very belief contents he acquired earlier. While it would be true that in the context of his reasoning he would no longer count as *knowledgeably* believing these contents (owing to the relevance of alternatives he can't rule out), nevertheless he would count as knowledgeably believing them in all contexts in which the now-relevant alternatives are (once again) irrelevant. This is because he would also be preserving the connection between his original grounds (at t_1 and t_2, respectively) and his current beliefs (at t_3). Part of what rendered Peter's beliefs at t_1 and t_2 *knowledge* was that they were appropriately caused by the very fact that made them true. It is true that, assuming he preserves these beliefs at t_3, this point about these beliefs' causal history is no longer sufficient, at t_3, to render those beliefs knowledge. Even so—and this is the key point—preserving the original beliefs should be of paramount importance to Peter: in so doing his memory operates so as to retain the sort of connection between a belief and the fact that renders it true, which, in at least some contexts, would be sufficient for counting the belief knowledge.[22]

Let us step back. Everyone can agree that it would be best if a reasoner in a Boghossian Situation simply refrained from engaging in the problematic reasoning. The difficulty arises because one can't tell whether one is in a Boghossian Situation. The question, then, is this: how ought one to reason, given the possibility of Boghossian Situations? I see this question as relevantly similar to a question about perception: how ought one to fix belief in perception cases, given the possibility that one can be misled by the perceptual appearances? That we can be led to form a false reasoning-based belief through being deceived by the validity-appearances in a Boghossian Situation should no more stop us from engaging in reasoning than the fact that we can be misled by the perceptual experiences into forming a false belief should lead us to refrain from forming perceptual beliefs. The reasonable thing to do in both cases is to take reasonable precautions. Of course, taking such precautions will no more ensure that one avoids the illusions associated with Boghossian Situations than taking the relevant precautions in the perceptual case ensures that we are never the victim of a perceptual illusion of any sort. It is here that we see the relevance of the foregoing reflections. These reflections suggest that it would be best—*most reasonable, all things considered*—for reasoners to reason with the fundamental intention of preserving their original belief contents. No doubt, the reasonable subject who does engage in deductive reasoning will also have the strong conviction that her expressions *are* being used univocally. But this is a belief that she should recognize to be fallible. I submit, then, that the best construal of Peter—the most

[22] I develop this theme further in Goldberg forthcoming.

charitable one—is one in which he preserves his original references throughout the piece of reasoning.

I have just defended the first of the three assumptions on which our result rests. Next, consider how one might think to challenge the second of the assumptions, to the effect that there is nothing wrong with the functioning of Peter's cognitive system throughout the process of reasoning. Given the way Boghossian described the case, it is not clear how this sort of challenge can be motivated. But suppose one favors the analysis that emerges from the Schiffer–Burge suggestion that subjects typically reason with univocality intentions. In that case, the contention that Peter's cognitive system is not functioning properly might be motivated by appeal to the fact that Peter's memory has not succeeded in preserving the content of his original beliefs, together with the claim that content preservation is a proper function of memory.[23] But this sort of argument should not convince. For one thing, the hypothesis of improper functionality would still be only as good as the analysis on which it is based; and I just gave reasons, independent of the present considerations, for thinking that this analysis (inspired by the Schiffer–Burge reaction to Boghossian's argument) is unacceptable. But even waiving this point the allegation of cognitive malfunction is not plausible. This is so for two further, related reasons. First, it would seem that preservative memory *is* functioning properly even in the case as described by Schiffer and Burge. Admittedly, in this case what is preserved is not the original belief contents themselves, but rather the single propositional contribution of Peter's uses of 'Pavarotti' across his reasoning at t_3. But Burge himself (1993) is quite explicit that this is a very important function of preservative memory, for it is the univocality-underwriting aspect of preservative memory that enables us to rely on reasoning even when it involves long chains of deduction. Of course the proper functioning of *this aspect* of preservative memory comes at the cost of the failure of memory to preserve the original belief contents acquired. But—and this is my second, related point—surely it is no strike against the proper functioning of preservative memory that, when simultaneously confronted with two demands *whose mutual satisfaction is not possible*, it satisfies one of them. I conclude, then, that there are reasons for accepting, and no compelling grounds for rejecting, the assumption that Peter's cognitive system is functioning properly throughout the process of reasoning.

This brings us to the last of the three assumptions underwriting the present argument, the core doctrine of SE itself: differences of reference suffice for differences of propositional contribution. If added to the two recently defended assumptions, making this assumption forces us to acknowledge the possibility of a kind of 'brutely external' flaw in the very process of reasoning. Should this lead us to reject the third assumption, and with it SE itself? I don't think so. The

[23] Indeed, it would seem that this is the *only* way by which to motivate the claim of improper functioning.

independent considerations supporting SE are overwhelming; and nothing about the present result gives us anything strong enough to motivate SE's repudiation. It is true that the reasoning exhibited by a subject in a Boghossian Situation will be cases in which the subject fails to be in a position to discern via reflection the flaw in her reasoning.[24] But, as Brown (2003: ch. 5) has argued at length, this result is independently defensible. Since I do not have anything more to add to what Brown herself has already said on this matter, I will leave it at this.

IV

In this chapter I have argued that, if SE is true, then there can be *brutely external conditions*—conditions that do not affect the proper functionality of the mature subject's cognitive system—that serve to expose her reasoning to the risk of invalidity, thereby exposing her to the risk of acquiring a false belief through that reasoning, and in ways that escape even searching reflection on her part. In establishing this result I have had little to say about how this bears on issues pertaining to the internalism/externalism debate in epistemology. In this connection one final point is worth making. If left to stand, our conclusion would force us to acknowledge a certain limitation in our ability to rationally police our reasoning-based beliefs via introspection alone. This would reveal yet another way in which what is available through even searching reflection is less epistemically robust than we might previously have thought. That such a result has been established in connection with *reasoning-based belief* is significant, since it is here that we would expect epistemically internalist accounts to have their strongest case. I leave it to the reader to decide what this says about the prospects for a viable epistemic internalism.[25]

REFERENCES

Boghossian, P. 1992: 'Externalism and Inference.' In E. Villanueva (ed.), *Philosophical Issues* 2: 11–28.

——1994: 'The Transparency of Mental Content.' In J. Tomberlin (ed.), *Philosophical Perspectives*, 8: *Language and Logic*, 33–50.

Bonjour, L. 1998: *In Defense of Pure Reason* (Cambridge: Cambridge University Press).

Brown, J. 2004: *Anti-Individualism and Knowledge* (Cambridge, Mass.: MIT Press).

[24] I should say that, if one favors the analysis inspired by the Schiffer–Burge suggestion about univocality intentions, then the relevant failure will be a failure to be in a position to discern via reflection that one's memory has failed to preserve the content of one's original belief—and so has failed to preserve the sort of connection between belief content and support which, in at least some circumstances, is sufficient for knowledge.

[25] See the penultimate section of Goldberg forthcoming for more discussion on the implications for epistemic internalism.

Burge, T. 1982: 'Other Bodies.' In A. Woodfield (ed.), *Thought and Object* (Oxford: Oxford University Press), 97–120.

——1986: 'Intellectual Norms and the Foundations of Mind.' *Journal of Philosophy* 83: 697–720.

——1993: 'Content Preservation.' *Philosophical Review* 102.4: 457–88.

——1996: 'Memory and Self-Knowledge.' In P. Ludlow and N. Martin (eds.), *Externalism and Self-Knowledge* (Palo Alto, Calif.: CSLI Press), 351–70.

Falvey, J. 2003: 'Memory and Knowledge of Content.' In S. Nucetelli (ed.), *New Essays on Semantic Externalism and Self-Knowledge* (Cambridge, Mass.: MIT Press), 219–40.

Gibbons, J. 1996: 'Externalism and Knowledge of Content.' *Philosophical Review* 105.3: 287–310.

Goldberg, S. 1999: 'The Relevance of Discriminatory Knowledge of Content.' *Pacific Philosophical Quarterly* 80.2: 136–56.

——2005: '(Non-standard) Lessons from World-Switching Cases.' *Philosophia* 31.1: 95–131.

——forthcoming: 'Anti-Individualism, Content Preservation, and Discursive Justification.' *Noûs*.

Goldman, A. 1976: 'Discrimination and Perceptual Knowledge.' *Journal of Philosophy* 73: 771–91.

Heal, J. 1998: 'Externalism and Memory II.' *Proceedings of the Aristotelian Society* suppl. vol. 72: 95–109.

Ludlow, P. 1995: 'Social Externalism, Self-Knowledge, and Memory.' *Analysis* 55.3: 157–9.

——1997: 'On the Relevance of Slow Switching.' *Analysis* 57: 285–6.

Pritchard, D. 2004: *Epistemic Luck* (Oxford: Oxford University Press).

Putnam, H. 1996: 'Introduction.' In A. Pessin and S. Goldberg (eds.), *The Twin Earth Chronicles* (Armonk, NY: M. E. Sharpe), pp. xv–xxii.

Schiffer, S. 1992: 'Boghossian on Externalism and Inference.' In E. Villanueva (ed.), *Philosophical Issues* 2: 29–38.

12

Psychological Externalism and the Role of Belief in the Analysis of Knowledge

Joseph Owens

Psychological externalism is a metaphysical thesis—one to the effect that physically identical individuals can differ in their psychologies, in their beliefs and desires, even though they are molecule for molecule physical replicas.[1] Metaphysical or not, an astonishing array of epistemological claims have been linked to externalism. Some of the better-known examples: Hilary Putnam has argued that it provides for a refutation of Cartesian skepticism; Michael McKinsey and Anthony Brueckner have argued that rather than leading to a refutation of Cartesian skepticism it deepens the skeptical quagmire, undermining self-knowledge; Davidson has argued for a split decision: the kind of externalism usually associated with Burge (arguments which turn on subjects having incomplete or incorrect conceptual understanding) tells against self-knowledge, while examples of the Putnam kind, examples in which the subject merely fails to have any scientific understanding of something (e.g. water), don't generate any new kind of skepticism, and indeed are part of the answer to skepticism.[2] Then there is Timothy Williamson, whose concerns are closer to mine in this chapter. Williamson has argued that psychological externalism undermines all theories of knowledge, internalist and externalist, which construe knowledge in terms of belief plus other factors (truth, justification, being the product of reliable mechanisms, etc.).[3] All such efforts are motivated by the intuition that knowledge, unlike belief, should not be thought of as a mental state because the question as to whether a subject S knows that P is not determined simply by how it is with the subject—we must also take into account the nature of the subject's world. This motivation is now undercut by the recognition that belief, a paradigm mental state, is externalist in an analogous sense. If the question as to whether S believes that P is in part determined by the character of S's world, without this telling

[1] See Burge 1979.

[2] See Putnam 1981; McKinsey 1991, 1994; Brueckner 1986, 1990, 1994; Davidson 1987, 1990, 1991.

[3] See Williamson 2000.

against beliefs being mental states, then why think the dependence of knowledge on external features of the subject's environment should tell against knowledge being a mental state. And Williamson, of course, has gone on to develop such an account—a disjunctive theory of knowledge.

Now one might, of course, take a very different lesson from this apparent similarity between belief and knowledge; one might retain the initial plausible intuitions that led one to deny that knowledge was a mental state, and infer rather that belief too is not a mental state. This is the avenue I want to explore in this chapter. This too has consequences for the epistemological debate, for it too will undercut much of the philosophical drive to analyze knowledge in terms of belief to the extent that that drive was motivated by the intuition that belief was a real internal state unlike the hybrid knowledge. It is not my intention, in arguing that the externalist examples tell against construing beliefs, etc. as mental states, to argue against ordinary usage, to argue that ordinary, non-philosophical talk of states of belief is somehow confused, wrong, or misleading. These ordinary locutions are not themselves the subject of the debate, but rather a certain philosophical interpretation of such talk. Philosophical talk of *psychological states* or states of the mind must commit the theorist to something of substance if that debate is to be of any consequence. What is at issue in talking of beliefs, etc. as mental states? What is one claiming when one claims that knowledge, for example, is a mental state? I don't intend to answer for all, but it seems to me that for most theorists this conception of belief, desire, etc. as genuine mental states is taken to commit them to thinking of *such states* as (i) being the designata of expressions such as 'Jones's belief that Mary is ill', and (ii) as serving to causally explain behavior. This is a fairly minimal requirement. The first component rules out antirealist, non-categorical construals of psychological ascriptions as, for example, ascriptions of behavioral dispositions. This requirement is one to the effect that such nominalizations denote, and they denote *mental states*—beliefs, desires, etc. This requirement commits one to the existence of beliefs, desires, etc. The second element serves to provide these real states with a real role to play; they are not mere epiphenomena, but serve to causally explain a subject's behavior. Let us call this kind of realism 'intentional realism'. I will argue that this intentional realism is incompatible with psychological externalism.

First, let me set out an externalism argument in a form that isolates the inference I want to focus on.

1. Alf, a competent speaker of English, has acquired the term 'aluminum' though he knows little formal chemistry. We may assume that the scientific community is well aware of the micro properties of aluminum.
2. Alf has heard the usual clichés about aluminum—that it is a light metal, used for making pots, pans, etc.—and he believes what he hears. In particular, Alf is disposed to sincerely assert, 'cans are made of aluminum'.

3. We have every reason to think that Alf believes that cans are made of aluminum.
4. Now consider a counterfactual situation (a non-actual possible situation or possible world). The inhabitants of this world use a language that is superficially very like English, one that sounds just like English. In particular, they use a word that sounds just like our 'aluminum'. Like us they use it to designate a light metal, one that they use in making pots, pans, etc. But this metal, the one they call 'aluminum', is not aluminum; though it has the same surface features as aluminum, it has a very different microstructure. We introduce an English term for this metal, which does not occur in our world. Call it 'twalum'. So if we want to translate their term 'aluminum' we should use this term 'twalum'.
5. Consider now an inhabitant of this other world, Alf*. Alf hears the common clichés about twalum, and he believes them. He has heard that twalum is a light metal, that it is used in the making of pots and pans, etc. And indeed he is disposed to assert that cans are made of twalum. They of course put things differently. Alf* hears what he hears in his own language, and he is disposed to speak in his own language: Alf* hears such things as 'cans are made of aluminum', 'aluminum is used in the making of pots and pans', and he is disposed to utter 'aluminum is a light metal'. In English, however, we have every reason to say that Alf* is sincerely disposed to say that cans are made of twalum and that he believes that that cans are made of twalum, etc.
6. Alf and Alf* thus may be said to have different beliefs; one may be said to believe something the other does not. One may truly say of Alf, but not of Alf*, that he believes that cans are made of aluminum.
7. They differ in this way—meriting different psychological characterizations–even if they are exactly alike from the non-intentional perspective, if they are particle for particle replicas, hear the same sounds, have the same dispositions to behavior, have the same internal streams of consciousness, etc.
8. Thus, they have different psychological states even though they are physically identical; psychophysical supervenience is false.

There is, of course, much that could and should be said about each of the steps here, but I want to focus on the move from (7) to (8).[4] There is today widespread agreement that something like (7) is true, that

(i) —— believes that cans are made of aluminum

[4] Wittgensteinians, for example, will object to steps 4 and 5 on a couple of related grounds. First, the philosopher who advances this externalist argument is assuming that our ordinary concepts of belief, desire, etc. are so fine tuned as to decide such issues, and for the Wittgensteinian there is really no reason other than philosophical prejudice to think this is so. Any philosopher worth his salt is able to imagine all kinds of problem cases, all kinds of challenging examples, examples far

is true of Alf, and

(ii) —— believes that cans are made of aluminum

is false of Alf*.

In line with this, let us also assume that (iii) gives us a true characterization of Alf*

(iii) Alf* believes that cans are made of twalum.

At this point a number of options arise: (1) one might just stop at (7), and understand the examples as serving simply to demonstrate the context-sensitive character of attitude ascriptions. But, while there is much to be said for this conservative reading of the examples, this is not where most theorists stop. Line (7), which is simply about our linguistic practice, seems to lead directly to a metaphysical result such as (8), and most of the recent argument has focused on line (8). In this debate philosophers have typically assumed that the nominalized expressions:

(a) 'Alf's belief that can are made of aluminum'

and

(b) 'Alf*'s belief that cans are made of twalum'

designate psychological states, propositional attitudes, states that bear content, and bear different content in each case. Then have gone on to discuss the kinds of states involved, the kinds of content involved, the question as to whether the content differences provide for a difference in causal powers, etc.[5]

My goal in this chapter is quite simple: it is to cast into doubt this brand of intentional realism, the assumption that (a) and (b) designate psychological states that play a causal role in the explanation of behavior. As far as I can tell, (7) tells against such realism.

Once one assumes that (a) and (b) designate states that bear contents, two bleak and blunt options present themselves:

removed from ordinary practice, and it is simply a mistake to suppose that our ordinary concepts are somehow inadequate if they fail to dictate an answer for each such imagined example. (See, for example, *Philosophical Investigations* 81.) Second, the argument strategy assumes that decisions as to whether this belief is the same as that belief can be made relatively free and independent of conversational context. Reflection on our practice seems to suggest that this too is a mistake. If, for example, our concern is with the truth and falsity of various claims made by Alf and Alf* we may indeed distinguish what they say and believe. If, however, our concern is with less finely individuated behaviors, with, for example, what they do when asked for an aluminum bat, we may not think that there is any difference of substance in what they believe. These remarks, too, are of course rough and in need of refinement, but I will not attempt to address such Wittgensteinian worries in this chapter. Given my concerns, there is no need to follow up on these issues, since most parties to the debate are not Wittgensteinians, and in any case, the Wittgensteinian picture is even less hospitable to intentional realism.

[5] See for example Fodor 1987; Owens 1993.

1. The content is taken to be non-essential to the state designated by the nominalized expression.
2. The content is taken to be essential to the state designated by the nominalized expression.

In the remainder of this chapter I want to investigate both options, and I will argue that neither option is a happy one for the intentional realist.[6] The first option undermines the intuition that (a) and (b) designate psychological states, and the second option undermines the intuition that the psychological states play a causal role in the genesis of behavior.

I

Consider the first option. The question of interest to me is this: What kind of picture of the mental results when one opts for the view that such nominalizations designate states that have content, but have the content they have contingently; when one assumes that the states designated could have had different contents than the ones they in fact have. There is I think little or nothing to be said for this conception of mental content; variations on this view have been proposed and all have been found wanting, and wanting in a systematic fashion. But my goal in the next few pages is not so much to argue against such views, but to try to gain a clearer perspective on the resulting picture of the mental; a better understanding of how, on this model, we are to think of the mental.

In the 1950s and 1960s the vision of mental properties (including content) as being in some sense contingent properties was integral to then popular type/type identity theories. True, most of these theorists were primarily concerned with debating the status of experiential states—pains, after-images, etc.—the states that seemed to pose the greatest challenge to the materialist picture. The conceptual distance between talk of pain and talk of physiological states seemed to many to tell against the very possibility of any mind/brain identity. Smart and others, however, argued that this kind of conceptual distance no more bars identity in this case than it does in more familiar cases such as the identity of water and H_2O, or the identity of lightning flashes and electrical discharges. These familiar examples were advanced as providing us with a model, a way of conceptualizing mental/physical identities. The claim was that the conceptual gap between experiential concepts and physiological concepts is on a par with the gap between the ordinary concept of water and the chemical concepts of oxygen and hydrogen. And just as they don't bar identity in the case of the case of water,

[6] In this discussion I will ignore appeals to specialized notions of content, 'narrow-content', 'character-like content', etc. I have argued elsewhere against such strategies, and I shall simply not concern myself with them here. See Owens 1987, 2004.

so they don't bar identity in the case of mind. In both cases we have something identified in a run-of-the-mill fashion, and then scientific investigation provides us with an alternative, scientific description of one and the same thing.[7]

Smart was not explicitly concerned with talk of belief and belief content. He assumed that the behaviorist could eventually give some kind of Rylean account of such talk. But if we take this kind of identity theory as a theory of mental states in general, as a theory of the attitudes as well as the experiential states, then the resulting theory is one that apparently construes nominalizations such as (a) as designating a physical state which happens to be contingently identical to a state identified by way of content. And, on this account, all we need to understand what is at stake in such contingent identities is to think of them on the model of the familiar identities of science, water = H_2O, etc. Despite the claims made for it, this supposed analogy between mind/brain identities and familiar physical/chemical identities is of no help in articulating the picture of mental we are concerned with here. The problem, of course, is the one highlighted by Kripke: the very identities selected by Smart and others are, if true, examples of necessary identities rather than contingent identities. And thus this kind of classical identity theory is of no help in understanding the nature of the contingent content interpretation.[8]

We can, however, gain some understanding of what this conception of the mental is like from a glance at David Lewis's identity theory. Lewis defended a specific account of the meaning of the mental vocabulary, an account of the mental vocabulary that construes nominalizations such as (a) as designating states that have content and have it contingently. I am not endorsing Lewis's account. It is, I have argued elsewhere, deeply misguided, and, indeed, the following remarks may be construed as telling against the contingent designator analysis, but it does afford us a window through which we can see our way to a better understanding of what is at stake in this talk of contingently possessed content, and this is what I want to focus on.

Lewis viewed mental terms as being analogous to the theoretical terms of science, and argued that both sets of terms are defined by the theories in which they figure—the mental terms being defined by the platitudes of folk psychological theory. Let me follow Lewis and use the example of the term 'pain'—assuming that what holds true for the experiential terminology also holds true for the propositional attitude terminology. The term 'pain', he claims, serves to abbreviate a description of the form 'the state that is *typically* induced by tissue damage—that in turn causes one to feel self-pity . . . that causes one to cry out, etc. etc.'.[9] If folk theory is a true theory then this description must pick out a unique state, one that any animal is in if and only if it is in pain. As to what kind of state is involved, Lewis argues that we, as materialists, have no choice but to suppose that the state in question is a physiological state. Pain,

[7] See Smart 1959. [8] Kripke 1980. [9] See Lewis 1966, 1970.

then, is said to be identical with some physiological state N_i, and an individual is in pain if and only if that individual is in N_i. This theory seemed to give Lewis just what he wanted: By assuming that the state in question need only *typically* play the characteristic role of pain, Lewis is able to allow for the obvious fact that paralytics can be in pain, even though it does not play its characteristic role in their lives (no moaning, wincing, etc.); and the theory allows for the fact that this state N_i, which does in fact play the role of pain, *need not* have played the role of pain—there is nothing absurd in the supposition that a creature might be in pain even though it lacks a physiology like ours, and there is nothing absurd in the supposition that this very state N_i might occur in a creature and that creature not be in pain. The description abbreviated by 'pain' just happens to be satisfied by N_i in the actual world; it is satisfied by various other states in different possible worlds.

This initial version of Lewis's theory encountered a number of problems; the most obvious being that we have no reason to suppose that there is some physical state common to all animals experiencing a state such as pain. In response to such concerns, Lewis amended the theory in various ways; in one paper he proposed that we should think of the term 'pain' as being indexed to individual subjects, designating in the case of each individual S the state in S that plays the role characteristic of pain; Then again, he has also argued that we should think of 'pain' as being indexed to species, as designating in each species the state that plays the characteristic role of pain in that species.[10] Each of these different accounts faces its own problems, and there is no reason to think that any of them are successful.[11] But Lewis's account is the clearest account we have of a theory that treats mental characteristics—including content—as non-essential. And I want to use this fairly specific account to better understand the picture that results when one treats nominalizations such as (a) as designating states that have content, but have it contingently.

On this account there are mental terms, and these terms are true of some states and events and not others, but there are no mental states in any interesting sense of the term. The mental terms abbreviate complex functional descriptions, descriptions that designate physical states (if they designate anything). Simply put, on this account there are physical states that happen to interact in various ways, and because of these contingent interactions, they get to be called 'mental'. If one looks to the case of the paralyzed individual and one asks how such a one can be in pain—given that no state in her actually plays the characteristic role of pain—the answer is direct: she is in pain because all that is needed is that she be in the appropriate physiological state N_i—the state that happens to play the characteristic role of pain in other members of the species, though it fails to do so in her. One can of course call it a mental state, because it is designated by 'pain', and 'pain' is a mental term, but this kind of theory is so deflationary that calling

[10] See Lewis 1969, 1980.

[11] I discuss some of these concerns in Owens 1986.

it a mental state amounts to no more than saying of this physiological state that it plays role R—where this role is entirely specified in non-mental vocabulary. *To be blunt about it, the token state designated does not have the identity conditions as pain—this very state can be present in the absence of pain, and absent in the presence of pain.*

To fully appreciate this point, it is important to sharply distinguish between the contingent theory of Lewis and the kind of functionalism once defended by Hilary Putnam—empirical or scientific functionalism.[12] Scientific functionalism differs from the Lewis-style functionalism along a number of parameters: it is advanced as an empirical thesis, to be evaluated on scientific rather than *a priori* grounds; since it is not an attempt to explicate the meanings of the psychological terms, it need not rely on commonsense vocabulary, but can employ a scientific vocabulary in formulating its definitions, etc. For our purpose, however, the most important point is that Putnam-style functionalism is a thesis about the nature of mental states or properties, not about the meanings of the psychological vocabulary; it is not a thesis to the effect that some states defined independently of psychological considerations—e.g. physiologically defined state—count as mental states—get to be called beliefs, etc.—if they happen to satisfy this or that accidental condition (as in Lewis's account), but rather one to the effect that mental states and properties are functionally individuated states and properties. On a Putnam-style account, nominalizations such as (a) are taken to designate a special kind of state, functionally individuated states, which are then identified with psychological states; *the identity conditions of these functional states are dictated by psychological considerations, and they are advanced as the identity conditions for psychological states not physiological states. These functionally individuated states are supposedly individuated with an eye to content—the role that serves to define the state is a role that gives it the content it has.* There is no such ontology of special kinds of states on a Lewis-type account; on such accounts nominalizations such as (a) do not designate functionally defined states; no, they designate neurological states—*states* whose identity conditions are of neurological character rather than a mental character. On the Lewis account, psychological considerations play no role in defining states designated by psychological terms; psychological factors are relevant only to the definition of the psychological terms which then ***serve to contingently designate states defined independently of any psychological considerations.***

Further, on an account such as Lewis's one cannot say that the physiological state N_i is an instance of the state of pain—that such physiological states serve to instantiate the state of pain. There is no such abstract state of pain to instantiate; there are physiological states and these same interact with each other in characteristic ways, and this is all we need to support our pain talk. Simply providing application conditions for psychological characterizations does

[12] See Putnam 1975b.

not mean that one has thereby provided a metaphysics of the mental, a theory of mental states; it does not commit one to mental states at all.[13]

The anti-realist character of Lewis's proposal is evident from an examination of the interplay between sortal and modal conditions. On a contingent content type account, such as Lewis's, one can say of the state designated by (a)—'Alf's belief that cans are made of aluminum'—that *it* could have had a different content; *it* could have had the content that cans are made of twalum. One who opts for this kind of analysis can perhaps swallow this—i.e. that this description designates a state that has a certain content, but could have had a different content. They can swallow this because they are thinking of the description as designating a physiological state. But it is another matter altogether to say that *Alf's state of believing that cans are made of aluminum could have had the content that cans are made of twalum*. This suggests some kind of incoherence, a failure to recognize that a difference in content makes for a different belief. If this belief has a different content than that belief, then surely that's enough to count them as different beliefs (see below). But if we think it proper to say of (a) not only that it designates some state or other, but that it designates *a psychological state, the state of believing that cans are made of aluminum,* then we should be willing to say of *this state, the state of believing that cans are made of aluminum*, that *it* could be the state of believing that cans are made of twalum. But we are simply not open to this possibility. We are able to coherently think of content as contingent to the state designated by (a) *only* when we think of this state as no more than the state, *whatever it is*, that is designated by this expression. We are ready to speak of the state designated by (a) as possibly having the content that cans are made of twalum only to the extent that we hide from ourselves the supposition that the state we are talking about is supposed to be the state of believing that cans are made of aluminum.

Now, one might be tempted to think that the differences between the sentences

(c) The state of believing that cans are made of aluminum might not have been have been the state of believing that cans are made of aluminum, but rather the state of believing that cans are made of twalum.

and

(d) The state designated by 'Alf's belief that can are made of aluminum' might not have been the state of believing that cans are made of aluminum, but rather the state of believing that cans are made of twalum

is due to the more general contrast between sentences of form:

1. The state of # might not have been the state of #.

[13] This point is evidenced by behavioristic analyses.

and those of the form:

2. The state designated by '#' might not have been the state of #.

That is, one might think that sentences of form (1) are typically problematic, and thus the issue here has nothing to do with the fact that (c) purports to be about psychological states. So, this objector urges, the contrast between (c) and (d) can't be used to support any substantive claims about psychological states or psychological talk; in particular, one cannot appeal to the contrast between (c) and (d) to argue that the non-problematic character of (d) is due to the fact that in it the expression 'the state designated by ''the state of believing that cans are made of aluminum'' ' is not being used to designate a psychological state. But this is a temptation to be resisted. When the expression 'the state of #' is clearly a contingent designator, as Lewis is construing it, sentences of form (1) are simply not problematic. Consider: 'The state favored by Susan might not have been the state favored by Susan'; here the state in question is mentioned by way of an obviously contingent descriptor, and here there is nothing odd or problematic about the sentence. The problem with (c) derives not from the general form of (c) but from the specific psychological vocabulary employed in (c).

We can gain some further appreciation of the subtleties and the tensions in this dialectic from Donald Davidson's discussion of Burge-like cases. Davidson argues that, contrary to Burge and others, these externalist examples do not tell against the kind of token identity theory he favors.

> It should be clear that it doesn't follow, simply from the fact that meanings are identified in part by relations to objects outside the head, that meanings aren't in the head. To suppose this would be as bad as to argue that because my being sunburned presupposes the existence of the sun, my sunburn isn't a condition of my skin. My sunburned skin may be indistinguishable from someone else's skin that achieved its burn by other means . . . yet one of us is really sunburned and the other is not. This is enough to show that an appreciation of the external factors that enter into our common ways of identifying mental states does not discredit an identity theory of the mental and the physical.[14]

For Davidson, then, these examples posed no problem: sure, these examples demonstrated that we would say of Alf and Alf* that they have different beliefs, but this is no reason to abandon token identity theory. It is true to say of Alf that his skin inflammation is a sunburn, and its false to say of Alf* that his corresponding inflammation is a sunburn but this in no way tells against identifying Alf's sunburn with his skin inflammation, the very thing which need not have been a sunburn. And, he argues, an analogous story holds for talk of Alf's beliefs. So far, this seems to be just another form of the contingent content thesis, the kind of account opted for by Lewis. Just as the state designated by 'Alf's sunburn' might have failed to be a sunburn, so the state designated by

[14] Davidson 1987: 451–2.

'Alf's belief that cans are made of aluminum' might not have been that belief, but another. But then, a paragraph later, Davidson writes, 'If two mental events have different contents they are surely different mental events.'[15] And this seems to be an outright rejection of the contingent content approach—a difference in content makes for a different mental state, a different belief. On the sunburn analogy we are led to think of the very state designated as being capable of having a content different from the one it has in fact. But now we are told that a mental state or event has the content it has essentially—that thing can't have a different content than it has, and hang around.[16] What are we to say here? Is Davidson simply being inconsistent?

There is one heroic line available to Davidson. He can avoid inconsistency—and retain (i) the claim that Alf's belief that cans are made of aluminum is identical to Alf's neurological state N_i, (ii) the claim that Alf* lacks this belief, and (iii) the claim that beliefs have their contents essentially–by simply stipulating that N_i (the state of Alf identified with the belief that cans are made of aluminum) cannot be a state of Alf*. One can stipulate that the changes in the twin's environments, no matter how remote, rule out the possibility that they agree in their physiological states. Thus Alf and Alf* are not physiological twins at all. Davidson seems to suggest this strategy in the following remark:

> I accept Burge's premise; I think its denial not merely implausible but absurd. If two mental events have different contents they are surely different events. What I take Burge's and Putnam's imagined cases to show is that people who are in all relevant physical respects similar . . . can differ in what they mean or think . . . But of course there is *something* different about them, even in the physical world; their causal histories are different.[17]

And, according to Burge, Davidson has privately confirmed this reading of his remark.[18] But this stipulation of Davidson is highly implausible; it rules out as metaphysically incoherent commonplace talk of what would happen if this event—this reaction, this invasion, this neuronal firing—were to occur in context C_i rather than in context C_j.[19] This is surely too much. In addition, this stipulation—that states and events are individuated in such a fashion as to rule out the possibility of the twins instantiating the same states and events—makes the entire sunburn story irrelevant, and indeed undermines that very story. That story seemed relevant because it seemed as though we were to think of one and the same skin condition occurring in the two contexts, simply meriting different descriptions in each—there was no suggestion that we could not think of the

[15] Ibid. 152.

[16] Davidson is not here making the simple logical point that x is not identical to y if x and y differ in content (or any other characteristic). The point is rather one about identity across possible worlds; the mental event x, in Wi, is not to be identified with the mental event y, in Wj, if x and y have different contents. He was granting Burge this point.

[17] Ibid. 152. [18] See Burge 1993. [19] See Ibid.

subject having the same physiological skin state in the two contexts—because of a difference in causal histories.[20]

The genuine attraction of Davidson's position, and the appeal of examples such as that of the sunburn, arises not from any such implausible rejection of ordinary counterfactual reasoning, but rather from the fact that sunburn type examples seem to be plausible models for the case at hand, *at least so long as we ignore the fact that we are supposed to considering mental states*. The sunburn case is advanced as a model for our counterfactual treatment of states and events: different contextual elements provide for different description of one and the same state or event, and such differences pose no problem for token-identity theory. But in advancing the sunburn type case as a general model in just this way, we are simply ignoring the possibility of limitations imposed by the nature of the vocabulary and the nature of the states and events considered; we are not allowing for the fact that it is mental states that we have in mind, and that certain kinds of counterfactual suppositions might not be coherent for such states—e.g. the supposition that the state of believing that cans are made of aluminum might have the content that cans are made of twalum. The sunburn type cases provide us with an apparently coherent model of contingent identity but the attraction of this model derives totally from ignoring the fact that the states under discussion are mental states, propositional attitudes, etc.

This, I suggest, is where all these contingent content theories leave us; they legitimize talk of the mental, and of mental content, but they inevitably result in eliminating the mental from our ontology. It is of course possible to extract a different message from these considerations. As noted above, they can be read as simply telling against the coherence of the contingent content analysis of psychological talk; one can insist, as some have, that it is simply counterintuitive to talk of a psychological state as possibly having a different content than the content it has. But, while this is correct, as indeed I have argued above, it tells against theorists such as Davidson or Lewis only if we assume that they are committed to thinking of psychological expressions such as (a) as designating *psychological states*, and elements in their accounts suggest that this is not how they understand expressions such as (a). Let me leave this then as a dilemma for the contingent content theorist. *Such a theorist may understand psychological expressions such as (a) as designating psychological states such as belief, but then that theorist is committed to what Davidson himself calls the absurd result of thinking that the very state of believing that P could be the state of believing that Q. On the other hand, the contingent content theorist may understand expressions such as (a) as not designating psychological states at all, but rather as designating physiological or*

[20] In addition, I have been assuming in this chapter that the original externalist argument goes through up to (7), and in doing so assuming that the twins are physiological twins. Almost all participants to the debate are ready to concede as much, and indeed in his discussion of Burge and Putnam cases, Davidson does not explicitly rely on such an implausible strategy.

some such states. On this second option, the absurd position, denounced by Davidson, is not forced on one, but one is opting for an ontology devoid of mental states. The attraction of the contingent content position is, I suspect, the product of ambiguously shifting back and forward between these two interpretations.

II

I turn now to the second option, that of supposing the description 'Alf's belief that cans are made of aluminum' designates a state that has content, and that state has the content it has essentially. I argue that this construal of the outcome undermines the confident assumption of the vast majority of theorists that explanation in terms of belief, desire, etc. is a form of causal explanation.

Following the Second World War, Wittgensteinian and Rylean arguments convinced much of the philosophical world that explanations of behavior in terms of reasons for acting, beliefs, desires, etc. were not causal explanations.[21] But then in 1963 Donald Davidson's 'Actions, Reasons and Causes' appeared, and the tables were turned.[22] So persuasive were Davidson's arguments that today there is something near consensus that such explanations are causal in character, and this consensus is also reflected in cognitive theorizing (the representational states of the system supposedly cause its behavior), in externalist epistemologies of beliefs as being justified by the character of their causal histories, by the reliability of the causal mechanisms, etc.[23] While Davidson wrote extensively on psychological explanation and causation, the two papers of special significance to us are 'Actions, Reason and Causes' and 'Mental Events'.[24] In the first Davidson attempted: (i) to expose the weaknesses in the Wittgensteinian-inspired arguments against the causal interpretation of psychological explanation, and (ii) to argue that only the causal account was able to do justice to the fact that an agent may have good reasons for acting in a certain way, act that way, and those same good reasons not explain his action (he does it for other good reasons). In the second paper Davidson assumes the causal understanding of psychological explanation, and argues that that understanding requires us to suppose that the token psychological states appealed to in the explanation of an individual's behavior are identical to token physical states.

[21] See for example Winch 1958; Anscombe 1959; and Melden 1961.

[22] Davidson 1963.

[23] So we find in Lepore and Loewer 1989: 175 '[I]t was said that propositional attitude explanations are causal explanations, and that beliefs, intendings, imaginings and the like are not even candidates to be causes. . . . The work of Davidson, Armstrong, Putnam and Fodor (amongst others) has reversed what was once orthodoxy and it is now widely agreed that propositional attitude ascriptions describe states and episodes which enter into causal relations.' See also Fodor 1981, 1987; Cummins 1983; Dretske 1988.

[24] Davidson 1963, 1970; both reprinted in Davidson 1980.

Most externalists subscribe to the re-entrenched causal interpretation of psychological explanation, and for them this last argument poses a direct challenge. If externalism undermines any identification of the mental and the physical, even token identities, then if one is to advance such a thesis and retain the causal interpretation one must turn aside this argument of Davidson's. Davidson's argument for the claim that the causal interpretation requires identities relies crucially on the premiss that a singular causal claim—of the form '*a* caused *b*'—is true only if events *a* and *b* both admit of descriptions d_1 and d_2 such that events of kind d_1 are lawfully correlated, without exception, with events of kind d_2. 'The second principle is that where there is causality, there must be a law: events related as cause and effect fall under strict deterministic laws.'[25] And in responding to Davidson, Burge and other externalists have defended retaining causality, while rejecting token identities, by simply arguing against this principle.[26] They point to the fact that the special sciences are invariably fashioned using laws that admit of all kinds of exceptions, laws that require ceteris paribus clauses, laws that rest on idealizations, etc. This is where most of the recent debate on the tension between externalism and causation has focused, but in the remainder of this chapter I want to focus on very different issues, issues that arise in the earlier paper, 'Actions, Reasons and Causes'.

As mentioned above, there are two central components to Davidson's argument in 'Actions, Reasons and Causes': (1) he attempts to defeat the objections posed by the Wittgensteinians to the causal model, and (2) he argues that the causal model, unlike the Wittgensteinian model, is able to do justice to the intuitive difference between acting in accordance with good reasons and acting out of or because of good reasons.

I have argued against (2) elsewhere, and I will limit my discussion here to (1).[27] Wittgensteinians offered a variety of objections to the causal model: they argued that reasons—beliefs, desires, etc.—are not events, and so cannot be construed as causes.[28] They argued that many psychological explanations appeal to states (e.g. intentions in acting) which are not antecedent to the behavior, and so cannot be causes;[29] they argued that on the causal model one truly explains S's doing A because of beliefs and desires R, only if there is a law-like regularity of the form 'one does A if and only if one has beliefs and desires R', and such laws they argue are not to be found, etc.[30] Davidson discusses a number of these different objections, and finds them all without serious merit.[31] The central objection of the Wittgensteinians, however, the one that Davidson focuses on and the one I will focus on, is of the form: there is a conceptual linkage between the language used in explaining human action, talk of the agent's beliefs, desires, intentions, etc., and the language used to describe the action being explained,

[25] Davidson 1970: 208 (as reprinted in Davidson 1980). [26] See Burge 1993.
[27] Owens 1998. [28] See Melden 1961. [29] See Malcolm 1968.
[30] See Hart and Honore 1959. [31] Davidson 1963.

and this linkage bars our understanding of the explanation as causal. Taylor, for example, offers the following:

> [F] or this is part of what we mean by 'intending X', that, in the absence of interfering factors, it is followed by doing X. I could not be said to intend X if, even with no obstacles or other countervailing factors, I still didn't do it. Thus, my intention is not a causal antecedent of my behavior.[32]

The idea is that an explanation such as:

> C.L. (Conceptual linkage) A ran to the store because she believed it was close to five p.m., and that the store closed at five p.m., and she wanted to buy milk before it closed

cannot be causal in character because of the conceptual linkage between the language used in describing the agent's reasons for acting, and the action itself. Here again, the issues are by no means all clear, but it is easy to get a sense of the disquiet. One who does not know that A's wanting milk and A's thinking that she will get milk only if she runs to the store serves to explain A's running to the store, does not understand talk of belief and desire, does not understand what is involved in rational action, and action explanation. The very concepts employed in talk of intention, belief, and desire are concepts whose primary function is to provide for the interpretation and explanation of rational behavior. Psychological concepts are not concepts that have a life of their own independent of their role in explaining behavior, concepts which fortunately happen to be also useful in explaining behavior. No, the linkage is much more intimate than this. The question as to what an agent was doing, for example, is decided once we know what he was aiming at, what his goals were. We can reject the claim that an agent was doing such and such by merely citing his intentions. How could one possibly do this if intentions are merely causal antecedents of behavior? The linkage between reasons and the actions they serve to explain is so intimate that we cannot conceive of its being challenged by empirical findings in ways in which causal linkages might be challenged. We can, of course, imagine individuals for whom such and such beliefs and desires don't engender behavior of the kind they engender in most of us, but the theory allows for such cases, instances that involve irrationalities, interferences, etc. What we can't imagine is the empirical falsification of explanatory generalizations such as 'individuals who want P, see no reason not to pursue P, etc. will tend to pursue P'.

This all remains somewhat vague, and the crucial notions all call out for further clarification and analysis. We certainly do use causal notions in talking of behavior; we say such things as 'What caused you to say that?' On the other hand when asked why we did what we did, we do not say that our actions were caused by such and such beliefs and desires. Or again, it is commonplace to think of there being different kinds of causation, event causation, agent causation,

[32] Taylor 1964: 33.

etc. and if these distinctions have any merit then they too must figure in the account. The notion of 'conceptual connection' is likewise unclear. How is one to distinguish between conceptual connections in the intended sense and mere truisms, entrenched, commonplace knowledge of the ways in which beliefs and desires happen to shape rational action? Despite these and other obscurities, the Wittgensteinian argument that the psychological and action vocabularies are both part of a whole, neither being understandable in the absence of the other, is a serious challenge to the familiar causal picture. Psychological explanation has all the appearances of being something very different from ordinary causal explanation.

Davidson's response is seemingly decisive, and appears to demolish this objection in one blow. He does not attempt any further analysis of the psychological concepts employed in action explanation in an effort to show that the conceptual linkages do not obtain. No, he dismisses the entire strategy of the Wittgensteinian, the very idea of appealing to conceptual linkage: the character of the linguistic terms used in fashioning 'A' and 'B' in the singular causal claim 'A caused B' cannot, he argues, be used to show that the claim is not a genuine (possibly true) causal claim. This is so even if 'A' might be said to entail 'B'. This is evident from the simple fact that we can truly characterize *any* contingent event B as being caused by 'the cause of B'. Here we have the most intimate conceptual linkage between the expression used to designate the cause and the expression used to designate the effect, and even though such a characterization of cause and effect does not provide for a very interesting explanation of B, it is, nevertheless, a true causal claim.

But while this kind of causal claim might be true, it is of course completely unexplanatory, and one does not want to suppose that 'explanations' of this empty sort provide a model for understanding those explanations in which we seem to genuinely explain an agent's behavior by citing the reasons with which she acted. But putting this aside, consider again the claim: 'The cause of B caused B.' Why does this not tell against the Humean requirement that causal linkage be contingent? The Davidsonian answer is clear. The causal linkage is a linkage between events, not between descriptions and not between events under descriptions. The claim '*a* caused *b*' is extensional, true if the event designated by *a* caused the event designated by *b*, and its truth or falsity does not hang on the way in which the two events are characterized. Events may be contingently linked as cause and effect, even though there are descriptions of them that are conceptually linked.

All this is well and good, *But what really lies behind the intuition that these descriptions a and b do not matter*—that it is the events themselves that are said to be related—*is the further supposition that the events are open to many descriptions, including descriptions that involve no problematic conceptual linkages*. We would not be willing to suppose that events *e* and *g* are contingently linked if they were such that every description 'A' and 'B' of *e* and *g*

were conceptually linked. What makes for the intuition that the descriptions employed in singular causal claim don't really matter is the ordinarily reliable intuition that other description's of the events are available, and these descriptions are such that (i) there is no deep conceptual linkage between them, and (ii) under these descriptions the events are seen as satisfying law-like generalizations.

This, of course, is just the line taken by Davidson in his 'Mental Events'. In this paper his concern was to establish the consistency of the following three claims: (a) Explanations of behavior by citing the agent's reasons in acting are causal explanations, (b) Causal claims of the form 'A causes B' are true only if the events designated by 'A' and 'B' can be subsumed under (exception-less) law-like generalizations, and (c) There are no such (exception-less) psychological laws. These claims can be seen to be consistent if we suppose that the states and events that constitute the agent's reasons are physical states and events, and that they are subsumed under exception-less laws when they are characterized in the appropriate physical vocabulary. In this paper Davidson's concern was not with defending the causal thesis, but even so, we see here the reliance on there being available other, non-psychological vocabularies in which to characterize the psychological states appealed to in reason explanations of behavior.

Davidson's response to the Wittgensteinian challenge then requires us to suppose that talk of reasons, beliefs, desires, etc. is talk of states and events which can be characterized in this non-problematic vocabulary. What kind of vocabulary? In 'Mental Events' Davidson suggests physical vocabulary, presumably the vocabulary of physiology. We, however, need not suppose that the vocabulary is at such a low level. It might employ concepts drawn from computer science, cognitive science, cybernetics, etc. This Davidsonian response, however, does presuppose that the states characterized using the problematic intentional psychological vocabulary can also be characterized and specified using non-intentional terms, terms which then serve to express the law-like linkages between these antecedent physical states and behavior; i.e. they can be characterized using terms drawn from physiology, neurochemistry, computer science, etc. To suppose that they can't be characterized in a non-intentional vocabulary is to suppose that we can't appeal to them in explanations without employing the problematic vocabulary—one that involves conceptual linkages between the events being explained and the explanans.

But this strategy is simply not available to the theorist who holds a description such as 'Alf's belief that cans are made of aluminum' designates a state that has content, and that state has the content it has essentially. Once we make this supposition we have to allow that this description designates a state that has a certain content P essentially, a content that is not common to Alf and Alf*. But by supposition we have it that Alf and Alf* do not differ in states under

non-intentional descriptions.[33] Hence the state designated by the intentional characterization is not a state characterized by a non-intentional description. If it admitted of a non-intentional characterization it would occur in Alf*, and have the essential content that-P. Hence, one cannot follow the second option, taking the content as essential, and respond to the Wittgensteinian challenge in a Davidsonian fashion: one cannot be an essentialist about content, and simply ignore the problematic character of the intentional vocabulary, by assuming that the events designated admit of further non-intentional, non-problematic descriptions.

This issue has been largely ignored in recent discussions of twin examples, and my primary concern is this last part of this chapter is to insist that these once dismissed Wittgensteinian considerations must be addressed today by those who think of the mental terms as designating mental states, states that have the content they have of necessity. One cannot simply take this position, and argue that the tension between the causal construal of psychological explanation and this brand of externalism can be mitigated by simply allowing that the laws need not be exception-less. ***No, making this point alone leaves untouched the conceptual linkage problems posed by the Wittgensteinians, and this is a problem that cannot now be dismissed in the way Davidson suggested.***

CONCLUSION

In *Knowledge and its Limits,* Timothy Williamson argued that externalism forces us to abandon a certain deeply loved picture—a picture of the agent as having a variety of mental states, beliefs, desires, etc., states of the mind. Knowledge on the other hand was denied this status—knowledge required more than one's mind being in a certain state—it also placed demands on the world: that it correspond. This picture then seduced us into thinking that knowledge is best construed as a combination of a true internal mental state such as belief and other factors. And, Williamson urges, with externalism in hand we no longer have reason to discriminate against knowledge in this way, we no longer have reason to deny knowledge its true status as a genuine state of mind. In this chapter I have argued for a very different reading of the externalism arguments. True, belief, like knowledge, is not internal, and hence this supposed difference cannot be exploited to motivate the traditional picture. But, as I read the arguments, they serve not to support a conception of knowledge as a state of mind, but to weaken the traditional view of belief as *a state of mind that is causally explanatory of behavior*. For the realist, the expressions 'Alf's belief that the cans are made

[33] Once again, one can avoid this objection by simply denying the very possibility of the twin example, by denying the possibility of Alf having the same token physical states in the counterfactual contexts. And, once again, I shall ignore this strategy.

of aluminum' and 'Alf*s belief that cans are made of twalum' designate states that have contents. The states designated have these contents essentially or non-essentially. If they have them non-essentially then the states designated are simply not psychological states, not beliefs, not desires. If they have such contents essentially then we have reason to think that explanations which cite such states are not causal explanations; we have reason to think that such states cannot play the causal role traditional theory assigns them. Rather than lending support to a conception of knowledge as a mental state, externalism tells against the entrenched philosophical understanding of belief as a robust, causally effective mental state.

REFERENCES

Anscombe, G. E. M. 1959: *Intention* (Oxford: Basil Blackwell).

Boghossian, P. 1989: 'Content and Self-Knowledge.' *Philosophical Topics* 17: 5–26.

Brand, M. 1984: *Intending and Acting: Towards a Naturalized Action Theory* (Cambridge, Mass.: MIT Press).

Brueckner, A. 1986: 'Brains in a Vat.' *Journal of Philosophy* 83: 148–67.

——— 1990: 'Scepticism about Knowledge of Content.' *Mind* 99: 447–52.

——— 1994: 'Knowledge of Content and Knowledge of the World.' *Philosophical Review* 103: 327–43.

Burge, T. 1979: 'Individualism and the Mental.' *Midwest Studies in Philosophy* 4: 73–121.

——— 1989: 'Individualism and Causation in Psychology.' *Pacific Philosophical Quarterly* 70: 303–22.

——— 1993: 'Mind-Body Causation and Explanatory Practice.' In Heil and Mele 1993: 97–120.

Cummins. R. 1983: *The Nature of Psychological Explanation* (Cambridge, Mass.: MIT Press).

Davidson, D. 1963: 'Actions, Reasons and Causes.' *Journal of Philosophy* 60: 685–700. Reprinted in Davidson 1980: 3–20.

——— 1970: 'Mental Events.' In Lawrence Foster and J. W. Swanson (eds.), *Experience and Theory* (London: Duckworth).

——— 1980: *Essays on Actions and Events* (Oxford: Clarendon Press).

——— 1987: 'Knowing one's Own Mind.' *Proceedings and Addresses of the American Philosophical Association* 60: 441–58.

——— 1990: 'The Structure and Content of Truth.' *Journal of Philosophy* 87: 279–328.

——— 1991: 'Epistemology Externalized.' *Dialectica* 45: 191–205.

Dretske, F. 1988: *Explaining Behavior* (Cambridge, Mass.: MIT Press).

Falvey, K. and Owens, J. 1994: 'Externalism, Self-Knowledge and Skepticism.' *Philosophical Review* 103: 107–37.

Fodor, J. 1981: 'Propositional Attitudes.' In *Representations* (Cambridge, Mass.: MIT Press).

——— 1987: *Psychosemantics* (Cambridge, Mass.: MIT Press).

Goldman, A. 1986: *Epistemology and Cognition* (Cambridge, Mass.: Harvard University Press).

Hart, H. L. A. and Honore, A. 1959: *Causation in the Law* (Oxford: Clarendon Press).
Heil, J. and Mele, A. (eds.) 1993: *Mental Causation* (Oxford: Clarendon Press).
Kripke, S. 1980: *Naming and Necessity* (Oxford: Blackwell).
Lepore, E. and Loewer, B. 1989: 'More on Making Mind Matter.' *Philosophical Topics* 18: 175–91.
Lewis, D. 1966: 'An Argument for the Identity Theory.' *Journal of Philosophy* 63: 17–25.
____ 1969: 'Review of Art, Mind, and Religion.' *Journal of Philosophy* 66: 22–7.
____ 1970: 'How to Define Theoretical Terms.' *Journal of Philosophy* 67: 427–46.
____ 1980: 'Mad Pain and Martian Pain.' In N. Block (ed.), *Readings in Philosophy of Psychology* (Cambridge, Mass.: Harvard University Press), i. 216–22.
Malcolm, N. 1968: 'The Conceivability of Mechanism.' *Philosophical Review* 77: 45–72.
McKinsey, M. 1991: 'Anti-Individualism and Privileged Access.' *Analysis* 51: 9–16.
____ 1994: 'Accepting the Consequences of Anti-Individualism.' *Analysis* 54: 124–8.
Melden, A. I. 1961: *Free Action* (London: Routledge and Kegan Paul).
Owens, J. 1986: 'The Failure of Lewis's Functionalism.' *Philosophical Quarterly* 36: 159–83.
____ 1987: 'In Defense of a Different Doppelganger.' *Philosophical Review* 96: 521–54.
____ 1993: 'Content, Causation, and Psychophysical Supervenience.' *Philosophy of Science* 60: 242–61.
____ 1998: 'Psychological Explanation and Causal Deviancy.' *Synthese*: 143–69.
____ 2004: 'Anti-individualism, Self-Knowledge and Character.' In M. Hahn and B. Ramberg (eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge* (Cambridge, Mass.: MIT Press).
Putnam, H. 1975a: *Mind, Language and Reality* (Cambridge: Cambridge University Press).
____ 1975b: 'The Nature of Mental States.' In Putnam 1975a: 429–40.
____ 1981: *Reason, Truth, and History* (Cambridge: Cambridge University Press).
Smart, J. C. 1959: 'Sensations and Brain Processes.' *Philosophical Review* 68: 141–56.
Taylor, C. 1964: *The Explanation of Behavior* (London: Routledge and Kegan Paul).
Williamson, T. 2000: *Knowledge and its Limits* (Oxford: Oxford University Press).
Winch, P. 1958: *The Idea of a Social Science* (London: Routledge and Kegan Paul).
Wittgenstein, L. 1958: *Philosophical Investigations*, 3rd edn. (Oxford: Blackwell).

Further Reading

Adams, F. 1991: "Audi on Structural Justification." *Journal of Philosophical Research* 16: 493–8.

——— and Aizawa, K. 2001: "The Bounds of Cognition." *Philosophical Psychology* 14.1: 43–64.

——— and Dietrich, L. 2004: "Swampman's Revenge: Squabbles among the Representationalists." *Philosophical Psychology* 17.3: 323–40.

——— and Kline, D. 1987: "Nomic Reliabilism: Weak Reliability is not Enough." *Southern Journal of Philosophy* 25: 433–43.

Adler, J. 2002: "Is the Generality Problem Too General?" *Philosophy and Phenomenological Research* 65.1: 87–97.

Almeder, R. 1995a: "Externalism and Justification." *Philosophia* 24.3/4: 465–9.

——— 1995b: "Dretske's Dreadful Question." *Philosophia* 24.3/4: 449–57.

——— and Hoff, F. 1989: "Reliability and Goldman's Theory of Justification." *Philosophia* 19.2/3: 165–87.

Alston, W. 1986: "Internalism and Externalism in Epistemology." *Philosophical Topics* 14: 179–221.

——— 1988: "An Internalist Externalism." *Synthese* 74: 265–83.

——— 1989: *Epistemic Justification: Essays in the Theory of Knowledge* (Ithaca, NY: Cornell University Press).

——— 1990: "Externalist Theories of Perception." *Philosophy and Phenomenological Research* 50: 73–97.

Amico, R. 2003: "Is a Fully General Theory of Knowledge Possible?" *Southern Journal of Philosophy* 41.3: 307–22.

Antony, L. 2002: "Embodiment and Epistemology." In P. Moser (ed.), *The Oxford Handbook of Epistemology* (Oxford: Oxford University Press), 463–78.

Armstrong, D. 1973: *Belief, Truth, and Knowledge* (Cambridge: Cambridge University Press).

Audi, R. 1989: "Internalism and Externalism in Moral Epistemology." *Logos* 10: 13–37.

——— 1991: "Structural Justification." *Journal of Philosophical Research* 16: 473–92.

——— 1992: *The Structure of Justification* (Cambridge: Cambridge University Press).

Axtell, G. 1997: "Recent Work on Virtue Epistemology." *American Philosophical Quarterly* 34.1: 1–26.

Bach, K. 1985: "A Rationale for Reliabilism." *Monist* 68: 246–63.

——— 1988: "Burge's New Thought Experiment: Back to the Drawing Room." *Journal of Philosophy* 85: 88–97.

——— and Elugardo, R. 2003: "Conceptual Minimalism and Anti-Individualism: A Reply to Goldberg." *Noûs* 37.1: 151–60.

Barber, A. (ed.) 2001: *Epistemology of Language* (Oxford: Oxford University Press).

Bar-On, D. 1990: "Scepticism: The External World and Meaning." *Philosophical Studies* 60: 207–31.

Bar-On, D. 2004: "Externalism and Self-Knowledge: Content, Use, and Expression." *Noûs* 38.3: 430–55.

Baumann, P. 1998: "Can Reliabilists Believe in Subjective Probability?" *Philosophical Quarterly* 48.191: 199–200.

Bealer, G. 1998: "A Theory of Concepts and Concepts Possession." In E. Villanueva (ed.), *Concepts* (*Philosophical Issues* 9): 261–301.

Becker, K. 2002: "Individualism and Self-Knowledge: *Tu Quoque*." *American Philosophical Quarterly* 39.3: 289–95.

____ 2004: "Knowing and Possessing Knowledge." *American Philosophical Quarterly* 41.1: 21–36.

Beebe, J. 2004a: "Reliabilism, Truetemp and New Perceptual Faculties." *Synthese* 140.3: 307–29.

____ 2004b: "The Generality Problem, Statistical Relevance and the Tri-Level Hypothesis." *Noûs* 38.1: 177–95.

Bender, J. 1988: "The Ins and Outs of Metaknowledge." *Analysis* 48: 167–76.

Ben Menahem, Y. 2005a: "Putnam on Skepticism." In Ben Menahem 2005b: 125–55.

____ (ed.) 2005b: *Hilary Putnam (Contemporary Philosophy in Focus)*. (Cambridge: Cambridge University Press).

Berg, J. 1998: "First-Person Authority, Externalism, and Wh-Knowledge." *Dialectica* 52.1: 41–4.

Bergmann, M. 1997: "Internalism, Externalism and the No Defeater Condition." *Synthese* 110.3: 399–417.

____ 2000: "Externalism and Skepticism." *Philosophical Review* 109.2: 159–94.

____ 2004: "Externalist Justification without Reliability." *Philosophical Perspectives* 14: 35–60.

____ 2005: "Defeaters and Higher Level Requirements." *Philosophical Quarterly* 55.220: 419–36.

Bernecker, S. 1996a: "Davidson on First-Person Authority and Externalism." *Inquiry* 39.1: 121–39.

____ 1996b: "Externalism and the Attitudinal Component of Self-Knowledge." *Noûs* 30.2: 262–75.

____ 2000: "Knowing the World by Knowing one's Mind." *Synthese* 123.1: 1–34.

____ 2004a: "Memory and Externalism." *Philosophy and Phenomenological Research* 69.3: 605–32.

____ 2004b: "Davidson on First-Person Authority and Externalism." *Inquiry* 39.1: 121–39.

Bilgrami, A. 1987: "An Externalist Account of Psychological Content." *Philosophical Topics* 15: 191–226.

____ 1989: "Realism without Internalism: A Critique of Searle on Intentionality." *Journal of Philosophy* 86: 57–72.

____ 1992: "Can Externalism Be Reconciled with Self-Knowledge?" *Philosophical Topics* 20.1: 233–67.

____ 1998: "Replies." *Philosophy and Phenomenological Research* 58.3: 635–47.

____ 2003: "A Trilemma for Redeployment." *Philosophical Perspectives* 13: 22–30.

Biro, J. 1992: "In Defence of Social Content." *Philosophical Studies* 67.3: 277–93.

——1995: "Dretske on Phenomenal Externalism." In E. Villanueva (ed.), *Perception: Philosophical Issues* 7: 171–8.

Bishop, M. 2000: "In Praise of Epistemic Irresponsibility: How Lazy and Ignorant Can You Be?" *Synthese* 122.1–2: 179–208.

Block, N. 2004: "Mental Paint." In Hahn and Ramberg 2004: 165–200.

Boghossian, P. 1989: "Content and Self-Knowledge." *Philosophical Topics* 17.1: 5–26.

——1992a: "Externalism and Inference." In E. Villanueva (ed.), *Philosophical Issues* 2: *Rationality in Epistemology*: 11–28.

——1992b: "Reply to Schiffer." In E. Villanueva (ed.), *Philosophical Issues* 2: *Rationality in Epistemology*, 39–42.

——1994: "The Transparency of Mental Content." *Philosophical Perspectives* 8: 33–50.

——1997: "What the Externalist Can Know A Priori." *Proceedings of the Aristotelian Society* 97: 161–75.

Bonjour, L. 1980: "Externalist Theories of Empirical Knowledge." *Midwest Studies in Philosophy* 5: 53–74.

——1985: *The Structure of Empirical Knowledge* (Cambridge, Mass.: Harvard University Press).

——1987: "Nozick, Externalism, and Skepticism." In S. Luper Foy (ed.), *The Possibility of Knowledge* (Savage, Md.: Rowman and Littlefield), 297–313.

——1988: 'Reply to Moser's "Internalism and Coherentism."' *Analysis* 48: 164–5.

——2002: "Internalism and Externalism." In Moser 2002: 234–63.

——2006: "Kornblith on Knowledge and Epistemology." *Philosophical Studies* 127.2: 317–35.

——and Sosa, E. 2003: *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues* (Oxford: Blackwell).

Bostrom, N. 2005: "The Simulation Argument: Reply to Weatherson." *Philosophical Quarterly* 55.218: 90–7.

Bradshaw, D. 1992: "On the Need for a Metaphysics of Justification." *Metaphilosophy* 23.1/2: 90–106.

Brandom, R. 1995a: "Knowledge and the Social Articulation of the Space of Reasons." *Philosophy and Phenomenological Research* 55.4: 895–908.

——1995b: "Perception and Rational Constraint: McDowell's *Mind and World*." In E. Villanueva (ed.), *Perception: Philosophical Issues* 7: 241–59.

——1998a: "Perception and Rational Constraint." *Philosophy and Phenomenological Research* 63.2: 369–74.

——1998b: "Insights and Blindspots of Reliabilism." *The Monist* 81.3: 371–92.

Brewer, B. 1996: "Internalism and Perceptual Knowledge." *European Journal of Philosophy* 4.3: 259–75.

——1997: "Foundations of Perceptual Knowledge." *American Philosophical Quarterly* 34.1: 41–55.

——1999a: "Self-Knowledge and Externalism." In R. Cobb-Stevens (ed.), *The Proceedings of the Twentieth World Congress of Philosophy*, 5: *Epistemology* (Bowling Green, Oh.: Philosophy Documentation Center), 39–47.

——1999b: *Perception and Reason* (Oxford: Oxford University Press).

Brewer, T. 2002: "The Real Problem with Internalism about Reasons." *Canadian Journal of Philosophy* 32.4: 443–74.

Brogaard, B. 2003: "Adhoccery in Epistemology." *Philosophical Papers* 32.1: 65–82.
Brown, J. 1995: "The Incompatibility of Anti-individualism and Privileged Access." *Analysis* 55.3: 149–56.
____ 1999: "Boghossian on Externalism and Privileged Access." *Analysis* 59.1: 52–9.
____ 2000a: "Reliabilism, Knowledge, and Mental Content." *Proceedings of the Aristotelian Society* 100: 115–35.
____ 2000b: "Against Temporal Externalism." *Analysis* 60.2: 178–88.
____ 2001a: "Anti-individualism and Agnosticism." *Analysis* 61.3: 213–24.
____ 2001b: "Externalism and the Fregean Tradition." In Barber 2001: 431–55.
____ 2004a: *Anti-individualism and Knowledge* (Cambridge, Mass.: MIT Press).
____ 2004b: "Wright on Transmission Failure." *Analysis* 64.1: 57–67.
Brueckner, A. 1988: "Problems with Internalist Coherentism." *Philosophical Studies* 54: 153–60.
____ 1990: "Scepticism about Knowledge of Content." *Mind* 99.395: 447–51.
____ 1992a: "If I Am a Brain in a Vat, Then I Am Not a Brain in a Vat." *Mind* 101.401: 123–8.
____ 1992b: "Semantic Answers to Skepticism." *Pacific Philosophical Quarterly* 73.3: 200–19.
____ 1993: "Skepticism and Externalism." *Philosophia* 22.1/2: 169–71.
____ 1994: "Knowledge of Content and Knowledge of the World." *Philosophical Review* 103.2: 327–43.
____ 1995: "The Characteristic Thesis of Anti-Individualism." *Analysis* 55.3: 146–8.
____ 1996: "Deontologism and Internalism in Epistemology." *Noûs* 30.4: 527–36.
____ 1997: "Externalism and Memory." *Pacific Philosophical Quarterly* 78.1: 1–12.
____ 1998: "Content Externalism and A Priori Knowledge." *Protosociology* 11: 149–59.
____ 2000a: "Ambiguity and Knowledge of Content." *Analysis* 60.3: 257–60.
____ 2000b: "Externalism and the *A Prioricity* of Self-Knowledge." *Analysis* 60.1: 132–36.
____ 2000c: "Transcendental Arguments from Content Externalism." In Stern 2000: 229–50.
____ 2001a: "Defending Burge's Thought Experiment." *Erkenntnis* 55.3: 387–91.
____ 2001b: "A Priori Knowledge of the World Not Easily Available." *Philosophical Studies* 104.1: 109–14.
____ 2002a: "The Consistency of Content Externalism and Justification Internalism." *Australasian Journal of Philosophy* 80.4: 512–15.
____ 2002b: "Anti-individualism and Analyticity." *Analysis* 62.1: 87–91.
____ 2003a: "Self-Knowledge via Inner Observation of External Objects?" *Australasian Journal of Philosophy* 81.1: 118–22.
____ 2003b: "The Coherence of Scepticism about Self-Knowledge." *Analysis* 63.1: 41–8.
____ 2003c: "Contents Just Aren't in the Head." *Erkenntnis* 58.1: 1–6.
____ 2004: "Brewer on the McKinsey Problem." *Analysis* 64.1: 41–3.
____ 2005: "Noordhof on McKinsey-Brown." *Analysis* 65.1: 86–8.
Bruns, M. and Soldati, G. 1994: "Object Dependent and Property Dependent Contents." *Dialectica* 48.3/4: 185–208.
Buekens, F. 1994: "Externalism, Content, and Causal Histories." *Dialectica* 48.3/4: 267–86.

Burge, T. 1979: "Individualism and the Mental." *Midwest Studies in Philosophy* 4: 73–121.
——1982: "Other Bodies." In A. Woodfield (ed.), *Thought and Object* (Oxford: Clarendon Press), 97–120.
——1986: "Intellectual Norms and the Foundations of Mind." *Journal of Philosophy* 83: 697–720.
——2003a: "Perceptual Entitlement." *Philosophy and Phenomenological Research* 67.3: 503–48.
——2003b: "Social Anti-individualism, Objective Reference." *Philosophy and Phenomenological Research* 67.3: 682–90.
Butler, K. 1996: "Individualism and Marr's Computational Theory of Vision." *Mind and Language* 11.4: 313–37.
——1997: "Externalism, Internalism, and Knowledge of Content." *Philosophy and Phenomenological Research* 57.4: 773–800.
——2000: "Problems for Semantic Externalism and A Priori Refutations of Skeptical Arguments." *Dialectica* 54.1: 29–49.
Cappelen, H. and Winblad, D. 1999: "Reference Externalized and the Role of Intuitions in Semantic Theory." *American Philosophical Quarterly* 36.4: 337–50.
Carruthers, P. 1988: "More Faith than Hope: Russellian Thoughts Attacked." *Analysis* 48: 91–6.
Cassam, Q. (ed.) 1994: *Self-Knowledge* (Oxford: Oxford University Press).
Casullo, A. 1988: "Revisability, Reliabilism, and A Priori Knowledge." *Philosophy and Phenomenological Research* 49: 187–213.
Chase, J. 2001: "Is Externalism about Content Inconsistent with Internalism about Justification?" *Australasian Journal of Philosophy* 79: 227–46.
——2004: "Indicator Reliabilism." *Philosophy and Phenomenological Research* 69.1: 115–37.
Chisholm, R. 1988: "The Indispensability of Internal Justification." *Synthese* 74: 285–96.
Chomsky, N. 2004: "Internalist Explorations." In Hahn and Ramberg 2004: 259–88.
Clark, A. and Chalmers, D. 1998: "The Extended Mind." *Analysis* 58.1: 7–19.
Clarke, D. 1990: 'Two Uses of "Know".' *Analysis* 50.3: 188–90.
Clarke, M. 1987: "Reliability and Two Kinds of Epistemic Justification." In N. Garver (ed.), *Naturalism and Rationality* (Buffalo: Prometheus), 159–70.
——1990: "Epistemic Norms and Evolutionary Success." *Synthese* 85: 231–44.
Cohen, S. 2003: "Greco's Agent Reliabilism." *Philosophy and Phenomenological Research* 66.2: 437–43.
Comesana, J. 2002: "The Diagonal and the Demon." *Philosophical Studies* 110.3: 249–66.
Conee, E. 1988: "The Basic Nature of Epistemic Justification." *The Monist* 71: 389–404.
——2002: "Innocuous Infallibility." *Philosophy and Phenomenological Research* 64.2: 406–8.
——2004: "Externalism, Internalism, and Skepticism." *Philosophical Perspectives* 14: 78–90.
——and Feldman, R. 1998: "The Generality Problem for Reliabilism." *Philosophical Studies* 89.1: 1–29
—— ——2004: *Evidentialism: Essays in Epistemology* (Oxford: Oxford University Press).

Conn, C. 2001: "Chisholm, Internalism, and Knowing That One Knows." *American Philosophical Quarterly* 38.4: 333–47.

Corbi, J. 1998: "A Challenge to Boghossian's Incompatibilist Argument." In E. Villanueva (ed.), *Concepts: Philosophical Issues* 9: 231–42.

Craig, E. 1986: "Privacy and Rule Following." In J. Butterfield (ed.), *Language, Mind, and Logic* (Cambridge: Cambridge University Press), 169–86.

Crane, T. 1991: "All The Difference in the World." *Philosophical Quarterly* 41: 1–25.

Crawford, D. 1991: "On Having Reasons for Perceptual Beliefs: A Sellarsian Perspective." *Journal of Philosophical Research* 16: 107–23.

Dancy, J. 1992: "Externalism for Internalists." In E. Villanueva (ed.), *Philosophical Issues* 2: *Rationality in Epistemology*, 93–114.

David, M. 1991: "Neither Mentioning "Brains in a Vat' nor Mentioning Brains in a Vat Will Prove that We Are Not Brains in a Vat." *Philosophy and Phenomenological Research* 51.4: 891–6.

Davidson, D. 1984: *Inquiries into Truth and Interpretation* (Oxford: Oxford University Press).

——1987: "Knowing one's Own Mind." *Proceedings and Addresses of the American Philosophical Association* 60.3: 441–58.

——1989: "The Conditions of Thought." In J. Brandl and W. Gombocz (eds.), *The Mind of Donald Davidson* (Amsterdam: Rodopi), 193–200.

——1991: "Three Varieties of Knowledge." In A. Griffiths (ed.), *A. J. Ayer: Memorial Essays* (Cambridge: Cambridge University Press), 153–66.

——1999: "Reply to Tyler Burge." In L. Hahn (ed.), *The Philosophy of Donald Davidson* (Chicago: Open Court Press), 251–4.

——2001: *Subject, Intersubjective, Objective* (Oxford: Oxford University Press).

——2003: "Quine's Externalism." *Grazer Philosophische Studien* 66: 281–97.

Davies, M. 1986: "Individualism and Supervenience: Externality, Psychological Explanation, and Narrow Content." *Proceedings of the Aristotelian Society* suppl. vol. 60: 263–83.

——1991: "Individualism and Perceptual Content." *Mind* 100.399: 461–84.

——1992: "Perceptual Content and Local Supervenience." *Proceedings of the Aristotelian Society* 92: 21–45.

——1993: "Aims and Claims of Externalist Arguments." In E. Villanueva (ed.), *Philosophical Issues* 4: *Naturalism and Normativity*, 227–49.

——1997: "Meaning and Semantic Knowledge." *Proceedings of the Aristotelian Society* suppl. vol. 71: 177–209.

——2004: "Epistemic Entitlement: Epistemic Entitlement, Warrant Transmission and Easy Knowledge." *Proceedings of the Aristotelian Society* suppl. vol. 78: 213–45.

De Gaynesford, M. 1996: "How Wrong Can One Be?" *Proceedings of the Aristotelian Society* 96: 387–94.

Devitt, M. 2001: "A Shocking Idea about Meaning." *Revue internationale de philosophie* 55.218: 471–94.

DeVries, W. 1990: "Burgeoning Skepticism." *Erkenntnis* 33: 141–64.

——1996: "Experience and the Swamp Creature." *Philosophical Studies* 82.1: 55–80.

Dretske, F. 1995a: "Reply to Commentators." In E. Villanueva (ed.), *Perception: Philosophical Issues* 7: 179–83.

——— 1995b: "Phenomenal Externalism." In E. Villanueva (ed.), *Perception Philosophical Issues* 7: 143–58.
——— 2004: "Externalism and Modest Contextualism." *Erkenntnis* 61.2–3: 173–86.
Ebbs, G. 1996: "Can We Take our Words at Face Value?" *Philosophy and Phenomenological Research* 56.3: 499–530.
Edwards, S. 1995: "Externalism in the Philosophy of Mind." *Australasian Journal of Philosophy* 73.4: 637–9.
Egan, F. 1992: "Individualism, Computation and Perceptual Content." *Mind* 101.403: 443–59.
——— 1999: "In Defense of Narrow-Mindedness." *Mind and Language* 14.2: 177–94.
Eilan, N. 1995: "The First Person Perspective." *Proceedings of the Aristotelian Society* 95: 51–66.
Elugardo, R. 1993: "Burge on Content." *Philosophy and Phenomenological Research* 53.2: 367–84.
Engel, M. 1992: "Is Epistemic Luck Compatible with Knowledge?" *Southern Journal of Philosophy* 30.2: 59–75.
Evans, G. 1982: *The Varieties of Reference* (Oxford: Clarendon Press).
Fales, E. 2004: "Proper Basicality." *Philosophy and Phenomenological Research* 68.2: 373–83.
Falvey, K. 2000: "The Compatibility of Anti-individualism and Privileged Access." *Analysis* 60.1: 137–42.
——— and Owens, J. 1994: "Externalism, Self-Knowledge, and Skepticism." *Philosophical Review* 103.1: 107–37.
Farkas, K. 2003: "What Is Externalism?" *Philosophical Studies* 112.3: 187–208.
Feldman, R. 1985: "Schmitt on Reliability, Objectivity, and Justification." *Australasian Journal of Philosophy* 63: 354–60.
——— 1993: "Proper Functionalism." *Noûs* 27.1: 34–50.
——— 1999: "Kvanvig on Externalism and Epistemology Worth Doing." *Southern Journal of Philosophy* 38 (supp.): 43–50.
——— 2003: "Chisholm's Internalism and Its Consequences." *Metaphilosophy* 34.5: 603–20.
——— 2004: "Foundational Justification." In J. Greco (ed.), *Ernest Sosa and his Critics* (Oxford: Basil Blackwell), 42–58.
——— 2005: "Justification Is Internal." In M. Steup and E. Sosa (eds.), *Contemporary Debates in Epistemology* (Oxford Blackwell), 270–84.
——— and Conee, E. 2001: "Internalism Defended." *American Philosophical Quarterly* 38.1: 1–18.
——— ——— 2002: "Typing Problems." *Philosophy and Phenomenological Research* 65.1: 98–105.
Fernandez, J. 2004: "Externalism and Self-Knowledge: A Puzzle in Two Dimensions." *European Journal of Philosophy* 12.1: 17–37.
Floyd, J. 2005: "Putnam's 'The Meaning of "Meaning" ': Externalism in Historical Context." In Ben Menahem 2005b: 17–52.
Fodor, J. 1980: "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology." *Behavioral and Brain Sciences* 3.1: 63–72.
——— 1987: *Psychosemantics* (Cambridge, Mass.: MIT Press).

Fodor, J. 1986: "Individualism and Supervenience." *Proceedings of the Aristotelian Society* suppl. vol. 60: 235–62.
——1994: *The Elm and the Expert* (Cambridge, Mass.: MIT Press).
Fogelin, R. 2000: "Contextualism and Externalism: Trading in One Form of Skepticism for Another." In E. Sosa (ed.), *Skepticism: Philosophical Issues* 10: 43–57.
Foley, R. 1990: "Skepticism and Rationality." In Ross and Roth 1990: 69–81.
——1993: "What am I to Believe?" In Kornblith 2001: 163–79.
Francescotti, R. 1991: "Externalism and Marr's Theory of Vision." *British Journal for the Philosophy of Science* 42: 227–38.
Frappoli, M. and Romero, E. (eds.) 2000: *Meaning, Basic Self-Knowledge, and Mind* (Palo Alto, Calif.: CSLI Press).
Fumerton, R. 1987: "Nozick's Epistemology." In S. Luper-Foy (ed.), *The Possibility of Knowledge* (Savage, Md.: Rowman and Littlefield), 163–81.
——1988a: "The Internalism/Externalism Controversy." In J. Tomberlin (ed.), *Philosophical Perspectives* 2: 443–59.
——1988b: "Foundationalism, Conceptual Regress, and Reliabilism." *Analysis* 48: 178–84.
——1990: "Metaepistemology and Skepticism." In Ross and Roth 1990: 57–69.
——1992: "Skepticism and Reasoning to the Best Explanation." In E. Villanueva (ed.), *Philosophical Issues* 2: *Rationality in Epistemology*, 149–69.
——1998: "Externalism and Epistemological Direct Realism." *The Monist* 81.3: 393–406.
——2004a: "Epistemic Probability." *Philosophical Perspectives* 14: 149–64.
——2004b: "Achieving Epistemic Ascent." In Greco 2004b: 72–85.
Gallois, A. 1996: *The World Without, the Mind Within* (Cambridge: Cambridge University Press).
——and O'Leary-Hawthorne, J. 1996: "Externalism and Scepticism." *Philosophical Studies* 81.1: 1–26.
Geivett, R. 2003: "Is 'Simple Reliabilism' Adequately Motivated?" *Philosophy and Phenomenological Research* 66.2: 444–50.
Georgalis, N. 1990: "No Access for the Externalist: Discussion of Heil's 'Privileged Access.'" *Mind* 99.393: 101–8.
——1999: "Rethinking Burge's Thought Experiment." *Synthese* 118.2: 145–64.
——2003: "Burge's Thought Experiment: Still in Need of Defense." *Erkenntnis* 58.2: 267–73.
Gertler, B. 2004: "We Can't Know *A Priori* That H2O Exists; But Can We Know That Water Does?" *Analysis* 64.1: 44–7.
Gibbons, J. 1993: "Identity without Supervenience." *Philosophical Studies* 70: 59–80.
——1996: "Externalism and Knowledge of Content." *Philosophical Review* 105.3: 287–310.
——2001: "Externalism and Knowledge of the Attitudes." *Philosophical Quarterly* 51.202: 13–28.
Ginet, C. 1985: "Contra Reliabilism." *The Monist* 68: 175–87.
Gleb, G. 1990: "The Trouble with Goldman's Reliabilism." *Australasian Journal of Philosophy* 68.4: 382–94.
Godfrey-Smith, P. 1992: "Indication and Adaptation." *Synthese* 92.2: 283–312.

Goldberg, S. 1997: "Self-Knowledge, Self-Ascription, and the Memory Argument." *Analysis* 57.3: 211–19.
____ 1999a: "The Relevance of Discriminatory Knowledge of Content." *Pacific Philosophical Quarterly* 80.2: 136–56.
____ 1999b: "Reply to Brueckner." *Analysis* 59.3: 212–17.
____ 2000a: "Externalism and Self-Knowledge of Content: A New Incompatibilist Strategy." *Philosophical Studies* 100.1: 51–78.
____ 2000b: "World-Switching, Word Ambiguity, and Semantic Intentions." *Analysis* 60.3: 260–4.
____ 2002: "Do Anti-individualistic Construals of the Attitudes Capture the Agent's Conceptions?" *Noûs* 36.4: 597–621.
____ 2003a: "Anti-individualism, Conceptual Omniscience, and Skepticism." *Philosophical Studies* 116.1: 53–78.
____ 2003b: "What do you Know when you Know your Own Thoughts?" In Nuccetelli 2003: 241–56.
____ 2003c: "On our Alleged *A Priori* Knowledge that Water Exists." *Analysis* 63.1: 38–41.
____ 2004: "Radical Interpretation, Understanding, and Testimonial Transmission." *Synthese* 138.3: 387–416.
____ 2005a: "The Dialectical Context of Boghossian's Memory Argument." *Canadian Journal of Philosophy* 35.1: 135–48.
____ 2005b: "(Non standard) Lessons from World Switching Cases." *Philosophia* 32.1: 95–131.
____ 2006a: "Brown on Self-Knowledge and Discrimination." *Pacific Philosophical Quarterly* 87.3: 301–14.
____ 2006b: "An Anti-individualistic Semantics for 'Empty' Natural Kind Terms." *Grazer Philosophische Studien* 70: 55–76.
____ 2007a: *Anti-Individualism: Mind and Language, Knowledge and Justification* (Cambridge: Cambridge University Press).
____ 2007b: "Anti-individualism, Content Preservation, and Discursive Justification." *Noûs.*
Goldman, A. 1967: "A Causal Theory of Knowing." *Journal of Philosophy* 64: 357–72.
____ 1980: "The Internalist Conception of Justification." *Midwest Studies in Philosophy* 5: 27–52.
____ 1986: *Epistemology and Cognition* (Cambridge, Mass.: Harvard University Press).
____ 1999: "Internalism Exposed." *Journal of Philosophy* 96.6: 271–93.
Glock, J. 1995: "Externalism and First Person Authority." *The Monist* 78.4: 515–33.
Greco, J. 1990: "Internalism and Epistemically Responsible Belief." *Synthese* 85: 245–77.
____ 1993: "Virtues and Vices of Virtue Epistemology." *Canadian Journal of Philosophy* 23.3: 413–32.
____ 1999a: "Skepticism, Reliabilism, and Virtue Epistemology." In R. Cobb-Stevens (ed.), *The Proceedings of the Twentieth World Congress of Philosophy*, v: *Epistemology* (Bowling Green, Oh.: Philosophy Documentation Center), 139–47.
____ 1999b: "Agent Reliabilism." *Noûs* 13 (suppl): 273–96.
____ 2000: *Putting Skeptics in their Place* (Cambridge: Cambridge University Press).
____ 2003a: "Virtue and Luck, Epistemic and Otherwise." *Metaphilosophy* 34.3: 353–66.

Greco, J. 2003b: "Further Thoughts on Agent Reliabilism: Replies to Cohen, Geivett, Kvanvig, and Schmitt and Lahroodi." *Philosophy and Phenomenological Research* 66.2: 466–80.

____ 2004a: *Ernest Sosa and his Critics* (Oxford: Basil Blackwell).

____ 2004b: "How to Preserve your Virtue while Losing your Perspective." In Greco 2004b: 96–105.

____ 2005: "Justification Is Not Internal." In M. Steup E. Sosa (eds.), *Contemporary Debates in Epistemology* (Oxford: Blackwell), 257–70.

Grimm, S. 2001: "Ernest Sosa, Knowledge, and Understanding." *Philosophical Studies* 106.3: 171–91.

Grundmann, T. 1999: "Bonjour's Self-Defeating Argument for Coherentism." *Erkenntnis* 50.2/3: 463–79.

Hacker, P. 1998: "Davidson on Intentionality and Externalism." *Philosophy* 73.286: 539–52.

Hahn, M. 2004: "When Swampmen Get Arthritis: 'Externalism' in Burge and Davidson." In Hahn and Ramberg 2004: 29–58.

____ and Ramberg, B. (eds.) 2004: *Reflections and Replies: Essays on the Philosophy of Tyler Burge* (Cambridge, Mass.: MIT Press).

Hale, B. 2000: "Transmission and Closure." In E. Sosa (ed.), *Skepticism: Philosophical Issues* 10: 172–90.

Harper, W. 1998: "Papier Mache Problems in Epistemology: A Defense of Strong Internalism." *Synthese* 116.1: 27–49.

Hawthorne, J. 1999: "Implicit Belief and A Priori Knowledge." *Southern Journal of Philosophy* 38 (suppl): 191–210.

Heal, J. 1998: "Externalism and Memory II." *Proceedings of the Aristotelian Society* suppl. vol. 72: 95–109.

Heil, J. 1984: "Reliability and Epistemic Merit." *Australasian Journal of Philosophy* 62: 327–38.

____ 1987: "Are We Brains in a Vat? Top Philosophers Say 'No.' " *Canadian Journal of Philosophy* 17: 427–36.

____ 1988: "Privileged Access." *Mind* 97: 238–51.

____ (ed.) 2004: *Philosophy of Mind: A Guide and Anthology* (New York: Routledge).

Henderson, D. and Horgan, T. 2001: "Practicing Safe Epistemology." *Philosophical Studies* 102.3: 227–58.

Hetherington, S. 1990: "Epistemic Internalism's Dilemma." *American Philosophical Quarterly* 27: 245–51.

____ 1991: "On Being Epistemically Internal." *Philosophy and Phenomenological Research* 51: 855–71.

Hickey, L. 1999: "The Chomskian Challenge to Externalism." *International Studies in Philosophy* 31.4: 39–51.

Higginbotham, J. 1992: "Skepticism Naturalized." In E. Villanueva (ed.), *Philosophical Issues* 2: *Rationality in Epistemology*, 115–29.

Hill, C. 1994: "Two Cheers for Process Reliabilism." *Pacific Philosophical Quarterly* 75.1: 12–28.

____ 1996: "Process Reliabilism and Cartesian Scepticism." *Philosophy and Phenomenological Research* 56.3: 567–81.

Hohwy, J. 2002: "Privileged Self-Knowledge and Externalism: A Contextualist Approach." *Pacific Philosophical Quarterly* 83.3: 235–52.
Hookway, C. 1998: "Normative Concepts and Epistemological Internalism." *Philosophy and Phenomenological Research* 58.4: 907–12.
Horowitz, A. 1996: "Putnam, Searle, and Externalism." *Philosophical Studies* 81.1: 27–69.
____ 2001: "Contents Just Are in the Head." *Erkenntnis* 54.3: 321–44.
____ 2005: "Externalism, the Environment, and Thought-Tokens." *Erkenntnis* 63.1: 133–8.
Horwich, P. 1995: "Comment on Dretske." In E. Villanueva (ed.), *Perception: Philosophical Issues* 7: 167–70.
Hurley, S. 1998: "Vehicles, Contents, Conceptual Structure, and Externalism." *Analysis* 58.1: 1–6.
Hymers, M. 1997: "Realism and Self-Knowledge: A Problem for Burge." *Philosophical Studies* 86.3: 303–25.
Jackman, H. 1999: "We Live Forward but Understand Backwards: Linguistic Practices and Future Behavior." *Pacific Philosophical Quarterly* 80.2: 157–77.
____ 2001: "Semantic Pragmatism and A Priori Knowledge (or 'Yes We Could All Be Brains in a Vat')." *Canadian Journal of Philosophy* 31.4: 455–80.
____ 2004: "Temporal Externalism and Epistemic Theories of Vagueness." *Philosophical Studies* 117.1–2: 79–94.
____ 2005a: "Intuitions and Semantic Theory." *Metaphilosophy* 36.3: 363–80.
____ 2005b: "Temporal Externalism, Deference, and our Ordinary Linguistic Practice." *Pacific Philosophical Quarterly* 86.3: 365–80.
Jacob, P. 1990: "Externalism Revisited: Is There Such a Thing as Narrow Content?" *Philosophical Studies* 60.3: 143–76.
____ 2001: "Is Self-Knowledge Compatible with Externalism?" *Mind and Society* 2.3: 59–75.
Jacobs, J. 1996: "The Virtues of Externalism." *Southern Journal of Philosophy* 34.3: 285–99.
Jacobson, S. 1992: "Internalism in Epistemology and the Internalist Regress." *Australasian Journal of Philosophy* 70.4: 415–24.
____ 1997: "Externalism and Action-Guiding Epistemic Norms." *Synthese* 110.3: 381–97.
Jager, C. 2004: "Skepticism, Information, and Closure: Dretske's Theory of Knowledge." *Erkenntnis* 61.2–3: 187–201.
Johnsen, B. 2003: "Of Brains in Vats, Whatever Brains in Vats May Be." *Philosophical Studies* 112.3: 225–49.
Johnston, M. 1995: "Is the External World Invisible?" In E. Villanueva (ed.), *Perception: Philosophical Issues* 7: 185–98.
Kaplan, D. 1997: "Demonstratives." In J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan* (Oxford: Oxford University Press), 481–563.
Kim, K. 1993: "Internalism and Externalism in Epistemology." *American Philosophical Quarterly* 30.4: 303–16.
____ 1995: "Dretske's Qualia Externalism." In E. Villanueva (ed.), *Perception: Philosophical Issues* 7: 159–65.

Kimbrough, S. 1998: "Anti-individualism and Fregeanism." *Philosophical Quarterly* 48.193: 470–82.
Kirk, R. 1994: "The Trouble with Ultra-externalism." *Proceedings of the Aristotelian Society* 94: 293–307.
____ 1996: "Why Ultra Externalism Goes Too Far." *Analysis* 56.2: 73–9.
Klein, M. 1996: "Externalism, Content and Causation." *Proceedings of the Aristotelian Society* 96: 159–76.
Klein, P. 2004: "Skepticism: Ascent and Assent?" In Greco 2004b: 112–25.
Kobes, B. 1990: "Individualism and Artificial Intelligence." In J. Tomberlin (ed.), *Philosophical Perspectives* 4: *Action Theory and Philosophy of Mind*, 429–59.
____ 1996: "Mental Content and Hot Self-Knowledge." *Philosophical Topics* 24.1: 71–99.
Kornblith, H. 1988: "How Internal Can You Get?" *Synthese* 74: 313–27.
____ (ed.) 2001: *Epistemology: Internalism and Externalism* (Oxford: Blackwell).
____ 2002: *Knowledge and its Place in Nature* (Oxford: Oxford University Press).
____ 2003a: "Can Internalism Be Saved?" *Metaphilosophy* 34.5: 621–9.
____ 2003b: "Roderick Chisholm and the Shaping of American Epistemology." *Metaphilosophy* 34.5: 582–602.
____ 2004: "Does Reliabilism Make Knowledge Merely Conditional?" *Philosophical Perspectives* 14: 185–200.
____ 2006: "Reply to Bermudez and Bonjour." *Philosophical Studies* 127.2: 337–49.
Kraay, K. 2002: "Externalism, Memory, and Self-Knowledge." *Erkenntnis* 56.3: 297–317.
Kripke, S. 1972: *Naming and Necessity* (Cambridge, Mass.: Harvard University Press).
____ 1979: "A Puzzle about Belief." In N. Salmon and S. Soames (eds.), *Propositions and Attitudes* (Oxford: Oxford University Press), 102–48.
Kvanvig, J. 2003: "Simple Reliabilism and Agent Reliabilism." *Philosophy and Phenomenological Research* 66.2: 451–6.
____ and Mezel, C. 1990: "The Basic Notion of Justification." *Philosophical Studies* 59.3: 235–61.
Lahroodi, R. and Schmitt, F. 2003: "Comment on John Greco's *Putting Skeptics in their Place*." *Philosophy and Phenomenological Research* 66.2: 457–65.
Lammenranta, M. 1996: "Reliabilism and Circularity." *Philosophy and Phenomenological Research* 56.1: 111–24.
Langsam, H. 2002: "Externalism, Self-Knowledge, and Inner Observation." *Australasian Journal of Philosophy* 80.1: 42–61.
Larkin, W. 1999: "Brute Error with Respect to Content." *Philosophical Studies* 94.1–2: 159–71.
Leclerc, A. 2001: "Fallibilism, Demonstrative Thoughts and Russellian Propositions." *Principia* 5.1–2: 43–54.
Lehrer, K. 1988a: "Coherence, Justification, and Chisholm." In J. Tomberlin (ed.), *Philosophical Perspectives* 2: 125–38.
____ 1988b: "Metaknowledge: Undefeated Justification." *Synthese* 74: 329–47.
____ 1989: "Knowledge Reconsidered." In M. Clay and K. Lehrer (eds.), *Knowledge and Scepticism* (Boulder, Colo.: Westview), 131–54.
____ 2000: *Theory of Knowledge* (Boulder, Colo.: Westview).

——— and Cohen, S. 1983: "Justification, Truth, and Coherence." *Synthese* 55: 191–208.
——— and Kihyoen, K. 1990: "The Fallibility Paradox." *Philosophy and Phenomenological Research* 50: 99–107.
Levin, M. 1997: "You Can Always Count on Reliabilism." *Philosophy and Phenomenological Research* 57.3: 607–17.
Lipson, M. and Savitt, S. 1993: "A Dilemma for Causal Reliabilist Theories of Knowledge." *Canadian Journal of Philosophy* 23.1: 55–74.
Lipton, P. 2000: "Tracking Track Records." *Proceedings of the Aristotelian Society* suppl. vol. 74: 179–205.
Loar, B. 1988: "Social Content and Psychological Content." In R. Grimm and D. Merrill (eds.), *Contents of Thought* (Tucson: University of Arizona Press), 99–110.
——— 1998: "Is There a Good Epistemological Argument against Concept Externalism?" In E. Villanueva (ed.), *Concepts: Philosophical Issues* 9: 213–17.
Losonky, M. 1995: "Embedded Systems versus Individualism." *Minds and Machines* 5.3: 357–71.
Ludlow, P. 1995a: "Social Externalism, Self-Knowledge, and Memory." *Analysis* 55.3: 157–9.
——— 1995b: "Externalism, Self-Knowledge, and the Prevalence of Slow Switching." *Analysis* 55.1: 45–9.
——— 1995c: "Social Externalism and Memory: A Problem?" *Acta Analytica* 14: 69–76.
——— 2001: "Externalism, Logical Form, and Linguistic Intentions." In Barber 2001: 399–414.
——— and Martin, N. (eds.) 1997: *Externalism and Self-Knowledge* (Palo Alto, Calif.: CSLI Press).
Ludwig, K. 1993: "Externalism, Naturalism, and Method." In E. Villanueva (ed.), *Philosophical Issues* 4: *Naturalism and Normativity*, 250–64.
——— 1996: "Duplicating Thoughts." *Mind and Language* 11.1: 92–102.
——— 1995: "Shape Properties and Perception." In E. Villanueva (ed.), *Philosophical Issues* 7: *Perception*, 325–50.
Luper-Foy, S. 1985: "The Reliabilist Theory of Rational Belief." *The Monist* 68: 203–25.
——— 1987: "Doxastic Skepticism." *Southern Journal of Philosophy* 25: 529–538.
——— 1988: "The Knower, Inside and Out." *Synthese* 74: 349–67.
Lycan, W. 2001: "The Case for Phenomenal Externalism." *Philosophical Perspectives* 15: 17–35.
Macarthur, D. 2003: "McDowell, Scepticism, and the "Veil of Perception"." *Australasian Journal of Philosophy* 81.2: 175–90.
McClamrock, R. 1991: "Methodological Individualism Considered as a Constitutive Principle of Scientific Inquiry." *Philosophical Psychology* 4.3: 343–54.
McCulloch, G. 1988a: "Carruthers Repulsed." *Analysis* 48: 96–100.
——— 1988b: "Faith, Hope, and Charity: Russellian Thoughts Defended." *Analysis* 48: 85–90.
——— 1990: "Externalism and Experience." *Analysis* 50: 244–50.
——— 1994: "Not Much Trouble for Ultra-externalism." *Analysis* 54.4: 265–9.
——— 2000: "Content Externalism and Cartesian Scepticism: A Reply to Brueckner." In Stern 2000: 251–70.

McCulloch, G. 2002: "Phenomenological Externalism." In N. Smith (ed.), *Reading McDowell: On 'Mind and World'* (New York: Routledge), 123–39.
——2003: *The Life of the Mind: An Essay on Phenomenological Externalism* (New York: Routledge).
MacDonald, C. 1995: "Externalism and First Person Authority." *Synthese* 104.1: 99–122.
——1998a: "Externalism and Authoritative Self-Knowledge." In Wright, Smith, and Macdonald 1998: 123–54.
——1998b: "Externalism and Norms." In A. O'Hear (ed.), *Current Issues in Philosophy of Mind* (New York: Cambridge University Press), 273–301.
McDowell, J. 1994: *Mind and World* (Cambridge, Mass.: Harvard University Press).
——1998: *Meaning, Knowledge, and Reality* (Cambridge, Mass.: Harvard University Press).
McEvoy, M. 2005: "Belief Independent Processes and the Generality Problem for Reliabilism." *Dialectica* 59.1: 19–35.
McGinn, C. 1982: "The Structure of Content." In A. Woodfield (ed.), *Thought and Object*. (Oxford: Clarendon Press), 207–58.
——1989: *Mental Content* (Oxford: Basil Blackwell).
McKinsey, M. 1991a: "Anti-individualism and Privileged Access." *Analysis* 51: 9–16.
——1991b: "The Internal Basis of Meaning." *Pacific Philosophical Quarterly* 72: 143–69.
——1994: "Accepting the Consequences of Anti-individualism." *Analysis* 54.2: 124–8.
——2002: "Forms of Externalism and Privileged Access." *Philosophical Perspectives* 16: 199–224.
McLaughlin, B. 2000: "Self-Knowledge, Externalism, and Skepticism." *Proceedings of the Aristotelian Society* suppl. vol. 74: 93–117.
——and Tye, M. 1998a: "Is Content Externalism Compatible with Privileged Access?" *Philosophical Review* 107.3: 349–80.
—— —— 1998b: "Externalism, Twin Earth, and Self-Knowledge." In Wright, Smith, and Macdonald 1998: 285–320.
Malpas, J. 1994: "Self-Knowledge and Scepticism." *Erkenntnis* 40.2: 165–84.
Mandelker, S. 1991: "An Argument against the Externalist Account of Psychological Content." *Philosophical Psychology* 4.3: 375–82.
Markie, P. 1996: "Goldman's New Reliabilism." *Philosophy and Phenomenological Research* 56.4: 799–817.
——2005: "Easy Knowledge." *Philosophy and Phenomenological Research* 70.2: 406–16.
Meeker, K. 1998: "Chisholming Away at Plantinga's Critique of Epistemic Deontology." *Australasian Journal of Philosophy* 76.1: 90–6.
——2004: "Justification and the Social Nature of Knowledge." *Philosophy and Phenomenological Research* 69.1: 156–72.
Mellor, D. 1977: "Natural Kinds." *British Journal for the Philosophy of Science* 28: 299–312.
Merceir, A. 1993: "Normativism and the Mental: A Problem of Language Individuation." *Philosophical Studies* 72.1: 71–88.
Miller, A. 2004: "Rule-Following and Externalism." *Philosophy and Phenomenological Research* 68.1: 127–40.
Miller, R. 1997: "Externalist Self-Knowledge and the Scope of the A Priori." *Analysis* 57.1: 67–75.

Millgram, E. 1996: "Williams' Argument against External Reasons." *Noûs* 30.2: 197–220.
Millikan, R. 1993: *White Queen Psychology and Other Essays for Alice* (Cambridge, Mass.: MIT Press).
____ 1998: "How We Make our Ideas Clear: Empiricist Epistemology for Empirical Concepts." *Proceedings and Addresses of the American Philosophical Association* 72.2: 65–79.
Montmarquet, J. 1985: "Epistemological Internalism." *Southern Journal of Philosophy* 23: 229–40.
Moore, J. 1991: "Knowledge, Society, Power, and the Promise of Epistemological Externalism." *Synthese* 88.3: 379–98.
Morton, A. 2003: "Saving Belief from (Internalist) Epistemology." *Facta Philosophica* 5.2: 277–95.
Moser, P. 1985: "Knowledge without Evidence." *Philosophia* 15: 109–16.
____ 1988a: "How not to be a Coherentist: Reply to Bonjour's 'Reply to Moser's "Internalism and Coherentism"'." *Analysis* 48: 166–7.
____ 1988b: "Internalism and Coherentism: A Dilemma." *Analysis* 48: 161–3.
____ 1989: "Reliabilism and Relevant Worlds." *Philosophia* 19.2/3: 155–64.
____ (ed.) 2002: *The Oxford Handbook of Epistemology* (Oxford: Oxford University Press).
Moya, C. 1998: "Boghossian's Reductio of Compatibilism." In E. Villanueva (ed.), *Concepts: Philosophical Issues* 9: 243–51.
Muller, O. 2001: "Does Putnam's Argument Beg the Question against the Skeptic? Bad News for Radical Skepticism." *Erkenntnis* 54.3: 299–320.
Nagasawa, Y. 2002: "Externalism and the Memory Argument." *Dialectica* 56.4: 335–46.
Naumann, R. 1997: "Events and Externalism." In G. Preyer, F. Siebelt, and A. Ulfig (eds.), *Language, Mind, and Epistemology: On Donald Davidson's Philosophy* (Dordrecht: Kluwer), 117–44.
Newman, A. 2004: "The Good, the Bad, and the Irrational: Three Views about Mental Content." *Philosophical Psychology* 17.1: 95–106.
____ 2005: "Two Grades of Internalism (Pass and Fail)." *Philosophical Studies* 122.2: 153–69.
Nöe, A. 2001: "Experience and the Active Mind." *Synthese* 129.1: 41–60.
Noonan, H. 1993: "Object-Dependent Thoughts." In J. Heil and A. Mele (eds.), *Mental Causation* (Oxford: Oxford University Press), 283–308.
____ 2000: "McKinsey-Brown Survives." *Analysis* 60.4: 353–6.
____ 2004: "Against Absence-Dependent Thoughts." *Analysis* 64.1: 92–3.
Noordhof, P. 2004: "Outsmarting the McKinsey-Brown Argument?" *Analysis* 64.1: 48–56.
____ 2005: "The Transmogrification of A Posteriori Knowledge: Reply to Brueckner." *Analysis* 65.1: 88–9.
Nordby, H. 2005: "Davidson on Social Externalism." *Pacific Philosophical Quarterly* 86.1: 88–94.
Nuccetelli, S. 2001: "Is Self-Knowledge an Entitlement? And Why Should We Care?" *Southern Journal of Philosophy* 39.1: 143–55.

Nuccetelli, S. (ed.) 2003: *New Essays on Semantic Externalism and Self-Knowledge* (Cambridge, Mass.: MIT Press).
O'Leary-Hawthorne, J. and Baergen, R. 1995: "Reliabilism and World Counting." *Philosophia* 24.3/4: 377–88.
Owens, J. 1989: "Contradictory Belief and Cognitive Access." *Midwest Studies in Philosophy* 14: 289–316.
____ 1992: "Psychophysical Supervenience: Its Epistemological Foundation." *Synthese* 90.1: 89–117.
____ 1994a: "Psychological Externalism and Psychological Explanation." *Philosophy and Phenomenological Research* 54.4: 921–8.
____ 1994b: "Psychological Externalism." In R. Warner (ed.), *The Mind-Body Problem: A Guide to the Current Debate* (Oxford: Blackwell), ch. 11.
Papineau, D. 1992: "Reliabilism, Induction and Scepticism." *Philosophical Quarterly* 42.166: 1–20.
Pappas, G. 1987: "Suddenly He Knows." In S. Luper-Foy (ed.), *The Possibility of Knowledge* (Savage, Md.: Rowman and Littlefield), 152–62.
Pasnau, R. 1996: "Who Needs an Answer to Skepticism?" *American Philosophical Quarterly* 33.4: 421–32.
Peacocke, C. 1992: *A Study of Concepts* (Cambridge, Mass.: MIT Press).
____ 1993: "Externalist Explanation." *Proceedings of the Aristotelian Society* 93: 203–30.
____ 1994: "Content, Computation and Externalism." *Mind and Language* 9.3: 30–5.
____ 1999: *Being Known* (Oxford: Oxford University Press).
____ 2004a: *The Realm of Reason* (Oxford: Oxford University Press).
____ 2004b: "Explaining Perceptual Entitlement." In R. Schantz (ed.), *The Externalist Challenge* (Berlin: Walter de Gruyter), 441–80.
Perlis, D. 1991: "Putting One's Foot in One's Head, Part I: Why." *Noûs* 25: 435–55.
Pessin, A. and Goldberg, S. (eds.) 1996: *The Twin Earth Chronicles* (Armonk, NY: M. E. Sharpe).
Pierson, R. 2003: "Alston's Concept of Justification." *Teorema* 22.3: 49–58.
Plantinga, A. 1986: "Epistemic Justification." *Noûs* 20: 3–18.
____ 1988a: "Positive Epistemic Status and Proper Function." In J. Tomberlin (ed.), *Philosophical Perspectives* 2: 1–50.
____ 1988b: "Chisholmian Internalism." In D. Austin (ed.), *Philosophical Analysis: A Defense by Example* (Dordrecht: D. Reidel), 127–51.
____ 1991: "Warrant and Designing Agents: A Reply to James Taylor." *Philosophical Studies* 64.2: 203–15.
____ 1992a: *Warrant and Proper Function* (Oxford: Oxford University Press).
____ 1992b: *Warrant: The Current Debate* (Oxford: Oxford University Press).
____ 1992c: "Justification in the 20th Century." *Philosophical Issues* 2: 43–77.
____ 1993: "Why We Need Proper Function." *Noûs* 27.1: 66–82.
Pollock, J. 1987: "Epistemic Norms." *Synthese* 71: 61–95.
Potrc, M. 1989: "Externalizing Content." In J. Brandl and W. Gombocz (eds.), *The Mind of Donald Davidson* (Amsterdam: Rodopi).
Preti, C. 1992: "Opacity, Belief and Analyticity." *Philosophical Studies* 66.3: 297–306.
____ 2005: "Externalism and Analyticity." *Philosophical Studies* 79.3: 213–36.
Pritchard, D. 2001: "The Opacity of Knowledge." *Essays in Philosophy* 2.1: 1–19.

——— 2002a: "McKinsey Paradoxes, Radical Scepticism, and the Transmission of Knowledge across Known Entailments." *Synthese* 130.2: 279–302.
——— 2002b: "Recent Work on Radical Skepticism." *American Philosophical Quarterly* 39.3: 215–57.
——— 2003a: "Virtue Epistemology and Epistemic Luck." *Metaphilosophy* 34.1/2: 106–30.
——— 2003b: "McDowell on Reasons, Externalism and Scepticism." *European Journal of Philosophy* 11.3: 273–94.
——— 2004a: *Epistemic Luck* (Oxford: Oxford University Press).
——— 2004b: "Some Recent Work in Epistemology." *Philosophical Quarterly* 54.217: 604–13.
——— and Kallestrup, J. 2004: "An Argument for the Inconsistency of Content Externalism and Epistemic Internalism." *Philosophia* 31.3–4: 345–54.
Proudfoot, D. 2004: "The Implications of an Externalist Theory of Rule Following Behaviour for Robot Cognition." *Minds and Machines* 14.3: 283–308.
Pryor, J. 2001: "Highlights of Recent Epistemology." *British Journal for the Philosophy of Science* 52.1: 95–124.
Puhl, K. 1997: "Davidson on Intentional Content and Self-Knowledge." In G. Preyer, F. Siebelt, and A. Ulfig (eds.), *Language, Mind, and Epistemology: On Donald Davidson's Philosophy* (Dordrecht: Kluwer), 339–52.
Putnam, H. 1970: "Is Semantics Possible?" In *Mind, Language, and Reality: Philosophical Papers*, ii (Cambridge: Cambridge University Press), 139–52.
——— 1975: "The Meaning of 'Meaning'." In *Mind, Language, and Reality: Philosophical Papers*, ii (Cambridge: Cambridge University Press), 215–71.
Recanati, F. 1994: "How Narrow is Narrow Content?" *Dialectica* 48.3/4: 209–29.
Rechenauer, M. 1997: "Individualism, Individuation and That-Clauses." *Erkenntnis* 46.1: 49–67.
Reiter, D. 1998: "Engel on Internalism and Externalism in Epistemology." *Erkenntnis* 49.2: 175–84.
Rey, G. 1992: "Semantic Externalism and Conceptual Competence." *Proceedings of the Aristotelian Society* 92: 315–33.
Riggs, W. 1998: "What Are the 'Chances' of Being Justified?" *The Monist* 81.3: 452–72.
——— 2002: "Reliability and the Value of Knowledge." *Philosophy and Phenomenological Research* 64.1: 79–96.
Roska-Hardy, L. 1997: "Internalism, Externalism, and Davidson's Conception of the Mental." In G. Preyer, F. Siebelt, and A. Ulfig (eds.), *Language, Mind, and Epistemology: On Donald Davidson's Philosophy* (Dordrecht: Kluwer), 255–97.
Ross, G. and Roth, M. (eds.) 1990: *Doubting: Contemporary Perspectives on Skepticism* (Dordrecht: Kluwer).
Roth, P. 1999: "Naturalizing Goldman." *Southern Journal of Philosophy* 37.1: 89–111.
Rowlands, M. 2003: *Externalism: Putting Mind and World Back Together Again* (Ithaca, NY: McGill-Queen's University Press).
Rudd, A. 1997: "Two Types of Externalism." *Philosophical Quarterly* 47.189: 501–7.
Sartwell, C. 1995: "Radical Externalism Concerning Experience." *Philosophical Studies* 78.1: 55–70.

Sawyer, S. 1998: "Privileged Access to the World." *Australasian Journal of Philosophy* 76.4: 523–33.
____ 1999: "An Externalist Account of Introspective Knowledge." *Pacific Philosophical Quarterly* 80.4: 358–78.
____ 2002: "In Defence of Burge's Thesis." *Philosophical Studies* 107.2: 109–28.
____ 2003a: "Sufficient Absences." *Analysis* 63.3: 202–8.
____ 2003b: "Conceptual Errors and Social Externalism." *Philosophical Quarterly* 53.211: 265–73.
____ 2004: "Absences, Presences and Sufficient Conditions." *Analysis* 64.4: 354–7.
Schantz, R. (ed.) 2004: *The Externalist Challenge* (Berlin: Walter de Gruyter).
Schiffer, S. 1992: "Boghossian on Externalism and Inference." In E. Villanueva (ed.), *Philosophical Issues* 2: *Rationality in Epistemology*, 29–38.
Schmitt, F. 1992: *Knowledge and Belief* (London: Routledge).
____ 1993: "Epistemic Perspectivism." In Kornblith 2001: 180–206.
____ 1994a: "The Justification of Group Beliefs." In Schmitt 1994b: 257–88.
____ 1994b: *Socializing Epistemology: The Social Dimensions of Knowledge* (Savage, Md.: Rowman and Littlefield).
Searle, J. 1983: *Intentionality* (Cambridge: Cambridge University Press).
Segal, G. 1991: "In Defence of a Reasonable Individualism." *Mind* 100.399: 485–94.
____ 2000: *A Slim Book about Narrow Content* (Cambridge, Mass.: MIT Press).
Sennett, J. 1992: "Toward a Compatibility Theory for Internalist and Externalist Epistemologies." *Philosophy and Phenomenological Research* 52.3: 641–55.
Shagrir, O. 2001: "Content, Computation and Externalism." *Mind* 110.438: 369–400.
Sidelle, A. 2001: "An Argument That Internalism Requires Infallibility." *Philosophy and Phenomenological Research* 63.1: 163–79.
____ 2002: "Innoculi Innocula." *Philosophy and Phenomenological Research* 64.2: 409–11.
Silverberg, A. 1998: "Semantic Externalism: A Response to Chomsky." *Protosociology* 11: 216–44.
Smith, A. 2003: "Semantic Externalism, Authoritative Self-Knowledge, and Adaptation to Slow Switching." *Acta Analytica* 18.30/1: 71–87.
Sosa, D. 1995: "Getting Acquainted with Perception." In E. Villanueva (ed.), *Perception: Philosophical Issues* 7: 209–14.
Sosa, E. 1980: "The Raft and the Pyramid." *Midwest Studies in Philosophy* 5: 3–26.
____ 1985a: "The Coherence of Virtue and the Virtue of Coherence." *Synthese* 64: 3–28.
____ 1985b: "Knowledge and Intellectual Virtue." *The Monist* 68: 226–45.
____ 1988a: "Methodology and Apt Belief." *Synthese* 74: 415–26.
____ 1988b: "Beyond Scepticism, to the Best of our Knowledge." *Mind* 97: 153–88.
____ 1991a: *Knowledge in Perspective: Selected Essays in Epistemology* (Cambridge: Cambridge University Press).
____ 1991b: "Between Internalism and Externalism." *Philosophical Perspectives* 1: *Consciousness*, 179–95.
____ 1991c: "External Realism and Philosophy in Transition." *Journal of Social Philosophy*, 183–6.
____ 1992: "Generic Reliabilism and Virtue Epistemology." In E. Villanueva (ed.), *Philosophical Issues* 2: *Rationality in Epistemology*, 79–92.

——— 1993a: "Skepticism and our Knowledge Circle." In E. Villanueva (ed.), *Philosophical Issues* 4: *Naturalism and Normativity*, 211–26.
——— 1993b: "Abilities, Concepts, and Externalism." In J. Heil and A. Mele (eds.), *Mental Causation* (Oxford: Oxford University Press), 309–28.
——— 1993c: "Proper Functionalism and Virtue Epistemology." *Noûs* 27.1: 51–65.
——— 1994: "Philosophical Scepticism and Epistemic Circularity." *Proceedings of the Aristotelian Society* suppl. vol. 68: 263–90.
——— 1997: "Chisholm's Epistemology and Epistemic Internalism." In L. Hahn (ed.), *The Philosophy of Roderick M. Chisholm* (Chicago: Open Court), 267–87.
——— 2001a: "Reliabilism and Intellectual Virtue." In Kornblith 2001: 147–62.
——— 2001b: "Human Knowledge, Animal and Reflective." *Philosophical Studies* 106.3: 193–6.
——— 2002: "Tracking, Competence, and Knowledge." In Moser 2002: 264–86.
——— 2004a: "Circularity and Epistemic Priority." In R. Schantz (ed.), *The Externalist Challenge* (Berlin: Walter de Gruyter).
——— 2004b: "Two False Dichotomies: Foundationalism/Coherentism and Internalism/Externalism." In W. Sinnott-Armstrong (ed.), *Pyrrhonian Skepticism* (Oxford: Oxford University Press), 146–60.
Spicer, F. 2004: "On the Identity of Concepts, and the Compatibility of Externalism and Privileged Access." *American Philosophical Quarterly* 41.2: 155–68.
Stalnaker, R. 1989: "On What's in the Head." In J. Tomberlin (ed.), *Philosophical Perspectives* 3: *Philosophy of Mind and Action Theory*, 287–316.
——— 1993: "Twin Earth Revisited." *Proceedings of the Aristotelian Society* 43: 297–311.
——— 1995: *Context and Content* (Oxford: Clarendon Press).
Stephens, J. and Russow, L. 1985: "Brains in Vats and the Internalist Perspective." *Australasian Journal of Philosophy* 63: 205–12.
Stern, R. (ed.) 2000: *Transcendental Arguments: Problems and Prospects* (Oxford: Oxford University Press).
Steup, M. 2004: "Internalist Reliabilism." *Philosophical Perspectives* 14: 403–25.
Stevenson, L. 1999: "First Person Epistemology." *Philosophy* 74.290: 475–97.
Stoneham, T. 1999: "Boghossian on Empty Natural Kind Concepts." *Proceedings of the Aristotelian Society* 99: 119–22.
——— 2003: "Temporal Externalism." *Philosophical Papers* 32.1: 97–107.
——— 2004: "Self-Knowledge." In I. Niiniluoto, M. Sintonen, and J. Wolenski (eds.), *Handbook of Epistemology* (Dordrecht: Kluwer), 647–72.
Stroud, B. 1994: "Scepticism, "Externalism," and the Goal of Epistemology." *Proceedings of the Aristotelian Society* suppl. vol. 68: 291–307.
Stueber, K. 2002: "The Problem of Self-Knowledge." *Erkenntnis* 56.3: 269–96.
Sturgeon, S. 1993: "The Gettier Problem." *Analysis* 53.3: 156–64.
——— 1997: "Rational Mind and its Place in Nature." In M. Sainsbury (ed.), *Thought and Ontology* (Milan: FrancoAngeli).
——— 2000: *Matters of Mind: Consciousness, Reason and Nature* (London: Routledge).
Sullivan, A. 2003: " 'Paging Dr. Lauben! Dr. Gustav Lauben!': Some Questions about Individualism and Competence." *Philosophical Studies* 115.3: 201–24.
Swain, M. 1988: "Alston's Internalistic Externalism." In J. Tomberlin (ed.), *Philosophical Perspectives* 2: *Epistemology*, 461–73.

Swank, C. 1988: "A New and Unimproved Version of Reliabilism." *Analysis* 48: 176–7.
Tappenden, P. 1996: "The Roundsquare Copula: A Semantic Internalist's Rejoinder." *Proceedings of the Aristotelian Society* 96: 395–400.
Taylor, J. 1992: "Plantinga's Proper Functioning Analysis of Epistemic Warrant." *Philosophical Studies* 64.2: 185–202.
Ten-Elshof, G. 2003: "An Internalist Rejoinder to Skepticism." *International Studies in Philosophy* 35.4: 105–14.
Thompson, E. 1995: "Colour Vision, Evolution, and Perceptual Content." *Synthese* 104.1: 1–32.
Tollefsen, D. 2002: "Challenging Epistemic Individualism." *Protosociology* 16: 86–120.
Tye, M. 1998: "Externalism and Memory I." *Proceedings of the Aristotelian Society* 72: 77–94.
Uebel, T. 1996: "An Internalist Dilemma Regained." *Australasian Journal of Philosophy* 74.1: 182–9.
Vahid, H. 1998: "The Internalism/Externalism Controversy: The Epistemization of an Older Debate." *Dialectica* 52.3: 229–46.
——2003a: "Externalism, Slow Switching and Privileged Self-Knowledge." *Philosophy and Phenomenological Research* 66.2: 370–88.
——2003b: "Content Externalism and the Internalism/Externalism Debate in Justification Theory." *European Journal of Philosophy* 11.1: 89–107.
——2004: "Doubts about Epistemic Supervenience." *Journal of Philosophical Research* 29: 153–72.
Van Cleve, J. 2003: "Is Knowledge Easy—or Impossible? Externalism as the Only Alternative to Skepticism." In S. Luper (ed.), *The Skeptics: Contemporary Essays* (Aldershot: Ashgate), 45–59.
Verheggen, C. 1997: "Davidson's Second Person." *Philosophical Quarterly* 47.188: 361–9.
Wallis, C. 1994: "Truth Ratios, Process, Task, and Knowledge." *Synthese* 98.2: 243–69.
Walsh, D. 1998: "Wide Content Individualism." *Mind* 107.427: 625–51.
Warfield, T. 1992: "Privileged Self-Knowledge and Externalism are Compatible." *Analysis* 52.4: 232–7
——1995: "Knowing the World and Knowing our Minds." *Philosophy and Phenomenological Research* 55.3: 525–45.
——1997: "Externalism, Privileged Self-Knowledge, and the Irrelevance of Slow Switching." *Analysis* 57.4: 282–4.
Weatherson, B. forthcoming: "Scepticism, Rationalism, and Externalism." In J. Hawthorne and T. Gendler (eds.), *Oxford Studies in Epistemology* (Oxford: Oxford University Press), i. 311–331.
Wedgwood, R. 2002: "Internalism Explained." *Philosophy and Phenomenological Research* 65.2: 349–69.
Weiner, M. 2005: "Why Does Justification Matter?" *Pacific Philosophical Quarterly* 86.3: 422–44.
Weitzman, L. 1996: "What Makes a Causal Theory of Content Anti-skeptical?" *Philosophy and Phenomenological Research* 56.2: 299–318.
Wikforss, A. 2001: "Social Externalism and Conceptual Errors." *Philosophical Quarterly* 51.203: 217–31.

——— 2003: "An *A Posteriori* Conception of Analyticity?" *Grazer Philosophische Studien* 66: 119–39.
——— 2004: "Externalism and Incomplete Understanding." *Philosophical Quarterly* 54.215: 287–94.
——— 2005: "Naming Natural Kinds." *Synthese* 145.1: 65–87.
Williams, M. 1998: "Bilgrami on Belief and Meaning: Is Fregean Externalism Possible?" *Philosophy and Phenomenological Research* 58.3: 607–11.
——— 2001: "Contextualism, Externalism and Epistemic Standards." *Philosophical Studies* 103.1: 1–23.
Williamson, T. 1995: "Is Knowing a State of Mind?" *Mind* 104.415: 533–65.
——— 1998: "The Broadness of the Mental: Some Logical Considerations." *Philosophical Perspective* (*Noûs Supplement*) 12: 389–410.
——— 1999: "Skepticism, Semantic Externalism, and Keith's Mom." *Southern Journal of Philosophy* 38 (suppl.): 149–58.
——— 2000: *Knowledge and its Limits* (Oxford: Oxford University Press).
——— 2001: "Comments on Michael Williams' Contextualism, Externalism and Epistemic Standards." *Philosophical Studies* 103.1: 25–33.
——— 2004: "Sosa on Abilities, Concepts, and Externalism." In Greco 2004b: 263–71.
Wilson, R. 1994: "Wide Computationalism." *Mind* 103.411: 351–72.
——— 1995: *Cartesian Psychology and Physical Minds: Individualism and the Sciences of the Mind* (Cambridge: Cambridge University Press).
——— 2004: *Boundaries of the Mind: The Individual in the Fragile Sciences: Cognition* (Cambridge: Cambridge University Press).
Woodfield, A. 1985: "Content and Community." *Proceedings of the Aristotelian Society* suppl. vol. 59: 197–214.
——— 1997: "Social Externalism and Conceptual Diversity." In J. Preston (ed.), *Thought and Language* (Cambridge: Cambridge University Press), 77–102.
Woodruff-Smith, D. 1986: "The Ins and Outs of Perception." *Philosophical Studies* 49: 187–211.
Worrall, J. 2000: "Relying on Meta-induction?" *Proceedings of the Aristotelian Society* suppl. vol. 74: 207–35.
Wright, C., Smith, B., and Macdonald, C. (eds.) 1998: *Knowing our Own Minds* (Oxford: Oxford University Press).
Wyler, T. 1994: "First-Person Authority and Singular Thoughts." *Zeitschrift für philosophische Forschung* 48.4: 585–94.
Yalowitz, S. 1999: "Davidson's Social Externalism." *Philosophia* 27.1–2: 99–136.
Yourgrau, P. 1987: "Information Retrieval and Cognitive Accessibility." *Synthese* 70: 229–46.
Zagzebski, L. 1999: "From Reliabilism to Virtue Epistemology." In R. Cobb-Stevens (ed.), *The Proceedings of the Twentieth World Congress of Philosophy*, v: *Epistemology* (Bowling Green, Oh.: Philosophy Documentation Center), 173–9.
Zalabardo, J. 1998: "Putting Reference beyond Belief." *Philosophical Studies* 91.3: 221–57.

This page intentionally left blank

Index

a posteriori, and knowledge 204
a priority:
  and contingency 209
  and empty thoughts 205–6
  and justification and the intellect 214–17
  and knowledge 201–2
    close world 204–5
    contingency 209
    environment independence 201–4
    experience independence 207–14
  and method-based conception 208, 209–14
  and overestimation of 201
  and social nature of meaning 206–7
  and test for 211
  and thought content 179
access:
  and epistemic internalism 15–16, 36, 55
  and introspection 47
  and semantic externalism 19–25
  and thought content 20–1
access internalism (AI) 14 n1, 17, 24
  and access to thought content 21
  and epistemic internalism 16
  and neo-Mooreanism 84
  and semantic externalism 21, 24
  and transglobal reliabilism 124–8
accessibilism 15 n2
Adams, F 185 n22
adverbial theory 37–8
  and epistemic internalism 38
agent internalism 126–7
Alston, W P 15, 16 n5, 16 n7, 55 n5, 84, 144 n24
amoralists 153, 154, 155
Anscombe, G E M 265 n21
anti-individualism, and perceptual entitlement 162
assertion 91–2
attitude contents, and content externalism 51
attributer contextualism, *see* contextualism
Audi, R 15
awareness, and epistemic internalism 39
Ayer, A J 38 n2

Baldwin, T 68 n1
basic rules, and rational belief 27–31
Bealer, G 214
belief:
  and agent internalism 126–7
  and basic rules 27–31
  and basing requirement 40
  and blameless belief 26, 27
  and causation 40–1, 270–1
  and conceptual requirements 52–3
  and *de re* ascriptions of 45–6
  and doxastic justification 54
  and entitlement 141, 144–5, 161
  and epistemic safety 107–9, 112–13
  and experiences 170–5
  and formation processes 235–6
  and grounds for 208
  and inference 227–9, 232
  and internal states 45
  and introspection 41
  and justification 14–16, 19–25, 26, 40, 41
  and knowledge 253–4
  and mental content 8, 261
  and mental states 30–1, 254
  and methods of formation 2, 209–14
  and modulation control 110–12
  and perceptual content 165–7
    analysis of Brewer's argument 168–70
    experiences 170–5
  and perceptual experience 215
  and propositional justification 54
  and psychological externalism 10–11
  and rationality of 26–30
  and reasons for 144
  and reliability 6–7
    local vs global 102, 104–5
  and safety-based theories 80
  and social nature of meaning 206–7
  and transglobal reliabilism 7, 101–2, 124–5
  and truth 145
  see also false belief
belief states 42–3
Bernecker, S 15
Bird, A 30 n24
blameless belief 26, 27, 29
Boghossian, P 9, 19 n12, 140, 141 n16, 179 n5, 180 n8, 184 n17, 185 n19, 188 n27, 189 n30, 208, 212, 241–2, 250
  and externalism and inference 221–4
Bonjour, L 16 n4, 18 n9, 135 n9, 138 n14, 179 n6, 215 n25, 216 n27, 240, 242
Brandom, R 85 n14, 141 n17, 143
Braun, D 53 n3, 185 n22
Brewer, B 8, 161, 165–75, 194 n33
  and analysis of argument 168–70

Brewer, B (*cont.*)
and perceptual content and reasons for belief 165–7
Brink, D 153 n36
Brown, J 19 n12, 22 n15, 23 n18, 178 n2, 179 n5, 180 n8, 183 n13, 187 n26, 223 n5, 241, 251
Brueckner, A 178 n2, 180 n8, 183 n14, 253
Burge, T 8, 20, 42, 132 n2, 133, 134 n6, 135 n7, 135 n8, 137 n11, 137 n12, 138 n13, 140 n15, 142 n19, 144, 145 n25, 145 n26, 146 n27, 147 n29, 148 n30, 149, 153 n34, 155 n37, 182 n12, 183 n13, 184 n17, 221, 230, 238 n7, 240, 241, 242, 243, 245, 246, 250, 253, 263, 266 n26
and analysis of argument 163–5
and externalism and inference 224–6
and perceptual content and epistemic entitlement 161–3

Cartesian internalism 15 n1
causation:
and belief 40–1, 270–1
and psychological explanation 265–70
Chase, J 18 n10
Chisholm, R 16 n5
Chisholmian internalism 15 n1
closure, and scepticism 70–1, 75–7
Cohen, S 77, 114, 183 n14
concepts, and propositions 52
conceptual requirements:
and content externalism 53
and mentalism 53
and propositional justification 54
and propositions 52–3
Conee, E 14, 15 n2, 25–6, 35, 37, 41, 52 n1, 55 n6, 84, 135 n9
Connection Problem 132, 133–7, 139
and entitlement 152
and epistemic externalism 137, 146–7
and epistemic internalism 135–7
and representation 152
and semantic externalism 133, 134–5, 146
and semantic internalism 133–4
and semantics 151
content externalism:
and attitude contents 51
and conceptual requirements 53
and empty names 53
and epistemic internalism 51, 66
and knowledge:
external parts of 64–6
justification from 62–4
and mental states 51
and partial understanding 59–62
and propositional justification 54
contextualism:
and evidential contextualism 81–3
and scepticism 77–81
contingency:
and knowledge 209
and mental states 257–65
contrastivism, and scepticism 76
counterpart propositions 52
critical reasoning, and self-knowledge 148
Cummins, R 265 n22

Davidson, D 42, 44, 138 n13, 253, 262–4, 265–6, 268, 269–70
Davies, M 141, 142, 144 n24, 180 n10, 183, 189 n30
Dennett, D 133 n4
DeRose, K 77, 80, 183 n14
Descartes, R 135, 137, 147
direct awareness, and epistemic internalism 39
disjunctivism, and evidential contextualism 82, 83
doxastic justification 54
Dretske, F 15, 76, 89, 141, 144 n24, 177 n1, 265 n22
dualism 37

empiricism:
and objective norms 145
and rationalism 150, 152, 157
and reasons internalism 157
and truth 38
empty names, and mentalism 53
entitlement 139–45
and belief 141, 144–5, 161
and conceptions of 140–5
and Connection Problem 152
and explanation of 143
and inference 140–1
and negative entitlement 142
and notion of 133
and objective norms 145
and Opacity Problem 152
and perceptual content 161–3
and perceptual entitlement 161–2
and rationalism 143–4
and truth 140–4
epistemic externalism 2
and Connection Problem 137, 146–7
and entitlement 161
and incompleteness of 7
and justification 131
and knowledge 88
and Mooreanism 88
and neo-Mooreanism 84–7
and norms 152
and Opacity Problem 138–9, 149–50

and transglobal reliabilism 124
epistemic internalism 2, 36–9
and access 15–16, 36, 55
and access internalism 16
and adverbial theory 38
and awareness 39
and belief 5
and Connection Problem 135–7
and content externalism 51, 66
and counterpart propositions 52
and definition of 13–14
and direct awareness 39
and epistemic properties 40–2
and forms of 4, 14–16
and incompleteness of 7
and internal states 36–7, 40
and introspective knowledge of content 20–1
and justification 14–16, 18, 32, 36, 41, 51, 52, 56, 131
partial understanding 59–62
and knowledge 87–8
and mental states 35, 36
and mentalism 5–6, 35, 51–2
advantage of 55–7
retains substance of 57–9
and neo-Mooreanism 84–7
and norms 151–2
and Opacity Problem 139
and reasons internalism 156
and semantic externalism 24, 25
and simple internalism 55–6
and slow switches 20–1
and supervenience on the accessible 15–16, 26
and supervenience on the mental 14–15, 25–32
and theses associated with 36
and transglobal reliabilism 124–8
epistemic justification 2
epistemic safety:
and belief 112–13
and knowledge 80, 202, 204
and justification 107–9, 114, 120–2
and modulation control 113–14
and reliability 107–9, 112, 114–17, 120–2
Truman Show scenario 117–20
and transglobal reliabilism 120–2
epistemology:
and internal/external distinction 2
and internalism/externalism debate 1, 13, 32, 35
and mind and world 2–3
and scepticism 3–4
essential content, and mental states 265–70
Evans, G 185 n23
evidential contextualism 81–3
evidential neo-Mooreanism 83
evidential scepticism, and Moorean anti-scepticism 73–5
evil demon problem, and transglobal reliabilism 100–1, 114–17
experience:
and belief 170–5
and knowledge 207–14
and perceptual experience 215
and thought content 191–2
external state, and epistemic internalism 36–7
externalism, *see* content externalism; epistemic externalism; psychological externalism; reasons externalism; semantic externalism

fallibilism 79
false belief 235
and cognitively immature processes 236
reasoning 239
and environmental factors 236, 237–8
generic factors 237, 238
local factors 237–8
memory 238–9
reasoning 239
semantic externalism 242
and improperly functioning cognitive processes 236, 237, 238
memory 238
reasoning 239
and preservative memory 246
and preservative processes 238–9, 240
external conditions 238–9
memory 238–9
reasoning 239–40
and semantic externalism 240–4
assumptions of 244–51
and sources of 236–7
and unreliable processes 236
Falvey, K 19 n11, 20, 21, 22 n14, 183 n13, 189 n29, 223 n5, 225 n10, 243 n11
Farkas, K 17 n8
Feldman, R 14, 15 n2, 25–6, 35, 37, 41, 52, 55 n6, 84, 135 n9
Field, H 142 n20, 213 n20
Fodor, J 133 n4, 256 n5, 265 n22
Foley, R 85 n14
Frances, B 23 n17
Frege, G 38, 39, 211
Fregean sense 22 n15
Fuller, G 185 n22
Fumerton, R 39 n5, 45 n16
functionalism 260

Gallois, A 187 n26, 189 n30
Gendler, T 204 n10

Gibbons, J 20, 246
Ginet, C 183 n16
'given', and concept of 39
Goldberg, S 19 n11, 22–3, 64 n11, 229 n18, 236 n5, 246 n19, 247 n20
Goldman, A 27, 137 n11, 204 n10, 237
Graham, G 101 n1
Greco, J 14, 15, 30, 94 n23
grounds, and knowledge 208

Hare, R M 153 n35
Harman, G 48, 183 n16
Hart, H L A 266 n30
Hawthorne, J 76 n9, 77, 180 n9, 183 n14, 187 n26, 189 n30, 204 n10, 209 n16
Heal, J 245 n16
Heil, J 20
Heller, M 78 n11
Honore, A 266 n30
Horgan, T 101 n1
Hume, David 38

identity theory, and mental states 257–60, 262
incompatibilism, and McKinsey's Argument 179–80, 188
independence, and knowledge 203–4
inference 219, 226–9
  and beliefs 227–9, 232
  as causal process 232
  and elements of 226–7
  and entitlement 140–1
  and externalism 229–33
    Boghossian on 221–4
    Burge on 224–6
  and ignorance insufficient for incoherence 219, 221, 230–1
  and inferential activation 227, 228–9
  and internalism about inference 219, 221, 231, 233
  and memory 225
    preservative memory 229, 231
    substantive memory 229–30, 232
  and semantic externalism 10
  and time 227–8
intellect, and justification 214–17
intentional realism, and psychological externalism 254, 256
intentional states 42–4
  and *de re* intentional states 45–6
  and internal states 43–4, 45, 48
  and introspection 49
  and mental states 46–7, 48
internal state:
  and adverbial theory 38
  and belief 45
  and epistemic internalism 36–7, 40
  and intentional states 42, 43–4, 45, 48
  and representation 46
  and sense data 38
internalism, *see* access internalism; agent internalism; epistemic internalism; metaphysical internalism; psychological internalism; reasons internalism; semantic internalism; simple internalism
introspection:
  and belief 41
  and intentional states 49
  and justification of belief 195–6
  and limitations of 251
  and McKinsey's Argument 178–9, 194
  and mental states 47
introspective knowledge of comparative content (IKCC):
  and semantic externalism 21–4, 32
  and thought content 21–4
introspective knowledge of content (IKC):
  and epistemic internalism 20–1
  and semantic externalism 19–20, 21, 32
intuition 79–80
judgement externalists 155
  and moral philosophy 153, 154
judgement internalists 155
  and moral philosophy 153, 154
justification 13
  and agent internalism 126–7
  and basing requirement 40
  and belief 40, 41
  and doxastic justification 54
  and epistemic internalism 14–16, 18, 32, 36, 41, 51, 52, 56
  and epistemic safety 107–9, 114, 120–2
  and evil demon problem 114–17
  and false belief 235
  and the intellect 214–17
  and internalism/externalism debate 14, 32
  and internalist fallacy 137
  and introspective knowledge of comparative content 21–3
  and introspective knowledge of content 19–21
  and local vs global reliability 104–5
  and mental content 160
  and mentalism 57–8
    knowledge 62–4
  and nature of 131
  and necessary truths 38
  and partial understanding 59–62
  and probability 41
  and propositional justification 54
  and reliabilism 114
  and reliability 131–2, 146–7

and sensation 57
and supervenience on the mental 25–6
and thought content 190–2, 194
and transglobal reliabilism 124–5
evil demon problem 100–1, 114–17
Truman Show scenario 117–20
and truth 132, 135, 136–7, 160
see also transglobal reliabilism

Keynes, J M 41
Kitcher, P 179 n5, 202 n5
Klein, P 183 n14
knowledge:
and belief 253–4
and close worlds 204–5
and contextualism 77–9
and contingency 209
and empty thoughts 205–6
and environment independence 201–4
and epistemic challenge 88–9
and epistemic internalism 35
and evidential contextualism 81–3
and experience independence 207–14
and external parts of 64–6
and factivity 154
and false belief 235
and grounds for 208
and indexical knowledge 220
and justification from 62–4
as mental state 62, 254, 270
and neo-Mooreanism 87–93
and psychological externalism 253
and safety-based theories of 80, 202, 204
and standards challenge 88
Kornblith, H 85 n14
Kripke, S 186, 258

Lehrer, K 40 n6, 114
Lepore, E 265 n22
Lewis, D 77, 80 n13, 85 n14, 177 n1, 258–62
Loar, B 133 n4
Loewer, B 265 n22
Ludlow, P 20, 43 n12, 221 n3, 245 n16
Luper-Foy, S 35

Majors, B 132 n2, 142 n19, 147 n28, 149 n31
Malcolm, N 266 n29
Martin, M F 173
Martin, N 43 n12
McDowell, J 153 n35, 155 n39, 185 n23
and neo-Mooreanism 93–6
McGinn, C 185 n23
McKinsey, M 19 n12, 178, 179 n5, 180 n8, 220 n2, 253
McKinsey-style reasoning 178–9
and constraints on 183–4
and environment of thought 180–3
hospitable/inhospitable 184–7, 189–90, 192–4
and incompatibilism 179–80, 188
and introspection 178–9, 194, 195–6
and justification of belief 190–2, 194, 195–6
and reflection 179, 193–4
and semantic externalism 8–9
and thought content 184–7
McLaughlin, B 179 n5, 183 n13, 185 n19, 187 n26
meaning, and social nature of 206–7
Melden, A I 265 n21, 266 n28
memory:
and false belief 238–9
and knowledge 221–2
and preservative memory 225, 229, 231, 232
false belief 246
and substantive memory 225, 229–30, 232
Mendeleev, Dmitri 180–1
mental content:
and belief 261
and empirical beliefs 8
and epistemic justification 160
and psychological externalism:
causation 265–70
contingent content 257–65
essential content 265–70
identity theory 257–60, 262
scientific functionalism 260
mental states:
and access 15–16
and belief 254
and belief revision 30–1
and content externalism 51
and epistemic internalism 35, 36
and intentionality 46–7, 48
and introspective accessibility 47
and knowledge 62, 254, 270
and psychological externalism
causation 265–70
contingent content 257–65
essential content 265–70
identity theory 257–60, 262
scientific functionalism 260
mentalism 14, 35
and advantage of 55–7
and conceptual requirements 53
and diversity of 59
and doxastic justification 54
and empty names 53
and epistemic internalism 5–6, 51–2, 55
retention of substance of 57–9
and inclusivity of 59
and justification 57–8

mentalism (*cont.*)
partial understanding 59–62
and knowledge:
external parts of 64–6
justification from 62–4
and propositional justification 54
and strong supervenience 55
metaphysical internalism 5
mind, and world 2–3
see also philosophy of mind and language
modulation control:
and epistemic safety 113–14
and reliability 110–12, 117–20
and transglobal reliabilism 125
Moore, G E 38, 48, 68–9
Mooreanism 68–70
and faults of 70–5
conversational impropriety objection 72
dialectical impropriety objection 70–1
evidential scepticism objection 73–5
impasse objection 71–2
no diagnosis objection 72–3
no supporting epistemology objection 72
moral philosophy 152–7
and internalism/externalism debates 152–3
and judgement externalists 153, 154, 155
and judgement internalists 153, 154, 155
and motivation and moral judgement 153, 154–5
and norms 152–3
and reasons externalism 155, 156
and reasons internalism 155–6
Moser, P 135 n9
motivation, and moral judgement 153, 154–5

neo-Mooreanism 6, 69–70, 97
and closure 75–7
and contextualism 77–81
and evidential contextualism 81–3
and evidential neo-Mooreanism 83
and externalism/internalism distinction 84–7
and knowledge 87–93
and McDowellian neo-Mooreanism 93–6
and sceptical challenge 97
see also Mooreanism; scepticism
Neta, R 66 n13, 81–2, 95 n24
Nilson, Lars Fredrick 180–1
norms 145, 150–2
norms, objective:
and empiricism 145
and entitlement 145
and epistemic externalism 152
and epistemic internalism 151–2
and internalism/externalism debate 157
and moral philosophy 152–3
and rationalism 145
and representation 150–1
and semantic externalism 150–1
and semantic internalism 151
Nozick, R 40 n7, 76
Nuccetelli, S 94 n22

Opacity Problem 132–3, 137–9, 147
and entitlement 152
and epistemic externalism 138–9, 149–50
and epistemic internalism 139
and semantic externalism 138
and semantic internalism 138
and thought content 151
Owens, J 19 n11, 20, 21, 22 n14, 189 n29, 223 n5, 224 n8, 256 n5, 257 n6, 266 n27

pain 258–60
as representational state 48
Parfit, D 155 n39
partial understanding, and justification 59–62
Peacocke, C 140, 143–4, 208, 212
perception 47, 48
as information source 238
and semantic externalism 8
perceptual content:
and epistemic entitlement 161–3
and reasons for belief 165–7
analysis of Brewer's argument 168–70
experiences 170–5
perspectival internalism 14 n1
philosophy, and related empirical disciplines 3
philosophy of mind and language:
and internal/external distinction 1–2
and internalism/externalism debate 1, 13, 32, 35
and mind and world 2
and scepticism 3
and semantic internalism 17–19
physicalism 37
Plantinga, A 136 n10, 144 n24
Pritchard, D H 70 n2, 71 n5, 73 n7, 74, 76 n9, 77 n10, 80 n12, 86 n15, 89 n18, 95 n24, 96 n25, 97 n26, 235 n1
privilege, and externalism 219–21
probability, and justification 41
propositional justification 54
propositions:
and concepts 52
and conceptual requirements 52–3
Pryor, J 15 n2, 16 n5, 17, 55, 86 n16, 135 n9, 136 n10, 142 n19, 144 n24, 190 n31
psychological explanation, and causal interpretation 265–70
psychological externalism 2, 43
and belief 10–11

and intentional realism 254, 256
and knowledge 253
and mental states:
causation 265–70
contingent content 257–65
essential content 265–70
identity theory 257–60, 262
scientific functionalism 260
as metaphysical thesis 253
psychological internalism 2, 42–9
Putnam, H 42, 133 n3, 179, 243, 253, 260

rationalism:
and belief 26–30
and empiricism 150, 152, 157
and entitlement 143–4
and objective norms 145
and reasons externalism 157
and reasons internalism 156
and semantic externalism 151
realism, *see* intentional realism
reasoning:
and false belief 239–40
semantic externalism 241–51
and self-knowledge 148
reasons externalism:
and moral philosophy 155, 156
and rationalism 157
reasons internalism:
and empiricism 157
and epistemic internalism 156
and moral philosophy 155–6
and practical rationality 156
recollection, and false belief 238–9
see also memory
reflection:
and epistemic internalism 56
and justification of belief 195
and knowledge of environment 179, 180
and McKinsey's Argument 179, 193, 195
and Putnam's Argument 179
Reimer, M 185 n22
reliabilism/reliability 103
and barn thought experiment 103–4, 109–10
epistemic safety 107
local vs global reliability 102, 104–5
and control processes 110–12
and epistemic safety 107–9, 112, 120–2
and global reliability 6–7
and justification 114, 131–2, 146–7
evil demon problem 114–17
Truman Show scenario 117–20
and local reliability 6
and modulation control 110–12, 117–20
and neoclassical reliabilism 112–14
and new evil demon problem 114–17
and robustness of reliability 105–7, 122–3
see also transglobal reliabilism
representation 131
and Connection Problem 152
and function of 134
and intellectual norms 150–1
and internal states 46
and representational states 48–9
and requirements of 146
see also Connection Problem
responsibility theory 58
Russell, Bertrand 38, 39, 42–3

safety, *see* epistemic safety
Salmon, N 185 n22
Sawyer, S 20, 132 n2, 142 n19, 147 n28, 147 n29, 149 n31, 179 n7, 201 n1
scepticism:
and closure 70, 75–7
and contextualism 77–81
and contrastivist response to 76
and evidential contextualism 81–3
and internalism/externalism debate 3–4
and McDowellian neo-Mooreanism 93–6
and Moorean anti-scepticism 68–70
conversational impropriety objection to 72
dialectical impropriety objection to 70–1
evidential scepticism objection to 73–5
faults of 70–5
impasse objection to 71–2
no diagnosis objection to 72–3
no supporting epistemology objection to 72
see also neo-Mooreanism
Schaffer, J 76
Schiffer, S 223, 241, 245, 246, 250
scientific functionalism 260
Segal, G 133 n4, 177 n1
self-knowledge 147–8
and externalism 148–9, 219–21, 232
and externalism and inference 221–4
and privileged self-knowledge 220
semantic content 3
semantic externalism 4–5
and access 19–25
and access internalism 21, 24
and Connection Problem 133, 134–5, 146
and definition of 240–1
and epistemic internalism 24, 25
and false belief 240–4
assumptions underlying 244–51
and inference 10
and intellectual norms 150–1

semantic externalism (*cont.*)
and introspective knowledge of comparative content 21–4, 32
and introspective knowledge of content 19–20, 21, 32
and knowledge, environment independence 203–4
and McKinsey-style reasoning 8–9
and Opacity Problem 138
and perceptual justification 8
and *a priori/a posteriori* distinction 9–10
and rationalism 151
and reasoning 241–2
and representation 131
and social nature of meaning 206–7
and supervenience on the accessible 24
and supervenience on the mental 32
and thought content 17, 18, 21, 24–5, 32
and transglobal reliabilism 100, 101, 128–9
semantic internalism:
and Connection Problem 133–4
and norms 151
and Opacity Problem 138
and philosophy of mind 17–19
and representation 131
and thought content 17–18
and transglobal reliabilism 101
sensation 37
and adverbial theory 37–8
and epistemic internalism 56
and justification 57
sense-datum theory 37, 38
and internal states 38
Shafer-Landau, R 153 n36, 155 n39
Shoemaker, S 185 n21
simple internalism 55–6
slow switch 221, 225, 231
and introspective knowledge of comparative content 22–3
and introspective knowledge of content 19–21
Smart, J C 258
Smith, M 153 n35
Sorensen, R 183 n16, 223 n7
Sosa, E 15 n1, 16, 80, 114, 137 n11, 202 n2
spatio-temporality:
and externalism 220, 221–2
and privileged knowledge of 220
Stanley, J 77
states:
and adverbial theory 37–8
and epistemic internalism 36–7
Stecker, R 185 n22
Steup, M 135 n9
Stine, G C 183 n14
Stoneham, T 185 n21
Strawson, P F 165 n3, 171
subject-sensitive invariantism 77
supervenience on the accessible (SA) 17, 26
and access to thought content 21
and epistemic internalism 15–16
and semantic externalism 24
supervenience on the mental (SM) 18, 25–32
and epistemic internalism 14–15
and justification 14
and semantic externalism 32
Svavarsdóttir, S 153 n36

Taylor, C 267
thought content 13, 177
and access 20–1
and environment of 180–3
hospitable/inhospitable 184–7, 189–90, 192–4
and experiences 191–2
and externalism 177
and incompatibilism 179–80, 188
and internalism/externalism debate 32
and introspection 178–9
and introspective knowledge of comparative content 21–4
and introspective knowledge of content 19–20
and justification 190–1
and McKinsey's Argument 178, 184–7
constraints on 183–4
and Opacity Problem 151
and a priori 179
and reflection 179, 193–4, 195
and semantic externalism 17, 18, 21, 24–5, 32
and semantic internalism 17–18
Tienson, J 101 n1
transglobal reliabilism 6–7, 100–2
and access internalism 124–8
and advantages of 7
and agent internalism 126–7
and belief 124–5
and epistemic externalism 124
and epistemic internalism 124–8
and epistemic safety 120–2
and experiences 127–8
and experientially possible global environments 101
and justification 124–5
evil demon problem 100–1, 114–17
Truman Show scenario 117–20
and lessons from thought experiments 120–4
and modulation control 125
and reliability:
form of 101
local vs global 102

robustness of 122–3
and strong semantic externalism 100, 128–9
truth:
and belief 145
and entitlement 140–4
and justification 132, 135, 136–7, 160
and necessary truths 38
Tye, M 177 n1, 179 n5, 183 n13, 185 n19, 187 n26
Tymoczko, T 179 n7

underdetermination principle, and scepticism 73–5
unexemplified properties 49

Vahid, H 20
Velleman, D 185 n20
Vogel, J 183 n14

Warfield, T 20, 179 n7
warrant, *see* entitlement
Wedgwood, R 14, 18, 25, 26–31
Williams, B 155 n38
Williams, M 141 n17
Williamson, T 27 n22, 28, 29, 39, 62, 64, 66, 87, 91, 92 n20, 140 n15, 173, 202 n2, 213, 216 n28, 253, 254, 270
Winch, P 265 n21
Wittgenstein, L 68, 72, 219
Woodfield, A 180 n8
Wright, C 70 n3, 72 n6, 86, 94 n23, 141, 142 n18, 183, 183 n15, 189 n30, 195 n35, 196 n36

Zagzebski, L 85 n14