

Moral Knowledge

Moral Knowledge

SARAH McGRATH

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

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

© Sarah McGrath 2019

The moral rights of the author have been asserted

First Edition published in 2019

Impression: 1

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, by licence 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 work in any other form
and you must impose this same condition on any acquirer

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2019937226

ISBN 978–0–19–880541–0

Printed and bound in Great Britain by
Clays Ltd, Elcograf S.p.A.

Links to third party websites are provided by Oxford in good faith and
for information only. Oxford disclaims any responsibility for the materials
contained in any third party website referenced in this work.

For Tom, Owen, Orla, and Hugh

Contents

1.	Introduction	1
	1.1 A Working Hypothesis	1
	1.2 Some Methodological Preliminaries and Assumptions	3
	1.3 Towards a More Social Moral Epistemology	8
2.	Reflective Equilibrium, Its Virtues and Its Limits	11
	2.1 Introduction	11
	2.2 The Method Is Too Demanding	20
	2.3 If the Method Is Too Weak, Then It Is Probably Not “The Best Method”	25
	2.4 The Method Is Too Weak	30
	2.5 Toward a More Defensible (But Less Ambitious) Version of the Method	37
	2.6 The Pursuit of Coherence as a Source of Moral Knowledge	42
	2.7 The Method and the Moral Philosopher	45
	2.8 Levels of Generality	48
3.	Moral Knowledge from Others	59
	3.1 The Moral Inheritance View	59
	3.2 Testimony, Deference, and Expertise	68
	3.3 Doubts About Moral Expertise	73
	3.3.1 Theory vs. Anti-theory	75
	3.3.2 Realism and Its Rivals	78
	3.3.3 Egalitarianism	83
	3.4 Opacity, Justification, and Understanding	88
	3.5 The Epistemology of Moral Expertise	99
4.	Observation and Experience	106
	4.1 Introduction	106
	4.2 Enabling and Triggering	110
	4.3 Confirmation and Disconfirmation	112
	4.3.1 Clarifying the Issue	112
	4.3.2 What the Possibility of Confirming Moral Claims by Observation Does Not Presuppose	116
	4.3.3 An Initial Argument in Favor of the Possibility of Confirming a Moral Claim by Non-Moral Observation	120
	4.3.4 Beyond the Case of Testimony	124
	4.3.5 When Facts About How People Actually Behave Provide Evidence for Claims About How We Should Behave	127

viii CONTENTS

4.4 Conditioning	134
4.5 Armchair Moral Knowledge	140
5. Losing Moral Knowledge	151
5.1 Introduction	151
5.2 Moral Forgetting and Ryle's Puzzle	152
5.3 Expertise, Competence, and Knowing the Difference Between Right and Wrong	154
5.4 Forgetting vs. Ceasing to Care	158
5.5 How to Forget the Difference Between Right and Wrong	161
5.6 Defeat	172
5.7 Dworkin Against Debunking	178
5.7.1 Counterfactuals, Conceivability, and Modality	178
5.7.2 Truth, Justification, and Explanation	182
5.7.3 "Hume's Principle"	186
6. Conclusions	192
6.1 General Theses and Methodological Assumptions	192
6.2 Reflective Equilibrium and Coherence as a Source of Moral Knowledge	193
6.3 Social Aspects of Moral Knowledge	195
6.4 Experience and Observation as Sources of Moral Knowledge	197
6.5 Knowing the Difference Between Right and Wrong, Ryle's Puzzle, and Losing Moral Knowledge	198
<i>References</i>	201
<i>Index</i>	211

Acknowledgements

I am grateful to the American Council of Learned Societies and to the John A. Templeton Foundation for supporting the research leave that I had during 2014–2015. I am also grateful to the Princeton University Center for Human Values which awarded me a semester of research leave as a Lawrence S. Rockefeller University preceptor in the Spring of 2012, during which time I developed some of the ideas about moral expertise that are drawn upon in Chapter 3.

I have been presenting many of the ideas contained in this book at workshops, seminars and conferences for several years now. I am grateful to many audiences for helpful discussion and feedback, including audiences at: the Bellingham Summer Philosophy Conference, Boston University, Bowling Green State University; Brown University; Cardiff University; Claremont-McKenna College; Georgetown University; Humboldt-ANU-Princeton workshop on normativity; the Madison Metaethics Workshop; New York University; Northeast Florida Philosophy Conference; the Oberlin Colloquium in Philosophy; the Ohio State University; Penn State University; Pittsburgh University; Princeton University; Princeton Theological Seminary; Rice University; Simon Fraser University; SUNY Buffalo; Syracuse University; the Templeton Intellectual Humility Workshop, Tulane University; Tufts University; University of Missouri; University of North Carolina, Chapel Hill; University of Nebraska, Lincoln; the University of Alabama, Birmingham; University of Edinburgh's Foundations of Normativity Workshop, University of Rochester; University of Sheffield; University of Utah; Virginia Tech; Western Michigan University; University of Nebraska at Omaha; and the University of Southern California.

I am also very much indebted to my Princeton colleagues, especially Elizabeth Harman, Thomas Kelly, Gideon Rosen, and Michael Smith, for detailed feedback on earlier drafts of this material. David Estlund, Geoff Sayre-McCord, and Russ Shafer-Landau have also been extremely generous in offering feedback and support. Three anonymous OUP referees provided enormously helpful comments from which the manuscript has benefited a great deal.

X ACKNOWLEDGEMENTS

Earlier versions of some of the material here has appeared elsewhere, in article form. In particular, portions of the following papers are incorporated into the chapters of this book: “The Puzzle of Pure Moral Deference” (2009) *Philosophical Perspectives* 23.1: 321–44; “Skepticism about Moral Expertise as a Puzzle for Moral Realism” (2010) *Journal of Philosophy* 108.3: 111–37; “Moral Knowledge and Experience” (2011) *Oxford Studies in Metaethics* 6: 107–25; “Is Reflective Equilibrium Enough?” (with Thomas Kelly) (2010) *Philosophical Perspectives* 24; “Forgetting the Difference Between Right And Wrong,” (2015) in S.G. Chappell (ed.) *Intuition, Theory, and Anti-Theory* (Oxford University Press): 104–126.

C1

1

Introduction

C1.P1 It is possible to doubt whether there is such a subject as moral epistemology.

C1.P2

Simon Blackburn

C1.S1

1.1 A Working Hypothesis

C1.P3 As its title suggests, this book is an exploration of moral knowledge: its possibility, its sources, and its characteristic vulnerabilities. Although it is far from a comprehensive survey of questions that might be pursued under the heading of “moral epistemology,” it does address, and argue for answers to, a relatively wide range of such questions. These questions include the following: What are the strengths and weaknesses of the *method of reflective equilibrium* as an account of how we should make up our minds about moral questions? Is this, as many moral philosophers believe, the correct account of how moral inquiry should proceed? What would count as evidence for or against a fundamental moral conviction? Are perception and testimony potential sources of moral knowledge? What, if anything, would be wrong with simply outsourcing your views about moral questions to a *moral expert*, if you had reason to believe that the expert would be more reliable than you are about such questions? How *fragile* is our knowledge of morality, compared to other kinds of knowledge? Is it true, as Gilbert Ryle once claimed, that knowledge of the difference between right and wrong fundamentally differs from knowledge of other kinds in that it cannot be forgotten? To what extent are our moral views vulnerable to being “debunked” by empirical discoveries about why we hold them? What is the relationship between being able to justify a moral judgment and knowing that it is true? Should we invest more confidence in relatively abstract, general moral principles that strike us as true, or more confidence in our judgments about the rightness and wrongness of particular actions?

C1.P4 My hope is that the most significant contributions of this book will prove to be its specific proposals about the topics that I address, as opposed to its

2 MORAL KNOWLEDGE

lending support to any more general thesis about morality or our cognitive relationship to it. For the most part, these proposals have a substantial degree of independence from one another, so that, for example, even if more or less everything that I say about the topic of moral expertise is off the mark, the discussions that I offer of the method of reflective equilibrium or the way in which observation contributes to moral knowledge might still be substantially correct. Moreover, for most potential readers, not every topic that I address will be of equal interest. With that in mind, I have made some effort to make the discussions of the various topics relatively self-contained. When an argument does appeal to some claim that I have argued for earlier in the book, I mark this in a way that makes it as easy as possible for a reader to selectively follow up on the earlier discussion.¹

C1.P5 That having been said, there *is* a larger theme that both unifies and informs my more specific proposals. I have a working hypothesis that is supported by the arguments offered here to the extent that they are successful. The working hypothesis is this: moral knowledge can be acquired in any of the ways in which we acquire ordinary empirical knowledge, and our efforts to acquire and preserve such knowledge are subject to frustration in all of the same ways that our efforts to acquire and preserve ordinary empirical knowledge are. Any source of ordinary empirical knowledge is also a potential source of moral knowledge, and any threat to our ordinary empirical knowledge (or our ability to acquire such knowledge in the first place) is also a threat in the moral domain. Given the many ways in which we gain empirical knowledge of the world around us, this working hypothesis is much stronger than anything that I can hope to argue for (let alone establish) here. Nevertheless, I hope to compensate for this to at least some extent by focusing on potential pressure points—areas where the hypothesis seems *least* plausible, and where the chances for its falsification seem relatively high.

C1.P6 For example, much of what we know about the world depends on evidence that is provided by observation. If the working hypothesis adopted here is true, then it follows that our moral views are susceptible to empirical confirmation and disconfirmation. This implication is one that many philosophers take to be false. Against this, I will defend the claim that even relatively fundamental moral convictions are in principle susceptible to being confirmed or disconfirmed by non-moral observations.² Similarly, it is clear

¹ In addition, a short concluding chapter contains a summary of the main claims endorsed in the book, together with references to the specific sections where those claims are discussed. This too can potentially be used as guide for selective reading.

² As will emerge below, I think that opinions to the contrary often rest on inadequate pictures of confirmation that we have reason to reject independently of anything having to do with morality.

on reflection that a great deal of our ordinary knowledge about the world depends on the testimony of others, including many cases in which we are not in a position to verify the relevant truths ourselves. But many have thought that full-fledged moral knowledge requires the knower to recognize or appreciate why the relevant moral claim is true. Against this, I will defend the possibility of arriving at full-fledged moral knowledge via testimony or (more broadly) deference to another person. I argue that what is legitimate in our unease about deferring to another person about morality does not derive from the impossibility of moral knowledge being transmitted or acquired in this way but rather from other sources.

C1.P7 Of course, the working hypothesis I endorse also entails that moral knowledge is susceptible to being undermined or lost in all of the ways in which non-moral knowledge is. Interestingly, some philosophers—Gilbert Ryle, Ronald Dworkin, and Thomas Nagel among them—have held that there are important respects in which our cognitive relationship to morality is *more* secure than our cognitive relationship to ordinary empirical knowledge.³ I explore several possibilities of this sort in Chapter 5, “Losing Moral Knowledge.” Although I defend the claim that moral knowledge is susceptible to being undermined in the same ways in which non-moral knowledge is, I also think that careful consideration of the proposed ways in which this might fail to be true both raises, and suggests answers to, a number of interesting and under-explored questions. These include questions about the extent to which some cases of moral corruption are best understood as cognitive processes (i.e. processes involving a loss of knowledge), as well as questions about the kinds of considerations that could in principle make it reasonable for us to lose confidence even in moral claims that strike us as obviously correct.

C1.S2 1.2 Some Methodological Preliminaries and Assumptions

C1.P8 A philosopher who announces that she intends to undertake an investigation of moral knowledge is apt to encounter skepticism about her project from a

Indeed, even those who defend the possibility of empirically confirming a moral claim sometimes assume an overly demanding picture of what this would involve. This has led to the view that the empirical confirmation or disconfirmation of a moral claim would require that the moral claim in question be in some sense *reducible* to non-moral, empirical claims. Against this, I will argue that even if a moral claim is *irreducibly normative*, it does not follow that non-moral observations cannot provide evidence for or against it.

³ For example, in the course of defending “normative realism” Nagel claims that “The truth here could not be radically inaccessible in the way that the truth about the physical world might be” (1986: 186).

4 MORAL KNOWLEDGE

number of different directions. Perhaps the most immediate and straightforward kind of skepticism that she confronts is skepticism about whether there is any such thing as moral knowledge at all. More generally, an inquiry into the ways in which we acquire moral knowledge might seem to presuppose answers to difficult metaethical questions that should be considered open, and therefore to beg the question against various metaethical views that are very much live theoretical options. Presumably, the existence of moral knowledge entails the existence of moral facts, no less than the existence of historical knowledge entails the existence of historical facts. And the claim that there are moral facts is denied, both by error theorists as well as by non-cognitivists who hold that the function of moral discourse is something other than the stating of facts. Even among philosophers who are prepared to countenance moral facts, the range of opinion as to the nature of such facts might seem too vast to allow for profitable discussion of moral knowledge as a relation that we sometimes stand in to these relata. In this respect, moral epistemology might seem *premature*—task that is better left until further progress has been made in moral semantics and moral metaphysics, areas which are sometimes treated as more fundamental.

C1.P9 While I feel the force of this line of thought, I do not believe that it warrants setting aside the kind of investigation that I undertake in this book. In part, this is because I am skeptical that moral metaphysics and moral semantics enjoy the kind of methodological priority to moral epistemology that this line of thought seems to assume. Consider the relationship between moral metaphysics and moral epistemology. While there is an obvious respect in which the question of whether there *are* any moral facts seems prior to the question of how we could come to know such facts, as a historical matter reflection on the second question has often influenced answers to the first. In particular, a prominent source of skepticism about whether there are any moral facts at all, or moral facts in anything like the relatively straightforward sense which common sense often seems to assume, has been dissatisfaction with existing accounts of how we might come to know such facts.⁴ (Although, of course, there are other sources of skepticism about the existence of moral facts.) Significantly, all of the mechanisms under consideration in this book—including interlocution, perception, empirical confirmation and

⁴ Thus, one of the main motivations for Mackie's (1977) error theory is his claim that, if we were aware of moral facts, "it would have to be by some special faculty of moral perception or intuition, utterly different from our ordinary ways of knowing everything else" (p. 38). Similarly, in the course of developing an argument against "evaluative realism," Street (2006: 143) objects to the postulation of a "highly specialized, sophisticated capacity, one specifically attuned to the evaluative truths in question."

disconfirmation, and reasoning toward a more coherent overall view—are generally acknowledged to be mechanisms that regularly deliver knowledge of non-moral facts. The only question is whether they also manage to deliver moral knowledge, or could deliver such knowledge if there were moral facts to be known. The more that it can be made plausible that these mechanisms are or would be suitable for delivering moral knowledge, the more one of the traditional reasons for doubting the existence of moral facts is weakened.⁵

C1.P10 Moreover, to the extent that the arguments that I offer here are successful, this would be relevant not only to the question of whether moral facts exist, but also to the question of what such facts would be like. For example, suppose that our moral beliefs are susceptible to empirical confirmation and disconfirmation regardless of whether they are reducible or equivalent to empirical claims. If this is the case, then at least one motivation for a reductionist conception of morality is undercut.⁶

C1.P11 Consider the following claims:

- C1.P12 Rape is wrong.
- C1.P13 Slavery in the antebellum American South was unjust.
- C1.P14 One shouldn't encourage children to smoke cigarettes.
- C1.P15 It was wrong for the 9/11 hijackers to deliberately fly planes into the Twin Towers.
- C1.P16 If one is in a position to help others in desperate need at little cost to oneself, then one should.
- C1.P17 It is right to keep one's promises.
- C1.P18 It is good for parents to vaccinate their children against polio.

⁵ Byrne (2018) usefully distinguishes between “extravagant” and “economical” accounts of our knowledge of a given domain. An account is *extravagant* just in case it invokes a special purpose, dedicated faculty or capacity in accounting for our knowledge of the relevant subject matter, on the grounds that more general purpose capacities are inadequate to deliver such knowledge. In contrast, an account is *economical* just in case it holds that our knowledge of the subject matter is delivered by general purpose epistemic capacities, capacities that also deliver knowledge of distinct subject matters. For example, Chomsky's (1998) account of our knowledge of language is extravagant as opposed to economical, because it entails that this knowledge requires a dedicated faculty, “a language organ.” In contrast, accounts of our knowledge of language that attempt to account for such knowledge in terms of the application of more general learning mechanisms count as economical. In terms of this dichotomy, the arguments of this book tend to support (to the extent they are successful) an economical as opposed to an extravagant account of our moral knowledge.

⁶ To be clear, and despite what might be suggested by the way in which the current study proceeds, I do not think that moral epistemology has any methodological priority to moral metaphysics, moral semantics, or moral psychology, either. My own view is that each of these sub-areas stands to gain from incorporating the genuine insights that ultimately emerge from each of the others, and that from the perspective of the division of cognitive labor it makes sense for different inquirers to set out from different starting points.

6 MORAL KNOWLEDGE

C1.P19 Plausibly, these claims are true, and known to be true by many people. Some philosophers hold that there are “Moorean truths,” philosophically incontestable fixed points that it is the job of the philosopher to respect, accommodate, and build upon in his or her theorizing, as opposed to either vindicate or refute.⁷ While I do not claim this status for the propositions listed above, they are among the closest approximations that the moral domain has to offer. Even if they are not beyond challenge, it seems reasonable to assume provisionally that they are true, in order to investigate how claims of the relevant kind might be known. This is a large project, to which the current study aims to make a modest contribution.⁸

C1.P20 Throughout the book, I invoke the concept of knowledge without offering an analysis (reductive or otherwise) of knowledge. In this respect, if not in others, this book is an exercise in “knowledge first” epistemology, in the tradition of Williamson (2000). Unlike in Williamson’s case, my proceeding in this way is not rooted in a conviction that there is no analysis of knowledge to be had, but rather in agnosticism about this question together with certainty that I myself do not know of such an analysis. I do, however, make the following two assumptions about knowledge.

C1.P21 First, regardless of what the standards for moral knowledge are—and even whether we have any moral knowledge at all—those standards are not materially different from the standards for knowledge in general. In the absence of some compelling reason for thinking that things are otherwise, we should prefer a unified account of knowledge, one on which the standards that must be met in order to count as knowing a proposition do not vary from domain to domain. Notice that, even if we adopt this as a default assumption, it does not in any way beg the question in favor of the working hypothesis

⁷ For representative statements, see especially Armstrong (1999) and Lewis (2001: 418). For a systematic treatment, see Kelly (2005b).

⁸ Thomson (2013: 54–5) suggests that G.E. Moore (1925) should have included some moral propositions among his famous list of commonsense platitudes. Cuneo and Shafer-Landau (2014) is a recent attempt to vindicate the idea that there are some truths that have the status of methodological fixed points in the moral domain, although not all of the claims listed above are candidates for such truths, given the account that they offer. By contrast, McPherson (2009) argues that appeals to (alleged) Moorean facts face greater obstacles in the moral domain than elsewhere, while Dreier (2007: 241–2) takes issue with the tendency of “moral absolutist philosophers” to see themselves as defenders of common sense against the bizarre, on a par with traditional epistemologists who defend our ordinary knowledge claims about the external world against radical skepticism.

Although the methodological default assumption adopted here (i.e. that some moral claims like those enumerated above are known to be true) is a substantive one, it is also much weaker than those that are sometimes made. For example, Enoch (2018) claims that the default position in metaethics is *non-naturalist realism* (see especially pp. 30–3, “Obviously, the default position”), a view that he characterizes as entailing that moral facts are “abstract, outside of space-time, causally inert [and] utterly independent of human responses and attitudes” (p. 35).

Moral knowledge
first break
when

that the sources of non-moral, empirical knowledge are also sources of moral knowledge. Generally speaking, even if the *standards* for knowing a proposition are the same across different domains, the *sources* of knowledge might vary greatly from domain to domain. For example, it is plausible that we can arrive at mathematical knowledge *via* a priori reasoning, although it is obvious that we cannot similarly arrive at geographical knowledge via such reasoning. Suppose then that a priori reasoning is a source of mathematical but not geographical knowledge. Even on that assumption, it does not follow that the standards or conditions one must meet in order to count as knowing a mathematical proposition differ from the standards or conditions that one must meet in order to count as knowing a geographical proposition. In both domains, knowledge might consist of (for example) sufficiently reliable true belief, or belief that is sufficiently *safe* (cf. Williamson 2000), or justified true belief that satisfies an additional “anti-Gettier” condition, or warranted true belief (in the sense of Plantinga 1993), or a belief that tracks the truth (cf. Nozick 1981), or....

C1.P22 Notice that this is a point on which we should expect even the moral skeptic, who denies that we have any moral knowledge, to agree. In fact, it seems that the skeptic should insist that the standards for moral knowledge do not differ in any significant way from the standards for knowledge of other subject matters. For the claim that we lack moral knowledge is potentially much less interesting if the sense in which we fall short of moral knowledge involves our falling short of standards that differ from those required for knowledge of other subject matters.

C1.P23 The second assumption concerns the relationship between knowledge and (epistemically) justified belief. I will assume that, if one knows that p, then one is justified in believing that p. (Equivalently: if one knows that p, then it is not the case that one is not justified in believing p.) This assumption is widely accepted, including by many philosophers who differ significantly in their substantive accounts of knowledge and justification. However, it is a substantive assumption that is not completely uncontroversial.⁹ Of course, on the traditional view according to which knowledge can be (partially) *analyzed* in terms of justified true belief, it follows immediately that one is justified in believing anything that one knows. However, even if we attempt to understand knowledge in other terms, or as conceptually fundamental, the idea that

*Mc Gregor
is a lumber
fisher
Helen.*

⁹ A notable dissenter is Robert Audi (2003: 235–9), who combines a strongly externalist account of knowledge with a strongly internalist account of justification. Audi’s discussion does not engage with the kind of consideration in favor of orthodoxy that I offer in the text above. For a more general critique of Audi’s account of justification, see Williamson, “On Being Justified in One’s Head” (2007).

8 MORAL KNOWLEDGE

one is justified in believing anything that one knows remains compelling. Consider the following assertion: “It’s true that you *know* that it rained this afternoon, but you’re not *justified in believing* that it rained this afternoon!” That assertion seems infelicitous, as does any assertion that replaces “it rained this afternoon” with an arbitrary proposition. The most straightforward explanation for this infelicity is the following: if one knows that p is true, then one is justified in believing that p is true. By contrast, on views on which unjustified beliefs can qualify as pieces of knowledge, we would expect that assertions of the relevant kind would be perfectly in order.

C1.S3

1.3 Towards a More Social Moral Epistemology

C1.P24 Another common thread that unifies many of the arguments and views presented in this book is that our access to moral knowledge has an important social dimension. This theme is most obvious in Chapter 3, “Moral Knowledge from Others.” There, I explore a number of philosophical issues raised by the possibility of arriving at moral views by relying on other people. I defend what I call the Moral Inheritance View, according to which a person whose earliest moral views are inherited from her social environment might very well have substantial moral knowledge even before she is in a position to begin critically reflecting upon or reasoning about those views. On this picture, a significant part of our earliest moral knowledge is due to our natural tendency to adopt beliefs that are held by those around us, or that are presupposed by common practices. More generally, other people are in principle potentially rich sources of moral knowledge. To the extent that we have reservations about the propriety of forming moral views by relying on others—as opposed to through the exercise of our own autonomous judgment—what is legitimate in those reservations does not derive from its being impossible to acquire moral knowledge in this way.

C1.P25 Although the theme is most obvious in Chapter 3, it plays a significant role in other chapters as well. For example, in Chapter 4, “Experience and Observation,” I argue that the phenomenon of moral testimony has important implications for the possibility of confirming moral views by non-moral observations (see especially section 4.3). I also argue that the fact that one is a member of a moral community, and thus is in a position to compare the moral opinions of others with one’s own, can contribute to moral knowledge, not only by affording evidence for or against those opinions, but in a more subtle way as well: by providing feedback that can serve to *condition* or

*What counts
as evidence?*

calibrate one's capacity for judgment so that future exercises of that judgment are more likely to deliver knowledge (see especially section 4.4). In Chapter 5, lessons drawn from the social dimensions of moral knowledge play a role in my attempt to solve Gilbert Ryle's puzzle about "forgetting the difference between right and wrong."

C1.P26 The fact that our access to moral knowledge has an important social dimension is both bad news and good news when it comes to our ability to acquire and maintain such knowledge. When other people hold mistaken moral views, this can amount to misleading evidence, and the existence of such evidence can undermine moral knowledge, or make such knowledge more difficult to acquire or maintain than it otherwise would have been (see sections 4.3 and 5.6). In the moral domain, a particularly interesting and potentially important source of misleading evidence is the following: people who either are or seem to be morally upstanding people *in other respects* might engage in a practice or behavior that is in fact seriously morally wrong, in a social context in which its wrongness is not generally recognized. From the perspective of a member of the society who is currently uncertain that the practice is wrong, the fact that a significant number of people who are morally upstanding in other respects regularly engage in the practice might constitute strong (albeit misleading) evidence that the practice is *not* wrong. This mechanism for generating misleading evidence can pose formidable practical difficulties for would-be moral reformers, who seek to convince other members of their society that some common behavior or practice is seriously wrong (see section 4.3.5).

C1.P27 I emphasize the idea that our access to moral knowledge has a significant social dimension not only because I think that it is important, but also because I think that it has been at least somewhat underemphasized within moral epistemology. Although it is clear that the contents of our moral views frequently concern relations that obtain between different people, the processes by which we arrive at those views are often depicted as though they were solitary enterprises. In part, this tendency within moral epistemology might simply reflect (or be a special case of) a historical tendency to neglect the social aspects of knowledge within epistemology more generally.¹⁰ A significant portion of the history of Western epistemology since Descartes

¹⁰ In the words of Alvin Goldman, arguably the most prominent living epistemologist:

Until recently, epistemology...was heavily individualistic in focus. The emphasis was on evaluating doxastic attitudes (beliefs and disbeliefs) of individuals in abstraction from their social environment. The result is a distorted picture of the human epistemic situation, which is largely shaped by social relationships and institutions. (2015: 1)

10 MORAL KNOWLEDGE

has been concerned with questions about how to respond to extremely radical forms of skepticism, for example, about whether and how someone can know anything beyond the present contents of her own mind. Unsurprisingly, such a focus tends to divert attention from epistemological questions that presuppose the existence of a rich social world of which we are members, questions that it is tempting to see as less basic or fundamental. In addition to this traditional concern with skepticism, many of the other topics that dominated the attention of mainstream epistemologists in the second half of the twentieth century—for example, the correct analysis of “S knows that p,” or the debates between foundationalists and coherentists and between internalists and externalists—similarly tended to encourage a focus on the individual knower in abstraction from the social environment that she occupies. In the final years of the twentieth century and the opening decades of the twenty-first, this situation has begun to change. We now see, in mainstream epistemology, sustained attention to topics such as the nature of testimony and the epistemology of “peer disagreement.”¹¹ Like many others, I think that this is a positive development, and I hope that this book might serve to further encourage this tendency in the branch of epistemology concerned specifically with our cognitive relationship to morality.

C1.P28 I suspect that there is a special reason why the social aspects of knowledge have been underemphasized within moral epistemology in particular. For many moral philosophers, the *method of reflective equilibrium* is the correct account of how moral inquiry should ideally be conducted, and thus stands at the very center of moral epistemology. At least as it is often presented and understood, the method naturally encourages an individualistic picture of moral inquiry: in principle, an individual could flawlessly execute the method while sitting alone at her desk.¹² Although the method of reflective equilibrium embodies genuine and important insights, I will argue that it does not have the kind of centrality for moral epistemology or moral inquiry that has been claimed for it. Exploring its virtues and its limitations is the central task of the next chapter.

¹¹ For an overview of developments in this area, see Goldman and Blanchard (2016); for a representative collection of work, see the anthology edited by Goldman and Whitcomb (2011).

¹² As we will see, there are at least some understandings of the method on which this is *not* the case; for discussion, see section 2.4.

C2

2

Reflective Equilibrium, Its Virtues and Its Limits

C2.S1

2.1 Introduction

C2.P1 If asked for an account of their own methodology, or for an account of the way in which we can arrive at justified views about the subject matter with which they are professionally concerned, many moral philosophers would appeal to *the method of reflective equilibrium*.¹ Indeed, prominent moral philosophers sometimes suggest that when it comes to moral inquiry, the method of reflective equilibrium is, in effect, the only game in town. Thus, according to Michael Smith, it is among the “platitudes” about morality that properly conducted moral inquiry has “a certain characteristic coherentist form” of a kind that was given systematic articulation by John Rawls in his seminal discussion of the method (Smith 1994: 40–1). Similarly, Thomas Scanlon says:

C2.P2 it seems to me that this method, properly understood, is in fact the best way of making up one’s mind about moral matters and about many other subjects. Indeed, it is the only defensible method: apparent alternatives to it are illusory.

C2.P3

(2002: 149)²

C2.P4 Kagan (1998: 16, 306) suggests that, in practice, something very much like the method of reflective equilibrium is accepted by anyone doing normative ethics, whether they realize it or not. Harman (1977: 79) argues that a fundamental division among moral philosophers is that between philosophical naturalists and practitioners of “autonomous ethics,” but claims that the method of reflective equilibrium is the shared method of both.

¹ See, e.g., Boyd 1998; Daniels 1996, 2003; DePaul 1987, 1998, 2006; Dworkin 1996: 119; Ebertz 1993; Harman 1977, 2004; Huemer 2005: 117; McMahan 2000; Rawls 1971, 1993, 1999, 2001; Scanlon 2002, 2014; Smith 1994; Sturgeon 1986: 74, 2002: 184–5; Tersman 2018; and Wallace 1998: 113–14.

² Compare DePaul (2006: 616) who argues that, when it comes to moral inquiry, “there is simply no reasonable alternative to reflective equilibrium.”

12 MORAL KNOWLEDGE

C2.P5 The popularity of the method of reflective equilibrium should not obscure the diversity of ways in which the method is understood by those who profess their allegiance to it. Thus, some view the method as a kind of coherentist account of justification (Daniels 1996, 2003; Brink 1989; DePaul 1987; Smith 1994; Sayre-McCord 1996), while others view it as a species of foundationalism (Harman 2004; McMahan 2000). Some view the method as a rival to moral intuitionism (Daniels 1996: 26, 83) while others think that the two are perfectly compatible or even complementary (Huemer 2005:117; cf. Audi 1997: 49–58). Some think that the domains that can be fruitfully investigated by the method are limited to those that admit of investigation from the armchair, while others insist that the method of reflective equilibrium is the method of the empirical sciences as well (see especially Boyd 1988: 199–200, 207).

*Rawls
LT is a
method
yelp
it's
what
moral*

Some think that the method descends from Aristotle (Nussbaum 1990:172–6), while others regard it as an invention of the twentieth century, and credit its discovery to Nelson Goodman (1953) and John Rawls (1971). Most view the method as a procedure for figuring out what one *should* believe, but its most famous and influential advocate, Rawls, often wrote of it as a procedure whose purpose is to unearth descriptive, psychological facts about our underlying “moral conceptions” (1971: 48–51; 1974: 288–91). Similarly, most think that the “considered judgments” on which the method operates are the considered judgments of the individual who is employing it, but others think that the only admissible considered judgments are those that are held in common by some larger group (see e.g. Smith 1994:40–1, following Rawls 1951).

C2.P6 There are, then, significant differences in how the method is understood by its proponents, some of which we will be concerned with below. Underneath these differences, however, lie common themes and commitments. At this point, it will be helpful to have a clear statement of the method that articulates some of these common themes and commitments. For this purpose, the recent account provided by Scanlon (2014: 76–7) is particularly useful, and worth quoting at length:

*Meth
it's
aq.*

C2.P7 In broad outline, the method of reflective equilibrium can be described as follows. One begins by identifying a set of considered judgments, of any level of generality, about the subject in question. These are judgments that seem clearly to be correct and seem so under conditions that are conducive to making good judgments of the relevant kind about this subject matter. If the subject in question is morality, for example, they may be judgments about the rightness or wrongness of particular actions, general moral principles, or judgments about the kinds of considerations that are relevant to determining

the rightness of actions... The method does not privilege judgments of any particular type—those about particular cases, for example—as having special justificatory standing.

C2.P8 The next step in the method is to formulate general principles that would “account for” these judgments. By this Rawls means principles such that, had one simply been trying to apply them, rather than trying directly to decide what is the case about the subject at hand, one would have been led to this same set of judgments. If, as is likely, this attempt to come up with such principles is not successful, one must decide how to respond to the divergence between these principles and considered judgments: whether to give up the judgments that the principles fail to account for, to modify the principles, in hopes of achieving a better fit, or to do some combination of these things. One is then to continue in this way, working back and forth between principles and judgments, until one reaches a set of principles and a reflective equilibrium.

*What is the
diff. between
public justification
and
reflective
equilibrium?*

C2.P9 This account of the method is broadly representative of the way in which the method has been understood by its proponents within moral philosophy.³ For example, here is a theme Scanlon emphasizes that would, I think, be generally accepted by proponents of the method of reflective equilibrium in the moral domain:

C2.P10 **COHERENCE:** In determining which moral views to hold, one should attempt to make one’s moral views more coherent and systematic.

C2.P11 In this context, the pursuit of greater coherence includes, but is not limited to, the elimination of conflicts among one’s moral views: the relevant notion of coherence is richer and more demanding than mere logical consistency. A set of moral views that consisted exclusively of views that did not overlap in content or in what they entailed would be perfectly consistent but would not on

³ See, for example, Daniels (1996: 48–9, 2011); DePaul (1998: 294–6 and 2006: 599–602); Harman (2004: 3); McMahan (2000: 100–1); Rawls (1971: 48–51, 1999: 289, 1980: 534, 2001: 29–31); Sayre-McCord (1996: 141); Sturgeon (2002: 184–5; and Tersman 2018: 2. Compare Goodman 1953: 63–4, Lewis 1983: x–xi), and Scanlon (2003): 140–1.

More generally, when concreteness is called for in what follows, I will often focus on Scanlon’s (2014) discussion and defense of the method. In addition to the representativeness of his characterization, there is a second reason why this focus seems appropriate: Scanlon’s defense of the method is at least as sophisticated as any that has yet been offered. It is, for example, highly sensitive to, and resourceful in addressing, the kinds of concerns about the method that have been raised repeatedly since its popularization among moral philosophers as a result of Rawls’ (1971) influential presentation.

14 MORAL KNOWLEDGE

that account have a high degree of coherence. In order for a set of views to have a high degree of coherence, its members must positively support one another. Such support occurs, for example, when an independently plausible general moral principle “accounts for” or explains the truth of a moral judgment about a particular case that is itself independently plausible.

C2.P12 A second theme emphasized by Scanlon that would be generally accepted by proponents of the method is this:

C2.P13 NO PRIVILEGE: In revising one’s moral views in the course of pursuing greater coherence and systematicity, one should not privilege any view over any other simply in virtue of its level of generality.

C2.P14 Suppose, for example, that one realizes that a general moral principle that one has accepted up until now is in conflict with one’s judgment about the moral status of a token action. At this level of abstraction, there is simply no answer to the question of whether one should favor the judgment about the token action or the judgment about the general principle. The fact that one judgment concerns a general principle while the other concerns a token action is not itself a reason for resolving the conflict in one way rather than the other.

C2.P15 In the discussion that follows then, I will assume that accepting both COHERENCE and NO PRIVILEGE is a necessary condition for accepting the method of reflective equilibrium.⁴ However, we should not assume that accepting these two ideas suffices for accepting the method. For COHERENCE and NO PRIVILEGE might be accepted by a very wide range of philosophers, including some whose views about moral epistemology and moral inquiry differ radically from those held by Scanlon, Rawls, and other paradigmatic proponents of the method. For example, consider a hypothetical philosopher whom I will refer to as the *Caricaturized Moral Intuitionist*. The Caricaturized Moral Intuitionist holds that our ability to arrive at moral knowledge depends

⁴ Although I take this assumption to be well-motivated by relevant texts, it would not be universally accepted. In particular, in his “In Defense of Reflective Equilibrium” (2013), Walden contends that “It is a mistake to try to give a positive characterization of the method of reflective equilibrium” (p. 244). Rather, acceptance of the method is best understood as acceptance of a “negative thesis” (p. 255) to the effect that nothing substantive can be said in advance of an inquiry about what inputs, methods, and goals are appropriate for that inquiry. While I think that this methodological thesis is an interesting one, I don’t think it’s plausible to identify it with the method of reflective equilibrium. In addition to failing to accord with the characterizations of the method offered by its leading proponents (see the specific passages referenced in footnote 15 for textual substantial of this point), notice that on this usage “the method of reflective equilibrium” does not actually refer to a *method*—and the connections with *reflection* and *equilibrium* also seem to be lost. Terminology aside, I note here that none of the criticisms that I offer of the method in what follows, nor any of the virtues that I attribute to it, have any direct bearing on the negative thesis defended by Walden, at least as far as I can tell.

entirely on the operation of an occult, *sui generis* faculty of moral intuition, and no account of moral inquiry that neglects to mention the central role of this faculty could possibly be adequate. On this view, just as it is the job of the more familiar five senses to deliver information about the empirical world, so it is the job of the faculty of moral intuition to deliver information about the moral realm. The Caricaturized Moral Intuitionist holds that this faculty, like the more familiar five senses, is a *fallible* source of information, although it is generally reliable when used in hospitable circumstances and regularly provides us with knowledge of both general moral principles and particular moral claims.⁵

C2.P16 Notice that the Caricaturized Moral Intuitionist might very well accept both COHERENCE and NO PRIVILEGE. First, she might very well accept that it is important to seek coherence and systematicity among our moral views. Precisely because she regards the faculty of moral intuition as fallible and susceptible to error, she thinks that the attempt to make our moral views more coherent serves as an important check on its deliverances, and a way of weeding out mistakes that have resulted from previous misfirings. (Compare the way in which seeking coherence among the views of our surroundings that we arrive at *via* sense perception can play a role in weeding out mistaken perceptual judgments.) Moreover, the Caricaturized Moral Intuitionist might very well see the pursuit of greater coherence and systematicity as a way of *positively extending* the moral knowledge that is delivered by our faculty of moral intuition. She conceives of this process as analogous to the process by which scientific theorizing can extend the empirical knowledge that is provided to us by the more familiar five senses. Her view is not that reasoning aimed at the achievement of coherence cannot extend our moral knowledge; rather, it is that we would not be in a position to attain any moral knowledge at all if not for the operation of the faculty of moral intuition.

C2.P17 Similarly, the Caricaturized Moral Intuitionist might happily accept the idea that no moral judgment is privileged over another simply in virtue of its level of generality. Perhaps her view is that while the faculty of moral intuition can deliver knowledge of both moral principles and more specific moral

⁵ Recall from Chapter 1 Mackie's claim that if we were aware of moral facts, "it would have to be by some special faculty of moral perception or intuition, utterly different from our ordinary ways of knowing everything else" (1977: 38). As I am imagining her, the Caricaturized Moral Intuitionist agrees with Mackie that it is a necessary condition of our having any moral knowledge at all that we have some such special faculty; unlike Mackie, however, she holds that we are fortunate enough to have been endowed with such a thing. In the terminology of 1.1 (cf. fn. 4), the Caricaturized Moral Intuitionist subscribes to an "extravagant" moral epistemology, while it's natural to understand the reflective equilibrium theorist as subscribing to an "economical" one.

16 MORAL KNOWLEDGE

claims, it is fallible with respect to both, and no more likely to be mistaken about one as opposed to the other. For this reason, she might hold that it would be a methodological mistake to privilege one moral judgment over another simply in virtue of its level of generality.

C2.P18 Thus, the Caricaturized Moral Intuitionist might very well accept both COHERENCE and NO PRIVILEGE. Nevertheless, she would adamantly deny that the method of reflective equilibrium is an adequate characterization of the epistemology of morals. When she reads accounts of moral inquiry offered by proponents of the method, these accounts strike her as at best radically incomplete. Indeed, she holds that we would have a great deal of moral knowledge even if we never engaged in anything like reflective equilibrium reasoning, and that on those occasions when such reasoning does deliver moral knowledge, this process is typically made possible by the fact that we have a substantial amount of moral knowledge that is provided by another source. By her lights, an account of how we acquire moral knowledge that neglects to mention the role of this other source is little better than an account of our empirical knowledge that leaves out the essential role played by sense perception.

C2.P19 The Caricaturized Moral Intuitionist offers us a badly inadequate moral epistemology, for we have no special faculty of moral intuition. Nevertheless, her attitude toward the method of reflective equilibrium is instructive. Like her, we should accept both COHERENCE and NO PRIVILEGE. But we should also agree with her that reflective equilibrium reasoning is less central to moral epistemology than it has sometimes been taken to be, and that this is true regardless of whether our concern is with the cognitive practices of ordinary moral agents or with those of more theoretically-minded philosophers. In particular, both ordinary moral agents and moral philosophers typically have a substantial amount of moral knowledge that does not derive from the method of reflective equilibrium. Moreover, on those occasions when the pursuit of coherence and systematicity does lead to an improvement in one's moral views (either by leading one to jettison mistaken moral views or by extending one's knowledge), this process is typically made possible by a background of substantial moral knowledge that is already in place; in the absence of moral knowledge from other sources, reflective equilibrium reasoning would be poorly suited to delivering such knowledge.

C2.P20 Two philosophically significant implications of this picture are the following. First, because standard accounts of the method of reflective equilibrium tend to neglect or at least obscure the role played by prior moral knowledge from other sources in the reasoning process, such accounts tend to *under-characterize*

good reasoning about morality. To the extent that such accounts are offered as accounts of how we should make up our minds about moral questions, they stand in need of amendment or at least supplementation, for the descriptions that they offer are consistent with paradigmatically bad reasoning as well as with paradigmatically good reasoning.

C2.P21 Secondly, to the extent that there is a genuine philosophical puzzle about “where moral knowledge comes from,” or how we could acquire such knowledge in the first place, “the method of reflective equilibrium” is not a good candidate for an answer to that puzzle. This is because the capacity of reflective equilibrium reasoning to deliver new moral knowledge typically depends on the reasoner’s already having substantial moral knowledge from other sources.

C2.P22 Perhaps some philosophers have understood the method of reflective equilibrium liberally enough so that anyone who accepts both COHERENCE and NO PRIVILEGE counts as accepting the method, including the Caricaturized Moral Intuitionist. While this strikes me as a suboptimal way of dividing up the terrain, I have no desire to argue over terminology. Especially if we allow for such an inclusive understanding of the method, we should distinguish between *modest* and *ambitious* interpretations of the method and its epistemological role. On a modest interpretation, anyone who endorses both COHERENCE and NO PRIVILEGE counts as endorsing the method. On the other hand, those who accept an ambitious interpretation of the method and its epistemological role tend to see what justification we ultimately have for our moral views as in some way arising out of, or dependent upon, reflective equilibrium reasoning. On an ambitious interpretation, the method of reflective equilibrium is a potential answer to an epistemological question about morality: how can we arrive at knowledge or justified beliefs about morality in the first place? Like the Caricaturized Moral Intuitionist, I accept the method of reflective equilibrium when it is interpreted modestly but reject it when it is understood ambitiously.

C2.P23 I will not engage in the exegetical exercise of attempting to determine which philosophers have accepted an ambitious interpretation of the method. But it seems pretty clear that at least some of its greatest proponents have tended to understand it in this way. Consider, for example, Nelson Goodman’s (1953: 61–3) seminal presentation of the method in the context of discussing the justification of logic and principles of inductive reasoning. There, Goodman explicitly champions the method as a preferable *alternative* to existing accounts of justification about the relevant subject matters, accounts that he takes to be unsatisfactory. Notably, Goodman’s discussion is frequently

18 MORAL KNOWLEDGE

cited with approval by moral philosophers who endorse the method as the correct account of justification in the moral domain.⁶ More recently, Scanlon (2014) claims that “the *only* way we have of establishing the truth of normative judgments is through direct, piecemeal application of the method of reflective equilibrium” (pp. 122–3, emphasis mine). Moreover, Scanlon explicitly presents the method of reflective equilibrium as an adequate answer to Mackie’s (1977) challenge of how we are able to achieve moral knowledge at all, given our apparent lack of any special faculty of moral perception or intuition (2014: 69–72). Both of these aspects of his discussion are at least strongly suggestive of what I have called an ambitious understanding of the method and its epistemological role.

C2.P24 According to the working hypothesis endorsed in Chapter 1, any source of ordinary empirical knowledge is also a potential source of moral knowledge. Given that one of the ways in which we acquire ordinary empirical knowledge is by attempting to make our beliefs about the world more coherent and systematic, the working hypothesis does not stand in any tension with the method when the method is interpreted modestly. However, when the method is interpreted ambitiously, the picture of moral epistemology that results *does* conflict with the picture suggested by the working hypothesis. For on an ambitious interpretation, the method is *the* source of our moral knowledge, or the primary source; and this idea conflicts with the alternative picture suggested by the working hypothesis, on which the sources of such knowledge are a diverse lot.

C2.P25 Let me conclude this section by summarizing the views about the method of reflective equilibrium for which I will be arguing in the rest of this chapter. On the one hand, I will argue that the method embodies genuine and important insights. These insights include the following:

C2.P26 (1) In moral inquiry, *we never start from scratch*. When we engage in moral inquiry, we typically come to the table already holding any number of substantive moral views, and there is no general requirement to set aside or bracket such views for purposes of the inquiry. On the contrary, exemplary moral reasoning typically makes heavy use of such convictions. It is not rationally required, or even desirable, to attempt to conduct moral inquiry from some more austere starting point (e.g. from some purely formal moral principle or principles, as suggested by Hare 1973, among others).

⁶ For an extended discussion of Goodman on the method, along with references, see Kelly and McGrath (2010): 329–34, 352.

C2.P27 (2) We can often improve our moral views by attempting to make them more coherent and systematic. Indeed, given that we have at least some substantial moral knowledge, the process of seeking greater coherence and systematicity among our moral views is likely to be a fruitful way of extending this knowledge.

C2.P28 (3) Contrary to what many philosophers have supposed, considerations having to do with the levels of generality of our moral judgments are not of any deep methodological significance.

C2.P29 (4) One consequence of (3) is that, when we pursue greater coherence and systematicity among our moral views, no moral judgment should be privileged over any other simply in virtue of its level of generality. For example, in cases in which an otherwise plausible general moral principle conflicts with an otherwise plausible specific claim about morality, there are no general norms which would favor one over the other simply in virtue of its level of generality.

C2.P30 (5) A further consequence of (3) is that there is no specific level of generality that is had in common by those propositions that constitute the proper starting point for moral inquiry. For example, and contrary to what some have supposed, it is not as though the proper starting point for moral inquiry consists exclusively of general moral principles as opposed to particular moral judgments, or vice versa.

C2.P31 Again, I take these to be insights of the method. But I will also argue that the method does not have the kind of epistemological significance that has sometimes been taken to have. More specifically, I will argue for the following claims:

C2.P32 (6) Neither the pursuit nor the achievement of reflective equilibrium is *necessary* for having substantial moral knowledge. An ordinary moral agent might have substantial moral knowledge even if she never engages in the pursuit of reflective equilibrium.

C2.P33 (7) Following the method is not *sufficient* for attaining moral knowledge, nor is it sufficient for attaining moral views that have any significant positive epistemic status. Following the method impeccably, and arriving at a state of reflective equilibrium in the canonical way, is perfectly consistent with having unreasonable moral views.

C2.P34 (8) Because of (6) and (7), the method of reflective equilibrium is not tenable when it is understood as an account of the circumstances in which our moral views are justified. Nor is the method of reflective equilibrium tenable when taken as a normative account of how we ought to make up our minds about moral questions. Given standard characterizations, the method is consistent with a great deal of good reasoning about morality, but it is equally

20 MORAL KNOWLEDGE

consistent with a great deal of bad reasoning about morality. Thus, on its standard characterizations, the method of reflective equilibrium undercharacterizes good moral reasoning.

C2.P35 (9) On its most defensible interpretations, the method of reflective equilibrium is characterized in explicitly normative terms: for example, the judgments which it treats as evidence are pieces of moral knowledge, or at least, things that we justifiably believe about morality *prior* to engaging in the method of reflective equilibrium. An upshot of this is that, on its most defensible interpretations, the method of reflective equilibrium takes for granted that we have some moral knowledge (or at least, justified moral beliefs) that we do not arrive at *via* application of the method itself. If we have any moral knowledge, we have some moral knowledge that is not due to engaging in reflective equilibrium reasoning.

C2.P36 (10) Because the ability of reflective equilibrium reasoning to deliver new moral knowledge generally depends on our already having substantial moral knowledge from other sources, the method is not a promising answer to the question of how we are able to acquire moral knowledge in the first place.

C2.P37 At the outset of this section, I noted the impressive popularity of the method among leading moral philosophers, including those who differ widely in their other commitments. Given this impressive popularity, it seems that someone who argues that the method does not have the kind of epistemological significance that has often been claimed for it inherits an additional dialectical burden: she needs to offer a plausible explanation of *why* the method has seemed to so many moral philosophers to be a compelling account of the kind of reasoning in which they routinely engage, a kind of reasoning that frequently seems to deliver genuine insights into the moral domain.⁷ It is hard to believe that so many moral philosophers could be radically deceived about what they have been doing all along. I attempt to discharge this burden in Section 2.7, “The Method and the Moral Philosopher.”

C2.S2

2.2 The Method Is Too Demanding

C2.P38 In my estimation, the most straightforward and compelling reason for thinking that engaging in the method of reflective equilibrium is *not necessary* for

⁷ As Sayre-McCord (1996: 142) says: “Recommending the method right off is the fact that it seems, in some sense, simply to work.”

acquiring moral knowledge is this: if there is any moral knowledge at all, the amount of it that is had by ordinary people is out of proportion to the frequency with which they engage in the activity of pursuing reflective equilibrium among their moral views, and the role that doing so plays in their lives.

C2.P39 Consider the claims that *slavery is unjust* and that *rape is wrong*. If there is anything that we might reasonably expect moral skeptics and non-skeptics to agree upon, perhaps it is this: these claims are not pieces of *esoteric knowledge*. If they are known by anyone, they are known by many. Suppose first that some variety of moral skepticism is true, so that no one knows that slavery is unjust (perhaps because *slavery is unjust* is simply not the right kind of thing to be the object of propositional knowledge). If no one knows that slavery is unjust, then *a fortiori*, no one knows that slavery is unjust on the basis of reflective equilibrium reasoning. Suppose on the other hand that moral skepticism is false, and that some people do know that slavery is unjust. Consider the description of reflective equilibrium reasoning offered by Scanlon, above. How plausible is it that all (or even most) of those who know that slavery is unjust know it on the basis of such reasoning? The claim that this kind of reasoning is responsible for the average person's knowledge that slavery is unjust seems to involve an unrealistic, hyper-intellectualized picture of the basis for such knowledge. And the same point seems to hold with respect to canonical statements of the method due to Rawls and others.⁸

*moral
skepticism
Nothing
be the
true or
propositional
knowledge*

C2.P40 More generally, the kind of knowledge that many proponents of the method of reflective equilibrium take it to deliver seems like a poor candidate for knowledge that is available only to relatively reflective individuals. For example, Scanlon (2014) touts the method as the way in which we arrive at knowledge of normative truths about reasons, a category that he takes to include the most fundamental moral truths. In the opening pages of that work, he provides the following examples of the kinds of truths that he has in mind:

- C2.P41 (1) For a person in control of a fast-moving automobile, the fact that the car will injure and perhaps kill a pedestrian if the wheel is not turned is a reason to turn the wheel.
- C2.P42 (2) The fact that a person's child has died is a reason for that person to feel sad.

⁸ Again, see the passages referenced in footnote 15 above. On the general point at issue here, compare Timmons (1999: 237) and Wedgwood (2007: 244). After noting that the method seems overly demanding when understood as an account that applies to ordinary moral agents, both Timmons and Wedgwood suggest that for that reason it is perhaps better interpreted as an account of theory acceptance that applies to moral theorists or philosophers. I consider this suggestion below.

22 MORAL KNOWLEDGE

C2.P43 (3) The fact that it would be enjoyable to listen to some very engaging music, moving one's body gently in time with it, is a reason to do this, or to continue doing it.

C2.P44 (2014: 2)

C2.P45 On the face of it, these examples seem like unlikely candidates for things that we know on the basis of any even moderately sophisticated reasoning. Notably, when Scanlon describes and defends the method in a later lecture, much of that lecture is devoted to a compelling description of the kind of holistic plausibility reasoning by which set theorists justify their beliefs in the less obvious axioms of set theory (including the axiom schema of replacement) which he also takes to be an application of the method (2014: 72–6). However, the kind of relatively subtle reasoning by which a set theorist might justify her belief in the axiom schema of replacement does not seem like a good model for the way in which an agent might come to justifiably believe that he should turn the wheel of his car when he realizes that there is a pedestrian directly in his path. If we look to mathematics for natural analogues to seemingly obvious and immediately grasped propositions about normative reasons like (1)–(3), better candidates for this role would seem to include the obvious and immediately grasped truths of elementary arithmetic such as $2+2=4$. But of course, at least on the face of it, the method of reflective equilibrium does not seem to be a particularly plausible account of how we know that $2+2=4$, either: it is no accident that our knowledge of such truths has historically inspired epistemological accounts that have appealed to notions like “self-evidence” (whatever the ultimate defects or merits of such accounts), as opposed to the more holistic picture of justification that is associated with the reflective equilibrium model.

C2.P46 In short, the view that whatever moral knowledge there is is knowledge that is arrived at via reflective equilibrium reasoning seems to predict that there is relatively little moral knowledge, and that what we have of it is disproportionately had by unusually reflective individuals. By my lights, that this is the actual situation is much less plausible than that either (i) there is relatively widespread moral knowledge, or (ii) there is no moral knowledge at all.

C2.P47 I will assume, then, that it is not a necessary condition for knowing that a moral claim is true either that that belief was arrived at or that it is sustained by reflective equilibrium reasoning. Even if you've never engaged in reflective equilibrium reasoning with respect to a moral belief that you hold, it does not follow that your belief falls short of knowledge.

C2.P48 Indeed, on my view, the plausibility of the claim that ordinary moral agents have at least some substantive moral knowledge (regardless of whether they've

engaged in reflective equilibrium reasoning or systematic moral theorizing) provides a compelling reason to believe one of the characteristic methodological claims of the reflective equilibrium theorist. Recall the idea that, in moral inquiry, *we never start from scratch*. According to this idea, whenever we engage in moral inquiry, we come to the table with moral convictions, and it is neither rationally required nor desirable for us to attempt to conduct the inquiry from some more austere starting point (e.g. one consisting exclusively of some purely formal principle or principles). One line of thought that supports this methodological claim runs as follows. Consider what we might call the *knowledge platitude*:

C2.P49 **The Knowledge Platitude:** If you know something that is relevant to a question that you are trying to answer, then you should take that information into account in arriving at a view.

C2.P50 The knowledge platitude is highly plausible.⁹ Now, on the assumption that people typically have at least some substantive moral knowledge prior to engaging in reflective equilibrium reasoning or systematic moral inquiry, it follows from the knowledge platitude that they should take that knowledge into account if and when they do engage in such inquiry. (It's not as though one loses whatever substantive moral knowledge one has simply by engaging in systematic moral inquiry.) Bracketing or setting aside substantive moral claims that one knows to be true is bad methodology, because it amounts to throwing away relevant information about the subject matter. Thus, the fact that a person might have substantive moral knowledge even without having engaged in systematic moral inquiry provides a good reason to believe that the reflective equilibrium theorist is correct in holding that the proper starting point for such inquiry need not be devoid of substantive moral claims.¹⁰

⁹ Notice that the knowledge platitude is not specifically concerned with moral knowledge, but with knowledge more generally. It seems clear enough that, if, for example, you were trying to make up your mind about which candidate is most likely to win some upcoming election, you should take into account any relevant information of which you have knowledge. On the face of it, there does not seem to be any reason to think that things are different when it comes to moral questions.

For more extended defenses of the knowledge platitude, see Kelly (2008) and especially Williamson (2000, chapter 9). The latter argues at length that the relevant norm remains compelling even if we abandon the so-called “KK thesis” and admit that we often know things without knowing that we know them.

¹⁰ Although I take the line of thought offered in this paragraph to amount to a compelling vindication of a major theme of proponents of the method, this particular way of vindicating the theme will not be available to anyone who interprets the method ambitiously as opposed to modestly. For the vindication depends on the idea that we have substantial moral knowledge prior to engaging in the method, an idea that will be rejected—mistakenly, if what has already been argued is correct—by

24 MORAL KNOWLEDGE

C2.P51 If what has been argued thus far is correct, engaging in reflective equilibrium reasoning is not necessary for moral knowledge. Similar considerations support the conclusion that engaging in reflective equilibrium reasoning is not necessary for having justified moral beliefs. In fact, there are two routes to this conclusion from the kinds of considerations canvassed in this section. In Chapter 1, I endorsed and briefly defended the common assumption that knowing that *p* entails justifiably believing *p*. If that assumption is correct, then, given that engaging in reflective equilibrium reasoning is not necessary for knowing some moral claim, it follows immediately that it is not necessary for justifiably believing that moral claim, either. However, even if we do *not* assume that knowing entails justifiably believing, the kinds of considerations canvassed in this section strongly suggest that one can have justified moral beliefs even in the absence of reflective equilibrium reasoning. Just as claims like *rape is wrong* or *slavery is unjust* seem like poor candidates for esoteric knowledge, so too they seem like poor candidates for being justifiably believed only by unusually reflective people: if they are justifiably believed by some, then they are justifiably believed by many.

C2.P52 Of course, even if it's true that engaging in reflective equilibrium reasoning is not necessary for either knowing or justifiably believing moral claims, it might still be that such reasoning has an important epistemological role to play. For example, it might be that when one impeccably follows the method of reflective equilibrium, this is *sufficient* for the reasonableness of the moral views at which one arrives. Moreover, notice that even if a great deal of our moral knowledge is not attributable to the method of reflective equilibrium, this is perfectly consistent with the claim, made by Scanlon and others, that the method is the best way of making up one's mind about moral questions. (Compare: it might be that the best procedures we have for arriving at knowledge of the empirical world are the procedures of the natural sciences. This is perfectly consistent with the fact that much of what one knows about the empirical world is *not* due to the application of the procedures of the natural sciences.) Finally, it might be that, even if the method of reflective equilibrium plays a limited role in the lives of ordinary moral agents, it is the correct normative account of how moral philosophers or theorists should conduct their inquiries. These are among the claims that I will argue against in what follows.

proponents of the method who understand it ambitiously. Thanks to an anonymous referee for making salient to me the need to clarify the dialectic on this point.

C2.S3 **2.3 If the Method Is Too Weak, Then It
Is Probably Not “The Best Method”**

C2.P53 In the last section, I noted that, although there are good reasons to think that the reflective equilibrium framework does not supply a necessary condition for being justified in holding a moral view, those considerations are perfectly consistent with the claim that the framework does supply a sufficient condition for being justified.¹¹ In the next section, I will argue against this sufficiency claim. But before doing that, I want to put on the table another way of thinking about the method of reflective equilibrium, on which it is not, at least in the first instance, an account of the conditions (either necessary or sufficient) under which our moral beliefs are justified. In particular, I want to look at Scanlon's suggestion that the method as an account of “the best way of making up one's mind about moral matters” and consider how that relates to the sufficiency question.

C2.P54 Let us start by considering a very straightforward sufficiency claim that a theorist might endorse:

C2.P55 (1) A person is justified in holding a moral belief if she holds that belief in a state of reflective equilibrium.

C2.P56 Now I suspect that few proponents of the method would find this principle attractive, because the idea that simply being in a certain state at a given time *suffices* for being fully justified in holding a moral belief fits poorly with the theme, frequently emphasized by proponents of the method, that the *process of seeking or pursuing* reflective equilibrium is as important (if not more so) than actually being in the state itself.¹² If (1) is true, then someone who ended up in a state of reflective equilibrium in virtue of being hit in the head by a falling tree branch, or in virtue of having been hypnotized, would be fully justified in her moral views so long as she remained in that state. But this seems to be at odds with the idea that it is engaging in a certain *procedure* or *way of reasoning* that is crucial. Indeed, it seems at odds with the idea that the method of reflective equilibrium is a *method*.¹³

¹¹ Several paragraphs in both this section and the next are adapted from Kelly and McGrath (2010). I utilize this material with the permission of my co-author.

¹² For representative statements of this thought, see Sayre-McCord (1996: 142) and Scanlon (2014: 79, 103).

¹³ Thus, the method of reflective equilibrium should be distinguished from any purely “current time slice” principle of epistemic justification. For example, the method should be distinguished from a principle according to which a person is justified in holding a moral belief a time t just in case that belief coheres well with the other beliefs that she holds at t, as well as from a principle according to

26 MORAL KNOWLEDGE

C2.P57 This line of thought might lead one to replace (1) with (2):

C2.P58 (2) A person is justified in holding a moral belief if (i) she holds it in a state of reflective equilibrium, and (ii) she reached that state by reasoning in the prescribed way.

C2.P59 where “the prescribed way” abbreviates the kind of procedure described by Scanlon in the passage quoted in Section 2.1, or something very similar. According to (2), holding a moral belief in a state of reflective equilibrium reached by reasoning in the prescribed way is a sufficient condition for being justified in holding that belief.

C2.P60 In the next section, I will argue that (2) is false. But for the moment, the issue is that of how (2) relates to the following claim:

C2.P61 (3) The method of reflective equilibrium is the best method for arriving at one’s moral views.

C2.P62 Suppose someone wanted to argue that (3) is false. To show that (3) is false, it would *not* be enough to show that a person who flawlessly follows the method could arrive at false moral views, or even that she could arrive at views that are radically mistaken. Because in general, it is *not* a condition of adequacy on a method of inquiry that it is guaranteed to deliver the truth, or even guaranteed not leave us much worse off with respect to the truth than we would have been had never applied it. Certainly, we do not hold our best scientific methods to the relevant standard. For example, contemporary cosmologists utilize relatively rigorous methods to arrive at theories about the history and structure of the physical universe on the basis of the evidence that is available to them. In a world in which the evidence that they have to go on is consistently misleading or unrepresentative, the impeccable application of their best methods to the data will not only fail to deliver the truth but will lead them further and further astray. But it would be a mistake to regard this fact as a good objection to their methods. We should not, I assume, hold a proposed method of *moral* inquiry to a higher standard than those to which we hold our best scientific methods. For this reason, even if there are circumstances in which employing the method of

which a person is justified in holding a moral belief at time t just in case that belief is supported by the other mental states she is in at t, including both her beliefs and her moral intuitions or “seemings.” Notice that putative principles such as these purport to be purely synchronic principles of justification. Thus, they are completely silent on how we should go about arriving at or revising our moral views over time; indeed, they lack any diachronic import at all. By contrast, the accounts offered by Scanlon, Rawls, and the other representative proponents of the method listed in footnote 15 clearly *do* have significant diachronic import.

reflective equilibrium could lead someone into error, even radical error, this would not be a good reason for thinking that (3) is false.

C2.P63 On the other hand, there *would* be a good objection to (3) if it turned out that by impeccably following the method a person could arrive at moral views that it would be *unreasonable* for her to hold, or to moral beliefs that are *unjustified*.¹⁴ For surely, if some method is in fact the best method for investigating the moral domain, and someone employs the method because she recognizes this, then the views at which she arrives by impeccably executing it would not be unreasonable. Thus, if one could arrive at unreasonable moral views by impeccably executing the method of reflective equilibrium, this *would* be a good reason to think that (3) is false.

C2.P64 Against this, it might be argued that requiring that the method not lead to unreasonable views is too stringent, for reasons analogous to those that speak against a requirement that the method not lead to false or even radically false beliefs. For imagine a person who begins with views about morality that are completely unreasonable. Suppose that he pursues and ultimately achieves a state of reflective equilibrium by reasoning flawlessly “downstream” from that rationally defective starting point. If the views at which he arrives are intuitively unreasonable, then it might seem like this should not be held against the method—it might seem like the method cannot be expected to deliver reasonable outputs given unreasonable inputs.

C2.P65 On this account, the goodness of the method of reflective equilibrium as a procedure would be something like the goodness of reasoning in accordance with *modus ponens*. If someone reasons from two unreasonable beliefs to a third belief in accordance with modus ponens, then the third belief might very well be unreasonable as well, but surely this is not a good objection to the practice of reasoning in accordance with modus ponens! Similarly, one might think, it is too much to require that the method of reflective equilibrium not lead to unreasonable beliefs when a person begins from a rationally defective starting point.

C2.P66 Although it is a natural response, this way of thinking about things sets the bar too low. In particular, comparisons such as the one to modus ponens do not do justice to the status that is assigned to the method by (3). If (3) is true, then the method of reflective equilibrium is *the appropriate* method for investigating the moral domain; it is not simply one norm or rule among many others (e.g. “One should seek coherence among one’s moral views”), which is what the comparison with modus ponens suggests. After all, as emphasized in Section 2.1, a philosopher who thinks that the method of reflective equilibrium

¹⁴ Here and below, I will use “justified” and its cognates and “reasonable” and its cognates interchangeably.

28 MORAL KNOWLEDGE

is hopelessly inadequate as a characterization of the way in which we should make up our minds about moral questions might very well agree that we should seek coherence among our moral beliefs. (Consider again the stance of the Caricaturized Moral Intuitionist.) That is, if (3) is true, then the method of reflective equilibrium stands to the moral domain as the cluster of procedures that are employed by (for example) physicists and biologists stand to their respective domains.

C2.P67 To see this point, suppose that, prior to embarking upon the systematic study of fruit flies, you held various baseless and ill-considered opinions about their nature. If you then devoted yourself to the study of fruit flies, and impeccably followed the best scientific procedures we have for arriving at accurate views about their nature, we would expect those earlier baseless opinions to be filtered out or corrected at some stage in the inquiry. In the unlikely event that some of those opinions were among the views that you held after having impeccably followed our best scientific methods, then those beliefs would no longer be unreasonable ones for you to hold. After all, if someone *did* criticize them as unreasonable, you would be in a position to reply as follows:

C2.P68 My views about fruit flies are ones that have withstood the impeccable application of our best methods for arriving at and correcting beliefs about fruit flies. Therefore, whatever else is true of these beliefs (e.g., even if later inquiry should show that they are false), they are not unreasonable views for me to hold as things stand.

C2.P69 This would be an excellent defense. Similarly, if (3) is true—that is, if the method of reflective equilibrium really is the best method for arriving at your moral views—then it would be an excellent defense of the reasonableness of your moral views that they either resulted from or withstood the impeccable application of that method. And therefore, if you could arrive at unreasonable moral views by impeccably following the method, that would be a reason for thinking that (3) is false.^{15,16}

¹⁵ The argument offered here oversimplifies things in one respect. Strictly speaking, that a given method is the best method for investigating a given domain (and is employed because it is recognized as such) is not a sufficient condition for the reasonableness of the views to which it leads. For suppose that we had *no* good methods for investigating a given domain: even our best method is highly unreliable, and known to be so. In that case, it would not be a good defense of the reasonableness of some belief to show that it was sanctioned by the best method. But of course, proponents of the method of reflective equilibrium typically do not think that it is a poor method that nevertheless manages to be the best of a bad lot. More importantly, I believe that once the way in which the method allows for the holding of unreasonable moral beliefs is made clear, it will also be clear that superior accounts are available.

¹⁶ The comparison with *modus ponens* is apt to mislead for another reason. As Harman (1986a) emphasizes, *modus ponens* is a principle of logic, not a norm of belief revision. Moreover, for reasons

C2.P70 Now, proponents of the method typically think that there are significant constraints on admissible starting points; if an inquirer simply sets out from baseless or ill-considered moral opinions she is *not* competently applying the method. In the broadly Rawlsian tradition, this is the idea that the correct starting point for moral inquiry consists of our *considered judgments*. I will explore this idea in some detail in the next section.

C2.P71 But again my main point here concerns the connection between (2) and (3). Recall:

C2.P58 (2) A person is justified in holding a moral belief if (i) she holds it in a state of reflective equilibrium, and (ii) she reached that state by reasoning in the prescribed way.

C2.P73 and

C2.P74 (3) The method of reflective equilibrium is the best method for arriving at one's moral views.

C2.P75 If the argument offered immediately above is compelling, then *if* (2) is false, that casts doubt on (3). In general, correctly reporting that you arrived at your current view about some question by impeccably employing the best method we have for arriving at a view about that question is a sufficient defense of the status of your belief as reasonable or justified. Thus, a person who can correctly report that he arrived at his belief about some moral question by impeccably employing the best method we have for arriving at a view about that question is in a position to offer a solid defense of his belief as reasonable or justified. For precisely that reason, if it turns out that an inquirer can impeccably execute the method yet arrive at moral views that are nonetheless unjustified—i.e. if turns out that (2) is false—then this would give us reason to think that the method of reflective equilibrium is *not* the best method for arriving at moral views—i.e. that (3) is false as well.

given by Harman, what is perhaps the most straightforward and obvious candidate for a norm of belief revision that corresponds to the logical principle—“If you believe that *p*, and if you believe that *if p, then q*, then you should conclude that *q*” is *not* a plausible candidate for being among the genuine norms of belief revision. My own view is that the best candidate for a genuine norm of belief revision that corresponds to the logical principle of modus ponens is one that explicitly invokes the concept of knowledge in characterizing the beliefs from which the reasoning proceeds: “If you know that *p*, and you know that *if p then q*, then you can justifiably conclude that *q*.” This version of the norm corresponds to what I will argue is the most defensible version of the method of reflective equilibrium, one which differs from existing accounts by characterizing the proper starting points for moral inquiry in explicitly normative terms, e.g., in terms of what one already knows or justifiably believes. See Section 2.5 below.

30 MORAL KNOWLEDGE

C2.P76 Thus far, no reason has been offered for thinking that an inquirer could arrive at unreasonable moral views by impeccably executing the method. Indeed, the possibility that the method is in this respect *too weak* might seem at odds with much of the general thrust of my discussion so far—I argued in the previous section that the method is *too demanding* to be a plausible account of the way in which ordinary agents arrive at much of the moral knowledge that they seem to have. But I also think that the method is not demanding enough.

C2.S4

2.4 The Method Is Too Weak

C2.P77 The idea that the method of reflective equilibrium is in some sense too weak has been a recurrent theme among its critics.¹⁷ A frequent thought among such critics is that considerations of coherence cannot bear the kind of justificatory weight that the reflective equilibrium theorist seems to place on them: even achieving a perfectly coherent and stable equilibrium among your moral views would count for little, unless your views had some measure of independent credibility. Scanlon (2014: 82–4) argues that this traditional criticism misses the mark, because it fails to take adequate account of the importance that proponents of the method give to the idea that the proper starting point for moral inquiry consists of our *considered judgments*. Properly understood, the method enjoins us to pursue a coherent equilibrium not among just any moral judgments that we might make, but rather among a select class of such judgments.

C2.P78 In the broadly Rawlsian tradition, your considered judgments are ones that you affirm in circumstances that are hospitable or conducive to judging well. (Of course, different theorists might offer different accounts of the circumstances that constitute good judging conditions.) I want to explore the idea that the proper starting point for moral inquiry consists of our considered judgments carefully, because I share Scanlon's sense that it is crucial to adjudicating the disagreement between defenders of the method and many of its critics. For purposes of illustration, I will focus on what I take to be Rawls' mature account of considered judgments. But I will argue that the conclusions that I reach have more general import, because they generalize to any account of considered judgments that shares certain features.

C2.P79 For Rawls, “considered judgment” is a technical term.¹⁸ Not everything that one believes or judges true, even on reflection, qualifies as a considered

¹⁷ See, for example, Brandt (1979), Lyons (1975), Raz (1982), Setiya (2012), and McPherson (2015).

¹⁸ And indeed, one whose stipulated meaning evolved considerably from work to work. For example, in the early “Outline of a Decision Procedure for Ethics” (1951) a considered judgment must concern actual (as opposed to merely hypothetical) cases (p. 5), and cannot be the object of disagreement

judgment. Rather, considered judgments are judgments of which one is confident (as opposed to uncertain or hesitant), that are issued when one is able to concentrate without distraction on the question at hand (as opposed to when one is “upset or frightened”) and with respect to which one does not stand to gain or lose depending on how the question is answered. In addition, such judgments must be stable over time.

C2.P80 Of course, the point behind the introduction of considered judgments is that “in deciding which of our judgments to take into account, we may reasonably select some and exclude others” (1971: 47). Thus, for Rawls, there are at least two different ways in which a moral conviction can be legitimately discarded: (i) it might fail to qualify as a considered judgment, or (ii) it might qualify as a considered judgment but be eliminated at some later stage in the course of pursuing reflective equilibrium. Because many moral judgments might fail to qualify as considered judgments, a significant amount of filtering might occur even before the process of seeking reflective equilibrium begins.

C2.P81 It is clear that Rawls’ conditions are substantial: they would potentially exclude a significant number of judgments from playing a role in the process. Nevertheless, they also seem quite weak: they could let in judgments that are utterly lacking in rational credibility. For example, given Rawls’ characterization, there is in principle nothing that would preclude the following from qualifying as a considered judgment for someone:

C2.P82 We are morally required to occasionally kill randomly.

C2.P83 There is nothing incoherent about the possibility that someone could confidently and stably subscribe to this judgment, even when he is not upset, frightened, or distracted, and so on. Not only do such conditions fail to preclude the possibility of someone’s having this among his or her considered judgments, but in principle such a proposition might score just as highly along the relevant dimensions as the following proposition does for the average person who believes it:

C2.P84 We are not morally required to occasionally kill randomly.

among “competent persons” (p. 6); both of these requirements are absent from later characterizations. In *A Theory of Justice* (1971), considered judgments concern particular cases; in that work, “considered judgment” is frequently juxtaposed with “general conviction” or “general principle.” However, by the time of “The Independence of Moral Theory” (1974), Rawls was clearly applying the term to judgments of all levels of generality. (See, for example, p. 289: “People have considered judgments at all levels of generality, from those about particular situations and institutions up through broad standards and first principles to formal and abstract conditions on moral conceptions.”) In what follows, I will work with this last and most general understanding of Rawls’ account, following the exegetical example of Scanlon’s “Rawls on Justification” (2002: 141).

32 MORAL KNOWLEDGE

C2.P85 Moreover, in principle, there does not seem to be anything in the Rawlsian conditions that would preclude a person from having any number of other, equally absurd moral judgments among her considered judgments.

C2.P86 Imagine a person who, beginning from a collection of such judgments, reasons flawlessly “downstream” in the canonical way, and manages to arrive at a coherent equilibrium. If her considered judgment that *we are morally required to occasionally kill randomly* survives that process, does it follow that she is justified in thinking this, as we are justified in thinking the opposite, or in thinking that *rape is wrong* or that *slavery is unjust*? That thought seems implausible, and its implausibility suggests the following: the epistemic standing of our convictions that *rape is wrong* or that *slavery is unjust* outstrips anything that might be bestowed on them in virtue of their having successfully survived a process of seeking reflective equilibrium from a starting point that consists of judgments that satisfy Rawls’ conditions. On the face of it, the Rawlsian conditions on considered judgments are not strict enough to allay the concern that the method is too weak, at least when it is understood as a normative account of how we ought to arrive at our moral views.

C2.P87 In defense of the Rawlsian account, it might be objected that the example that I’ve used in attempting to cast doubt on it is unrealistic: no actual person has or would have the claim that *we are morally required to occasionally kill randomly* among her considered judgments. Although the objection is a natural one, I think that it is misguided. In fact, I think it is methodologically crucial to consider examples of this kind, for the following reason. In my view, the weakness of the conditions typically imposed on “considered judgments” is sometimes obscured by the choices of examples that are given when the method of reflective equilibrium is illustrated. The theorist illustrating the method typically proceeds from the first person perspective, and speaks of (for example) “our” considered judgments; she thus selects one of her own considered judgments that she expects her audience to share. This is of course natural enough—certainly, it would be strange for a proponent to illustrate the method by proceeding from a judgment that she does not hold, or which she does not expect her audience to share.¹⁹ But nevertheless, it’s a dangerous procedure. For proceeding in this way runs the risk that what we are responding to, in agreeing that a certain judgment is part of an appropriate starting

¹⁹ In some cases talk of “our” considered judgments might signal commitment to an additional, substantive constraint on the relevant class of judgments: namely, a constraint to the effect that only judgments which are *generally shared* are among the appropriate starting points for pursuing reflective equilibrium. I address the possibility of understanding the method as incorporating this kind of interpersonal constraint below.

point for conducting moral inquiry, is *not* its status as a considered judgment in some stipulated sense, but rather our perception that it has some more significant positive epistemic status: for example, that we have *good reason* to believe it, or even that it is *among the things that we know to be true*. How might one discriminate between these two possibilities? In order to test the claim that it is the fact that the judgment in question is a considered judgment which is doing the work in this context, it is important to consider cases from the third person perspective, in which the starting points of the person pursuing reflective equilibrium are (i) his considered judgments but (ii) *perverse* considered judgments, at least when judged by one's own lights. (That is, judgments which, when judged by one's own lights, are clear cases of non-knowledge, or propositions lacking in rational credibility.) But when we perform this experiment, the idea that the normatively appropriate starting point for any person consists of all and only her considered judgments seems less appealing.

C2.P88 Does the line of criticism offered here presuppose an overly individualistic conception of the method of reflective equilibrium? Although the method is typically understood so that the appropriate starting point for any individual consists of his or her *own* considered judgments, an obvious way to rule out the possibility of highly idiosyncratic judgments such as *one is morally required to occasionally kill randomly* is to impose interpersonal constraints on permissible starting points. For example, perhaps if there is any significant disagreement about a given moral judgment, then that judgment is *ipso facto* disqualified as a permissible starting point. That is, perhaps we should treat it as an (additional) necessary condition of a judgment's being a “considered judgment” in the honorific sense that there is *general agreement* that it is true.²⁰

C2.P89 However, there are good reasons to reject the requirement that permissible starting points must be generally accepted. Consider again the knowledge platitude:

C2.P90 **THE KNOWLEDGE PLATITUDE:** If you know something that is relevant to a question that you are trying to answer, then you should take that information into account in arriving at a view

²⁰ Something close to this seems to be Smith's view (1994: 176–7). There, Smith cites with approval Rawls' earliest account of considered judgments (1951), according to which such judgments must be held in common by “the class of competent judges” (where “competent judge” is understood in relatively minimal terms). Notably, Rawls subsequently seemed to drop this requirement on “considered judgments”: it does not appear in any of his several later explications of the notion. Of course, it's possible that Rawls' earliest formulation was the best version, or at least, superior in this respect. In any case, quite apart from issues of Rawls' exegesis, an objection along these lines seems well worth considering.

34 MORAL KNOWLEDGE

C2.P91 The knowledge platitude is inconsistent with strong social requirements on permissible starting points. Suppose that before beginning the process of seeking reflective equilibrium among your moral views, you genuinely know that ϕ -ing is wrong. (Perhaps you originally learned that ϕ -ing is wrong from your parents, who knew it themselves.) If the wrongness of ϕ -ing is not generally accepted, then a norm that would have you set aside any moral judgment that is not generally accepted would have you ignore this piece of knowledge. Doing so would involve violating the knowledge platitude, because your knowledge that ϕ -ing is wrong is clearly relevant if you are attempting to determine (for example) which general principles about wrongness you should accept. Since the knowledge platitude is true, we should not accept putative norms that would lead to its violation. Therefore, the best version of the method of reflective equilibrium will not make it a necessary condition on permissible starting points that they are generally accepted.²¹

C2.P92 Consider the following view:

C2.P93 (i) For any individual, the appropriate starting point from which to pursue reflective equilibrium is the class of judgments consisting of all and only her considered judgments, i.e., judgments that she affirms in conditions that are hospitable or conducive to good judging.

C2.P94 For the reasons offered here, this characterization of the appropriate starting point is *overly inclusive*: even if a judgment qualifies as “considered” in the relevant sense, it does not follow that it is part of an appropriate starting point for pursuing reflective equilibrium. Because that claim is disputable, it is worth noting that (i) also seems to suffer from the opposite problem: that of being *overly exclusive*. That is, if one restricts the starting point to *only* considered judgments, then there is a risk that certain judgments might be excluded from playing a role in subsequent deliberation, in virtue of failing to

²¹ Notice that the general line of argument offered here against social requirements on permissible starting points does not actually require a principle as strong as the Knowledge Platitude; weaker principles would also suffice. For example, suppose that one rejects the Knowledge Platitude on the following grounds: even if you know something that is relevant to a question that you are trying to answer, it doesn't follow that you should take that information into account, for you might lack higher order knowledge that this is your situation. (Thanks to an anonymous referee for raising this issue.) Consider then the following, weaker norm, which is entailed by (but does not entail) the Knowledge Platitude:

If you know that (i) you know that p and that (ii) p is relevant to a question you are trying to answer, then you should take that information into account in arriving at a view.

Given that it's possible to stand in this cognitive relationship to a moral claim that is not generally believed, this weaker norm, like the Knowledge Platitude, is inconsistent with strong social requirements on permissible starting points.

qualify as “considered,” despite the fact that, intuitively, the judgments in question *should* play a role in one’s deliberations about which theory to accept.

C2.P95 For example, as we’ve noted, Rawls suggests that one set aside judgments that are heavily bound up with one’s own interests. But consider the following proposition:

C2.P96 A person of color should not receive lesser consideration in virtue of being a person of color.

C2.P97 Notice that, for a person of color, this judgment is heavily bound up with his or her own interests.²² Offhand, it seems like this judgment might fail to qualify as a considered judgment for a person of color for that reason, and thus should be excluded from her subsequent deliberations. But this seems like the wrong result. Rather, it seems as though it would be perfectly reasonable for a person of color to give a great deal of weight to this proposition in working towards a reflective equilibrium.²³ I think that the reason for this is the following: despite the fact that it is very much in the self-interest of a person of color that this proposition is true as opposed to false, it will typically be reasonable for her to think that this proposition is true. Indeed, at least in the usual case, we would expect her to *know* that this proposition is true. Because of this, not only would it be rationally permissible for her to take this proposition into account in her deliberations, but it would be a mistake for her to set it aside.

C2.P98 Significantly, both the “overly inclusive” and “overly exclusive” concerns about (i) issue from the same source. Considered judgments are ones that are held in conditions that are *hospitable* or *conducive* to judging well. However, even if one is in conditions that are hospitable or conducive to good judgment, there is no guarantee that the judgment at which one arrives will be reasonable as opposed to unreasonable. Not all unreasonableness is due to the operation of the kind of general corrupting factors (e.g., being personally invested in how a given question is answered) that the relevant conditions exclude. Conversely, even when one is in conditions that are in some respects *inhospitable* to good judgment, there is no guarantee that the judgment at which one arrives will be less than perfectly reasonable, or that it will fall short of knowledge. Here as elsewhere, there is a substantive gap between the

²² Of course, I do not mean to suggest that this proposition is not bound up with the self-interests of others as well.

²³ I do not mean to suggest that Rawls himself would disagree with this verdict, only that, given the conditions that he offers, its status as a considered judgment is at the very least problematic. Moreover, even if there are resources within the Rawlsian account for blocking this particular example, the underlying problem (which I spell out in the next paragraph) would persist.

36 MORAL KNOWLEDGE

quality of the conditions under which one performs and the quality of one's performances: both good performances in adverse conditions and bad performances in favorable conditions are perfectly possible, even if less likely than other combinations. This element of slack causes difficulties for the view that all and only one's considered judgments constitute the appropriate starting point for inquiry. For in deciding which theory to accept, it will seem wrong to give weight to unreasonable considered judgments, and (perhaps even more clearly) wrong to give no weight to perfectly reasonable judgments, or judgments that one knows, that do not qualify as "considered."

C2.P99 Moreover, it would be a mistake to think that the line of criticism developed here depends on some particular or idiosyncratic conception of what it is to be a considered judgment, as though the difficulties could be avoided by (for example) tweaking the Rawlsian conditions in various ways. Although of course particular counterexamples can be blocked by amending the conditions in various ways, the underlying problem is a more general one, and will arise for any account that explicates "considered judgment" in terms of conditions that are only *generally* hospitable or conducive to good judgment.

C2.P100 For example, consider again the account of the method offered by Scanlon (2014). There, Scanlon directly addresses the traditional worry that "the justificatory force of the method depends on the credibility of considered judgments with which the process begins, not just the credence attached to these judgments by the person carrying out the process" (p. 82). In response, he says the following:

C2.P101 this way of putting the matter rests on a misunderstanding... of what it is for something to be a considered judgment. In order to count as a considered judgment about some subject matter it is not enough that that judgment be very confidently held. It is necessary also that it should be something that seems to me to be clearly true *when I am thinking about the matter under good conditions for arriving at judgments of the kind in question*. My belief, sitting in my armchair with no information about conditions on the moon, that there is a rock there with my name on it, does not count as a considered judgment in the required sense, no matter how certain I may be that it is true.

C2.P102 (p. 82, emphasis his)

C2.P103 What Scanlon says here seems fair enough. But in the present dialectical context, the crucial point is this: even if *I am* in good conditions for making judgments about how things are on the moon, it does not follow that whatever seems true to me about how things are on the moon is a reasonable thing for me to think. Suppose that on a given occasion I am in good conditions for

making such judgments (whatever exactly that would involve) and that the evidence available to me strongly suggests that p is true, (where p is some proposition about conditions on the moon). Given that I am fallible in responding to evidence, these facts do not guarantee that I will not mistakenly judge that $\neg p$; and if I believe against my evidence, then the resulting judgment is unreasonable (even if the content of that judgment seemed clearly true to me). Thus, in the *non-moral* case, it seems clear enough that making a judgment in good conditions for making judgments of the relevant kind does not guarantee the reasonableness of one's judgment. Given one's fallibility, there is no guarantee against believing contrary to one's evidence. And it is unclear why we should think that things are otherwise when it comes to the moral domain—why we should think that *here*, something's seeming true to a person in good conditions guarantees the reasonableness of the corresponding judgment. Unlike Rawls, Scanlon does not offer an account of what counts as “good judging conditions” for moral judgments. Even in the absence of an account of those conditions, however, we can offer a recipe for generating problem cases for the version of the method that results once such an account is filled in. Namely: consider a hypothetical case of a person who is in the specified conditions, but who nevertheless has arrived at intuitively absurd moral judgments that seem true to her. Next, consider the question of whether the moral views of such an individual would enjoy the same epistemic standing as our views that *slavery is unjust* or *rape is wrong* given that her views survive the process of seeking reflective equilibrium. To the extent that it is implausible that this is so, this suggests that the standing enjoyed by our beliefs that *slavery is unjust* and that *rape is wrong* exceeds anything that could be bestowed upon them by the method of reflective equilibrium so understood.²⁴

C2.S5

2.5 Toward a More Defensible (But Less Ambitious) Version of the Method

- C2.P104 In light of the concerns adduced in the last section, it's worth considering what the method of reflective equilibrium would look like if we simply dropped the apparatus of considered judgments, and instead characterized the appropriate starting point in terms of judgments that have some substantive, positive

²⁴ Suppose that we characterized considered judgments not as judgments made in conditions that are *conducive* or *hospitable* to reasonable judging (e.g. being free from distractions) but in terms that *entailed* or *guaranteed* reasonable judging. For example, at the limit, one might characterize good judging conditions as those in which *one is thinking rationally about the question at issue, or responding correctly to the available evidence*. This would yield an account of the method that is equivalent to a version discussed in the next section, one that I take to be more defensible.

38 MORAL KNOWLEDGE

epistemic status. For example, suppose that instead of endorsing (i) a proponent of the method endorsed (ii):

C2.P105 (i) For any individual, the appropriate starting point for pursuing reflective equilibrium is the class of judgments consisting of all and only her considered judgments, i.e., judgments that she affirms in conditions that are hospitable or conducive to good judging.

C2.P106 (ii) For any individual, the appropriate starting point for pursuing reflective equilibrium is the class of all and only those judgments that she is justified in holding at that time.

C2.P107 Alternatively, a theorist drawn to “knowledge-first” accounts of evidence, on which a person’s evidence consists of their knowledge (see esp. Williamson 2000: 184–208), might very well be drawn to a version of the method that incorporates (iii) rather than (ii) or (i):

C2.P108 (iii) For any individual, the appropriate starting point for pursuing reflective equilibrium is the class of all and only those judgments that she *knows*.²⁵

C2.P109 Notice that a version of the method that incorporates either (ii) or (iii) as opposed to (i) would be significantly less vulnerable to both the “overly inclusive” and “overly exclusive” worries. For this reason, versions of the method that characterize the starting point in terms of judgments or beliefs that have some positive epistemic status seem much more defensible than versions that do not. Nevertheless, proponents of the method generally do not characterize the starting point in this way.²⁶ Why not?

²⁵ On a version of the method that incorporates (iii), the appropriate starting point for any inquirer will not contain any internal conflicts (nor will it conflict with the appropriate starting point for any other inquirer) given that anything that is known is consistent with anything else that is known. However, notice that even on such a version of the method, internal conflicts (as well as interpersonal conflicts) can easily arise in the course of pursuing reflective equilibrium, because attempts to formulate principles that account for initial judgments might very well generate inconsistencies.

Assuming that knowledge entails justified belief (cf. Section 1.2) but that justified belief does not entail knowledge, a version of the method that incorporates (iii) as opposed to (ii) will generally license less inclusive starting points. Whether a version that incorporates (iii) is preferable to a version that incorporates (ii), or vice versa, is a question that raises subtle issues; because these are orthogonal to the main line of argument developed here, I will not pursue them. For my current purposes, the most important points are that (1) either version would be less vulnerable to the kinds of objections that I’ve raised to standard versions of the method; and that (2) this is because of a feature that they share, viz. the fact that they characterize the class of judgments that constitute the appropriate starting point in terms of judgments that have some substantive positive epistemic status.

²⁶ Given that a version of the method that invokes *knowledge* would represent a particularly radical departure from early, canonical statements of the method, it’s perhaps not surprising that proponents of the method do not appeal to the concept of knowledge in characterizing it. More noteworthy is the

C2.P110 Here is one possible motivation that would amount to a *decisive reason* for eschewing notions like justified or reasonable belief, or knowledge, in characterizing the proper starting point: a theorist might have an *ambitious* understanding of the method and its epistemological role, in the sense described in section 2.1. Again, a theorist who adopts an ambitious interpretation of the method and its epistemological role tends to see our moral knowledge, and whatever justification we have for our moral views, as in some way arising out of, or dependent upon, the method. On an ambitious interpretation, the method of reflective equilibrium is the answer to an epistemological question about morality, one that is sometimes framed as a challenge: how can we arrive at knowledge or justified beliefs about morality in the first place? Notice that, for a theorist who accepts an ambitious interpretation of the method, the suggestion that we understand the method in terms of (ii) or (iii) rather than (i) will seem to be a complete non-starter. Because the theorist aspires to account for the possibility of moral justification and knowledge in terms of the method itself, her project is undermined when the method is understood and characterized in a way that presupposes that we *already have* some justified moral views, prior to engaging in the method.

C2.P111 However, as we have seen, there are good reasons to reject an ambitious interpretation of the method. As section 2.2 argued, given that there is at least some moral knowledge, it is implausible that all the moral knowledge that is had by individuals is in some way dependent on reflective equilibrium reasoning; and the same is true of justified moral beliefs. The arguments offered there are independent of the critique of “considered judgments” developed in the last section. Because there are independent reasons to think that not all moral knowledge or justified moral beliefs are due to the method of reflective equilibrium, the kind of consideration at issue here does not provide a good reason for eschewing such notions in characterizing the proper starting point for moral inquiry.

C2.P112 Is there some other more compelling reason for eschewing such notions? Consider again Rawls’ conditions for being a considered judgment. It is a notable feature of those conditions that it is typically quite *easy to tell* whether a given judgment satisfies them. For example, we are typically in a good position to tell whether we are upset or frightened when thinking about a question,

fact that they also consistently eschew appeals to weaker notions, such as justified or reasonable belief. For example, none of the accounts representative accounts referenced in footnote 15 make use of such notions in characterizing the starting point. Moreover, and significantly, none of those accounts seems like a good candidate for being extensionally equivalent to an account that does make use of such notions, for the same reasons why (i) is not plausibly equivalent to (ii).

40 MORAL KNOWLEDGE

whether we are relatively confident as opposed to uncertain whether a given judgment is true, and so on. Thus, it seems that we would generally be quite reliable in determining whether a given judgment qualifies as *considered* or not. If the method of reflective equilibrium requires each of us to be able to reliably identify which of our judgments are considered in the relevant sense, it seems like that task will be a manageable one. On the other hand, it is natural to worry about our ability to successfully follow a method that requires us to identify those judgments that have some more objective positive epistemic status. How after all, are we supposed to assess the initial, rational credibility of some proposition, or whether it would be reasonable to believe it?

C2.P113 While I appreciate the appeal of this line of thought, I think we should resist it. Notice first that, although we can expect that people will generally be quite good at identifying which of their moral opinions count as Rawlsian considered judgments, they will not be infallible. (And the same is true for any explication of “considered judgment” that has yet been proposed.) Indeed, on any plausible explication, errors will be possible in both directions: in a given case, one might mistakenly take something to be a considered judgment that is not, or fail to recognize something that is in fact a considered judgment as such. This fallibility with respect to identifying considered judgments follows from our fallibility with respect to questions about whether the relevant conditions are satisfied in particular cases.

C2.P114 Consider next a version of the method that characterizes the proper starting point in terms of some positive epistemic status, for example, as *those propositions that you are justified in believing as the search for reflective equilibrium begins*. In order to correctly apply this version of the method, you will have to have a reasonably competent grip on which moral claims it is reasonable for you to take to be true given your current situation; if we were hopelessly unreliable or at sea with respect to such assessments, then, given that the method is supposed to be something that we can *apply*, this would not be a viable construal. But in fact, setting aside a radical and as of yet unmotivated skepticism, there is no reason to suppose that we are hopelessly unreliable with respect to such assessments. Granted, given the relative ease with which Rawlsian considered judgments can be identified as such, we would expect people to make more mistakes about which propositions it is reasonable for them to believe than about whether something is among their Rawlsian considered judgments. But given that the difference is a matter of degree as opposed to kind, it is unclear why that should motivate interpreting the method in terms of Rawlsian considered judgments, in light of the problems that beset the method when it is interpreted in that way.

C2.P115 By way of illustration let us look at how this issue plays out with respect to concerns about self-interest. Above, I suggested that, even if it is very much in your self-interest that p is true rather than false (where p is some moral claim), it does not follow that you should set aside your judgment that p in evaluating rival moral theories. This is because, even if you are in less than ideal circumstances for judging whether p, and you appreciate this fact, your judgment that p might nevertheless be fully justified; indeed, it is consistent with that fact that it is among the things that you know to be true. Of course, the fact that you should not automatically set aside any moral judgment that strongly aligns with your self-interest might make it significantly more difficult in practice to correctly manage your biases: from the inside, a case in which you hold an unjustified belief *because* it aligns with your self-interest might feel very much like a case in which you are genuinely justified in holding a belief that happens to align with your self interest. A norm such as “One should set aside any moral judgment that strongly aligns with one’s self-interest” would have us treat these difficult to distinguish cases alike; it is in that respect relatively easy to apply. (Although again, we should expect honest failures of compliance to occur even with respect to such easy to apply norms.) But that is a poor reason for adopting this as a norm of inquiry, as opposed to a norm of inquiry that would have us distinguish, among judgments that align with our self-interest, between those that are justified and those that are not.

C2.P116 Once we abandon an ambitious interpretation of the method and its epistemological role, we should be open to the following natural thought: for any individual, the moral views that constitute the proper starting point for moral inquiry are those views that it is reasonable for her to think true at the outset of inquiry. On a version of the method that accords with this natural thought, when you give weight to an unreasonable judgment, or neglect to take into account a judgment that is both reasonable and relevant, you fail to apply the method correctly. Thus, on the proposed understanding of the method, someone who gave weight to an unreasonable judgment that *we are morally required to occasionally kill randomly* by (for example) treating it as a reason to reject proposed moral principles with which it is inconsistent, would *ipso facto* be reasoning badly, and violating the method (even if that judgment qualifies as “considered” for the person). In contrast, when your pursuit of greater coherence and systematicity is in accord with the method, the judgments to which you are responding are ones to which it is reasonable to be sensitive, given the point to which the inquiry has progressed.

C2.P117 There is, I think, nothing special about morality here. When an empirical scientist engages in armchair theorizing, he might attempt to make progress

42 MORAL KNOWLEDGE

by pursuing a more coherent and systematic theory of the domain with which he is concerned. In engaging in such armchair theorizing, it makes sense for the scientist to begin from what is reasonably believed to be true at the time. Perhaps some of these propositions are things that he has observed to be true in his own laboratory, while others are things that he reasonably believes by relying on the reports of other competent scientists. Notice that in this case, the analogue to an *ambitious* interpretation of the method seems like a complete non-starter: there is no temptation for us to think that what justification the scientist has for his beliefs about the domain arises exclusively from his own deliberative, reflective activity. Rather, he is already justified in believing much that is of direct relevance before he engages in the relevant theorizing. The fact that an ambitious interpretation is more tempting in the moral domain might be partially an artifact of the facts that (i) there does not seem to be a clear analogue to empirical observation in the moral domain, and that (ii) at least for a mature moral agent, relying on moral testimony seems like at best a relatively marginal phenomenon.²⁷

C2.S6

2.6 The Pursuit of Coherence as a Source of Moral Knowledge

C2.P118 On the account suggested in the last section, in typical cases in which you reason *well* in pursuing a more coherent overall view, you rely on claims that you were *already* justified in believing, prior to engaging in the process itself. Indeed, we can go further: in typical cases in which you gain knowledge via the pursuit of greater coherence and systematicity among your moral views, the process typically depends on your having *prior moral knowledge*. Imagine someone who initially believes that *One should always tell the truth*. She is then presented with a hypothetical case in which an agent can only save the life of an innocent person by declining to tell the truth to a would-be murderer. In response, she judges that the agent in the hypothetical case should *not* tell the truth to the would-be murderer, a judgment that she recognizes is in conflict with a belief that she has held up until now. She resolves the conflict by abandoning her prior belief that *One should always tell the truth* and replacing it with the view that *in some circumstances, it is morally permissible not to tell the truth*, a view that she takes to be entailed by her judgment about the hypothetical case.

²⁷ For further discussion of (i), see Chapter 4; for further discussion of (ii), see Chapter 3.

C2.P119 Here, the pursuit of coherence leads to a change in view of a perfectly familiar kind. What would have to be the case in order for this change in view to amount to an *improvement* in the person's moral views, as opposed to a mere change? More specifically, what would have to be the case in order for the moral view at which she ultimately arrives to count as something that she *knows*? Given the way in which she arrives at the view that *in some circumstances, it is morally permissible not to tell the truth*, it is plausible that she knows that this view is true only if her judgment that the agent in the hypothetical case is not required to tell the truth is itself a piece of knowledge: given that she infers the former from the latter, she knows the former only if she knows the latter. I think that in typical cases in which a person acquires new knowledge in the course of attempting to improve the coherence of her views, her prior knowledge plays a crucial role in the process. (It is not as though she comes to know the truth about the agent in the hypothetical case as a *result* of attempting to bring her moral views into a coherent equilibrium.) And this, I think, is true in both the moral and the non-moral domains.

C2.P120 In claiming that prior knowledge *typically* plays a role in cases in which a person acquires new knowledge via the pursuit of coherence, I do not mean to endorse the stronger claim that the pursuit of coherence can *never* lead to the acquisition of new knowledge in the absence of prior knowledge. Indeed, I think this is perfectly possible, although relatively unusual.

C2.P121 Consider first the non-moral case. Suppose that, although viewing conditions are relatively poor, you seem to observe that O. Suppose further that the apparent observation O is quite unlikely and unexpected given the rest of what you currently believe. Perhaps given the relative poverty of your viewing conditions and the low prior probability of O relative to the rest of what you believe, you are not in a position to know that O is true. However, your glimpse was strong enough to justify significant credence in O, and if your epistemic position with respect to O were any stronger than it currently is, you would be in a position to know that O is true. (That is, as things stand, the case is a marginal case of non-knowledge.) Still, the fact that you seemed to see that O, something that would be quite surprising given the rest of your beliefs, occasions further thought on your part: is there anything that would explain O's being true, and that would allow you to reconcile its truth with the rest of what you believe in a satisfying way? Suddenly you hit upon a novel hypothesis that would explain O's being true, and which is itself perfectly plausible given the rest of what you believe. Given that you were almost in a position to know that O even in the absence of the novel hypothesis (and one of the things that prevented you from knowing that O is true was the absence

44 MORAL KNOWLEDGE

of any plausible account of how it could be true), it seems perfectly possible that hitting on the novel hypothesis in this way might put you in a position to know that O is true. However, it is not as though you had prior knowledge that the hypothesis is true; indeed, your main reason for believing it now might very well be the apparent observation that O. In this way, the search for a more coherent overall view—one that reconciles the apparent observation O with the rest of what you currently believe—leads to new knowledge, even though the process is not underwritten by *prior* knowledge.

C2.P122 Plausibly, the same phenomenon is possible in the moral case. Suppose that you make a number of plausible moral judgments, each of which is a bit too uncertain to count as an instance of knowledge (although each falls short of that standard only marginally). You then hit upon a plausible moral principle which accounts for the truth of these moral judgments in a satisfying way; you thus infer that the principle is true, and your awareness that there is a plausible principle that offers a satisfying explanation is enough to boost the credibility of the original judgments to the point that they are now among the things you know to be true. Perhaps by the end of this process you also know that the principle is true as well. However, even in that case, it is not as though either your knowledge of the principle or your knowledge of the judgments is prior to the other; rather, gaining knowledge of the one travels hand in hand with gaining knowledge of the other. Thus, it seems that in the moral domain too, the pursuit of coherence can in principle lead to new knowledge, in a way that does not essentially involve prior knowledge.

C2.P123 We should not, however, mistake this interesting epistemic phenomenon for the usual case, in either the non-moral or the moral domain. When a scientist learns that a given hypothesis is false by making an observation that contradicts it, her ability to recognize that the hypothesis is false depends on her prior, observational knowledge. Similarly, when the scientist comes to know that some theory is true because it is the only plausible explanation of some body of data, the theoretical knowledge at which she arrives typically depends on her prior knowledge of the data that it explains. In paradigmatic cases, the acquisition of knowledge via the pursuit of coherence and systematicity in one's views depends on the possession of prior knowledge, and there does not seem to be any reason to think that things are otherwise in the moral domain. As noted above, in a case in which one learns an existentially quantified proposition to the effect that “It is sometimes morally permissible to ϕ ” by inferring it from a judgment to the effect that “In *this* case, it is morally permissible to ϕ ,” one typically knows the former proposition only if one knows the latter proposition. Is there some reason to think that the prior knowledge

which one exploits in drawing this conclusion can always be traced back to some earlier process in which moral knowledge was brought into existence by the pursuit of greater coherence and systematicity among moral beliefs that were not known at the start of the process? As far as I can tell, there is none. The mistake made by a philosopher who accepts an ambitious interpretation of the method of reflective equilibrium (on which the pursuit of greater coherence and systematicity among initially plausible judgments is *the source* of our moral knowledge) is to treat what is in fact a relatively exceptional case as though it were the usual one.

C2.P124 According to the working hypothesis endorsed in Chapter 1, any source of non-moral, empirical knowledge is also a potential source of moral knowledge. An important source of empirical knowledge is the pursuit of coherence and systematicity: not infrequently, we arrive at new knowledge about the empirical world by attempting to make our existing views about it more coherent and systematic. On the view sketched here, the process by which this occurs does not differ in epistemologically relevant respects from the way in which the pursuit of coherence and systematicity yields new moral knowledge.

C2.S7

2.7 The Method and the Moral Philosopher

C2.P125 In section 2.1, I suggested that a philosopher who contends that the method of reflective equilibrium does not have the kind of centrality for moral epistemology that it is often regarded as having incurs an additional dialectical burden given the popularity of the method among practicing moral philosophers. That burden consists in the need to offer a plausible explanation of why the method has seemed to many moral philosophers to be a compelling account of their “best practices,” or a kind of reasoning in which they routinely engage and which frequently seems to deliver genuine insights into the moral domain. Here is my attempt to discharge this burden.²⁸ When a moral philosopher succeeds in gaining moral insight or knowledge by critically reflecting upon her moral views, the description that the reflective equilibrium theorist would offer of this process is often *a correct description* of what has taken place. That is, when a moral philosopher arrives at moral knowledge or

²⁸ After strongly criticizing the method on a number of fronts, McPherson (2015: 668) provocatively conjectures that its continuing popularity among moral philosophers is due to the fact that it provides a sociologically respectable (although ultimately intellectually indefensible) license to continue doing what they are already doing without “becoming bogged down in methodological inquiry.” In what follows, I offer an alternative hypothesis.

46 MORAL KNOWLEDGE

insight by critically reflecting upon her existing moral views, this involves (at least often enough) attempting to achieve greater coherence and systematicity among judgments that satisfy Rawls' conditions for "considered judgments," or broadly similar conditions.

C2.P126 However, that is consistent with the following possibility: in a case in which a philosopher acquires moral knowledge in this way, to describe the process in terms of the method of reflective equilibrium will typically be to *under-characterize* that reasoning. Compare: in a case in which a philosopher arrives at a new piece of moral knowledge by critically reflecting on her existing moral views, it will typically be correct to describe that process as "reasoning"—but in an obvious sense, *that* description would be a significant under-characterization of the way in which she has arrived at the knowledge in question.

C2.P127 Suppose that a particular episode of knowledge acquisition can be correctly described both as an instance of "reasoning" as well as an application of reflective equilibrium reasoning in Rawls' or Scanlon's sense. Because the generic description "reasoning" is equally applicable to episodes of good reasoning and bad reasoning, its use in this context brings what is in fact a case of good reasoning under a perfectly *neutral* description, one that indicates nothing about its quality. When we describe the same episode in terms of the reflective equilibrium model, we offer a better description, not only in the sense that the description is more specific and less generic, but also in the sense that it is at least somewhat *less neutral* with respect to the quality of the reasoning so described. For to describe the episode as an instance of reflective equilibrium reasoning is to exclude the possibility that the episode is a case of bad reasoning in at least some of the ways that reasoning can be bad. (For example, given that the reasoning conformed to the reflective equilibrium model, the believer did not rely on judgments of hers that are fleeting as opposed to stable, or on premises that did not even seem true to her at the time that she reasoned from them. By contrast, even such paradigmatically bad reasoning is not excluded by the more generic description "reasoning.") Still, despite the fact that a description in terms of the method is an improvement over the more generic description "reasoning," it is, I think, ultimately defective in exactly the same way: while both descriptions are consistent with (and correct descriptions of) much good reasoning about morality, they are equally consistent with (and correct descriptions of) a great deal of intuitively *bad* reasoning about morality, reasoning that would be manifestly unsuited for delivering knowledge or even justified beliefs about the subject matter. In this respect, the kinds of descriptions that the reflective equilibrium theorist offers fail to characterize good moral reasoning as good moral reasoning—even though the reflective

equilibrium story might very well be *true* of much moral reasoning that is in fact good. My suggestion is that it is this last fact that accounts for much of the popularity that the method enjoys among practicing moral philosophers, and for their sense that there is something right about it, as an account of their practice at its best.

C2.P128 Still, it would be a mistake to infer from this that the kinds of evaluatively neutral descriptions offered by the reflective equilibrium theorist are the correct *normative* account of that practice, or an adequate characterization of the kind of reasoning about morality that is well-suited to deliver moral knowledge or even justified moral beliefs. Consider again a comparison with scientific knowledge. Suppose that a scientist initially (and reasonably) believes some theory T. She then subjects theory T to a *crucial experiment*, one that is designed to falsify it just in case the theory is in fact false: aware that the theory predicts observational consequence O in circumstances C, she arranges for circumstances C in her laboratory. She then observes not-O; on the basis of this observational knowledge, the scientist comes to know that theory T is actually false. Suppose that we wanted to offer an account of how the scientist managed to arrive at this piece of theoretical knowledge. We could of course offer a *neutral description* of how her beliefs changed over time, along the following lines: initially, she believed theory T because it seemed true to her; at a later time, she came to believe that not-O is true; because she believed that not-O and T are inconsistent, she changed her mind about T and came to believe that T is false.

C2.P129 At one level of description, this is a perfectly accurate account of how the scientist's psychology evolved over time, and how a concern for coherence led to a change in her stance towards theory T. However, if what we are interested in is an epistemologically illuminating account of how the scientist arrives at *knowledge* that theory T is false, then there is an obvious respect in which an account along these lines badly under-describes what has taken place. For if what we seek to understand is how the scientist ends up knowing (as opposed to merely believing or judging) that theory T is false, then it is crucial that the scientist knows that not-O is true. Moreover, there is an informative story to be told about how the scientist knows that not-O is true: namely, she knows that it is true in virtue of having observed this in her laboratory. My view is that, when the reflective equilibrium story is offered as an account of how we know moral truths, it is inadequate for the same reason that any account that failed to make reference to the scientist's observational knowledge would be inadequate. At the same time, there is a good reason for why the method of reflective equilibrium might seem correct to many practicing moral

48 MORAL KNOWLEDGE

philosophers as an account of their “best practices”: namely, descriptions in terms of the method are in fact *true of* much of the reasoning in which they regularly engage, reasoning which sometimes delivers genuine moral knowledge and insight.

C2.S8

2.8 Levels of Generality

C2.P130 As noted in Section 2.1, a characteristic claim of proponents of the method of reflective equilibrium is that no judgment should be privileged over any other in virtue of its level of generality. In Scanlon’s words, “What the method of reflective equilibrium prescribes is, so to speak, a level playing field of intuitive justification on which principles and judgments of all levels of generality must compete for our allegiance” (2002: 151). In Section 2.1, I emphasized that although a claim of this sort will be accepted by any reflective equilibrium theorist, it might be also be accepted by others, including philosophers who differ quite radically in their views about moral inquiry. (The Caricaturized Moral Intuitionist described in that section is an example of such a philosopher.) At the same time, this aspect of the method has frequently been criticized.²⁹ Notably, for example, some consequentialists have objected to the method on the grounds that it encourages us to treat some of our relatively low-level moral judgments as casting doubt on more general and over-arching consequentialist principles, principles that are alleged to enjoy something like the status of self-evident axioms in comparison to the lower-level moral judgments with which they conflict.³⁰

C2.P131 On this issue, I think that reflective equilibrium theorists are correct. In this section, I will defend the claim that no moral judgment should be privileged over any other in virtue of its level of generality. Before defending this claim, I want to say a bit about how I understand it, and what I take it to entail. First, if no moral judgment should be privileged over any other in virtue of its level generality, then it follows that, when conflicts arise among

²⁹ Notice, however, that resistance to this particular aspect of the method is *unlikely* to come from either moral skeptics or nihilists. Indeed, there is an obvious way in which moral skeptics and nihilists will agree with the idea that none of our moral judgments is privileged over any other in virtue of its level generality: for such theorists, the *equal* standing of moral judgments at different levels of generality follows from the general *lack* of standing of such judgments.

³⁰ See, for example, Singer (1974, 2005). To be sure, debates about the methodological significance of the fact that our moral judgments differ with respect to their level of generality predate contemporary debates about the method of reflective equilibrium. For a recent discussion of this issue as it arose in the British moral philosophy of the immediately preceding periods, see Hurka (2014).

moral judgments that differ in their level of generality, the mere fact that the conflicting judgments have the levels of generality that they do is not a reason for resolving the conflict in one way rather than another. Second, and contrary to what some moral philosophers have thought, there is no specific level of generality that is had in common by the judgments that constitute the proper starting for systematic theorizing about morality. On the contrary, no moral judgment is precluded from being part of the proper starting point for such theorizing simply in virtue of its level of generality.

C2.P132 The substantial nature of these commitments becomes even clearer when we bear in mind just how diverse our moral judgments actually are with respect to their level of generality. In discussions of moral methodology, philosophers often distinguish between “judgments about particular cases” and “judgments about general principles.” But this simple bipartite division tends to obscure the actual diversity of our moral judgments along the relevant dimension (even if those who talk in these terms are not themselves under any illusions about the point). Consider the common philosophical practice of describing a hypothetical scenario and then offering a judgment about that scenario—for example, a judgment to the effect that it would be morally wrong to sacrifice an unwilling innocent bystander in order to save five hospitalized patients who are dying from organ failure. Philosophers frequently refer to such judgments as “judgments about particular cases,” but as Shelly Kagan (2001: 61–2) has pointed out, such judgments are actually judgments about a *type* of case: the situation to which they purport to apply is one that could, at least in principle, be instantiated multiple times. In this respect, such judgments are significantly more abstract and theoretical than the most specific moral judgments that we commonly make and affirm in the course of ordinary life. Consider, for example, judgments such as *The 9/11 hijackers acted wrongly in flying planes into The World Trade Center*, or *I did not do anything morally wrong by writing this book*, or *George Washington was a good man*. These moral judgments concern the moral quality of a token action or individual; their content does not extend to any other actions or individuals, even other actions or individuals that it would be natural to regard as relevantly similar. In this respect, such judgments are considerably less general than the judgment that one should not sacrifice the innocent bystander in the Organ Harvesting case, or that one should flip the switch in the Trolley case (cf. Foot 1967; Thomson 1976), judgments which are best interpreted as judgments about act types.

C2.P133 Cases like the Organ Harvesting or Trolley cases, along with the intuitive judgments that they tend to elicit, are often treated, controversially, as

50 MORAL KNOWLEDGE

counterexamples to more general philosophical principles and theories that entail that we should act differently in the relevant circumstances. Standard philosophical practice might naturally suggest that the attendant judgments are among the *least* general judgments that are relevant to moral theorizing. However, it is clear that moral judgments to the effect that some token action has a certain moral quality, no less than “judgments about particular cases,” can conflict with general moral principles and theories. Consider, for example, the moral judgment that *I did not do anything wrong by writing this book*. On some uncompromising forms of consequentialism (cf. Kagan 1989), we act wrongly whenever we act in some way other than the way that would maximize the good; on this view, almost all of us act wrongly, almost all of the time. On the plausible assumption that there was some alternative course of action open to me during the writing of this book that would have better promoted the good than my writing it, I acted wrongly in writing this book, according to such theories. Similarly, the judgment that the *The 9/11 hijackers acted wrongly in flying planes into the World Trade Center* conflicts with the general claim that no action is ever morally wrong, a claim that is central to various error theories. Given that judgments of this kind potentially have direct import for which theories we should accept, a theorist who holds that no moral judgment is disqualified from playing a role in moral inquiry simply in virtue of its level of generality should hold that there are judgments of relevance to moral inquiry that are even less general than what are often called “judgments about cases.”

C2.P134 In the same way that the expression “judgments about cases” can be ambiguous between judgments about token cases and judgments about case types, and in this way mask the actual diversity among such judgments, talk of “judgments about principles” can similarly cover a range of judgments that differ significantly in their level of generality. The generalizations about morality that one might affirm range from the sweeping and unrestricted (“One ought always to act so as to make the world as good as possible”) to very general, exceptionless prohibitions on relatively unqualified act types (e.g. “It is always wrong to torture people”) to highly restricted, qualified, and hedged moral principles (e.g. “Any man who tortures children to death solely for the pleasure of watching them die is a bad man”³¹). Indeed, if in fact what philosophers

³¹ I take this last example from Scott Soames (2014: 200). According to Soames, “Restricted generalizations like these are the platitudes that constitute our starting points in ethics” (p. 200). Although he does not discuss the method of reflective equilibrium, I take Soames’ view that the judgments that constitute our proper starting points in ethics have some specific level of generality to be incompatible with the more egalitarian outlook endorsed both here and by the reflective equilibrium theorist.

call “judgments about particular cases” are often judgments about case types as opposed to judgments about token events, then there is a sense in which these judgments are *also* judgments about general principles (i.e. they are judgments about how one should act in *any* case which satisfies the corresponding description) albeit judgments with significantly narrower scope than the judgments that are typically called “judgments about principles” by philosophers.

C2.P135 But even when the full extent to which our moral judgments differ in their level of generality is taken into account, the kind of egalitarian methodological approach associated with the reflective equilibrium picture remains compelling. I will argue for the following two theses:

C2.P136 **DIVERSITY:** There is no specific level of generality that is common to all of the propositions that make up the proper starting point for theorizing about morality.

C2.P137 **NO PRIVILEGE:** When conflicts arise among moral judgments that differ in their level of generality, the mere fact that the conflicting judgments have the levels of generality that they do is not itself a good reason for resolving the conflict in one way rather than another.

C2.P138 Here is a straightforward line of thought in support of DIVERSITY. When one engages in systematic theorizing of any kind, one should take into account any relevant knowledge that one has. Thus, in theorizing about morality, one should take into account any pre-theoretical moral knowledge that one has, i.e., anything that one knows about morality prior to engaging in systematic theorizing about it. Although in principle it might be maintained that full-fledged moral knowledge arises only as the *output* of such theorizing, we should reject that idea, for the sorts of reasons having to do with hyper-intellectualization rehearsed in Section 2.2. Given that we have at least some pre-theoretical moral knowledge, the claim that DIVERSITY is false amounts to the claim that all of our pre-theoretical moral knowledge has some specific level of generality in common. However, that claim is less plausible than the alternative view, that our pre-theoretical moral knowledge is not restricted in this way. For example, an ordinary person might know both that *it was wrong for the 9/11 hijackers to fly planes into the World Trade Center*, and that *it is*

For further discussion of Soames on this point and related issues, see McGrath and Kelly, “Soames and Moore on Method in Ethics and Epistemology” (2015) and Soames’ (2015) reply.

52 MORAL KNOWLEDGE

wrong to intentionally hurt other people, even without having engaged in any systematic moral theorizing.^{32,33}

C2.P139 The basic line of thought can be detached from the knowledge theoretic framework in which I have presented it. For example, one might appeal to *reasonable belief* rather than knowledge: plausibly, our starting point for moral theorizing should include anything that we reasonably believe about morality, prior to engaging in such theorizing. But the moral opinions that it is reasonable for us to hold prior to engaging in such theorizing do not have some one level of generality in common.

C2.P140 A similarly straightforward line of reasoning can be used to support NO PRIVILEGE. In any case in which one becomes aware of a conflict in one's judgments (whether those judgments concern morality or any other subject matter), the relevant questions to ask will be about the comparative epistemic standing of the judgments that conflict. For example, is either judgment among the things that one knows to be true? If not, which of the two conflicting judgments is it more reasonable to think is true? It would be strange to think that much progress could be made in answering questions of this sort by reflecting on the level of generality of the judgments at issue, for the following reason. If we think of a judgment as something that is made by a particular person at a particular time (or held by that person through some interval of time), then its level of generality is something that it shares with any other token judgment of the same type (i.e. any other token judgment with the same propositional content), regardless of who makes it or how she is embedded in the world at the time that she makes it. In contrast, the epistemic status of a token judgment will typically be quite sensitive to such facts, since the epistemic status of a token judgment will be sensitive to the evidence that the

³² *Objection:* No one could know that *it is wrong to intentionally hurt other people*, because the relevant generalization is not even *true*: there are, after all, exceptional circumstances in which it is not wrong to intentionally hurt other people. *Reply:* That there are such circumstances would show that the generalization is false only if it is interpreted as a universal generalization, as opposed to a generic generalization that tolerates exceptions. But interpreting this and similar moral generalizations as generics rather than as universal generalizations is both the more plausible and the more charitable way of understanding these claims, as they are believed by most of those who believe them. (Note also that even when such claims are understood as generics, they conflict with substantive claims that have been defended in both normative ethics and metaethics.) On the importance of moral generics, see especially Lerner and Leslie (2013).

³³ Here is a reason for thinking that not every piece of pre-theoretical moral knowledge has the same level of generality: it is implausible that the potential sources of such knowledge would discriminate on the basis of this feature. For example, in the next chapter, I will defend the claim that testimony can deliver full-fledged moral knowledge. However, it is implausible that testimony would be able to transmit knowledge of moral propositions at some specific level of generality but not at any other. (Thus, a young child might come to know via the testimony of her parents that some specific action that she performed was wrong, as well as the more general claim that "It is wrong to intentionally hurt other children.")

person has at the time she makes it, and it is a truism that different people can have different evidence, and that one and the same person can differ in the evidence that she has at different times. Token judgments of the same type (i.e. with the same propositional content) will not differ in their level of generality, but will frequently differ in their epistemic status. This suggests that epistemic status is largely independent of level of generality.

C2.P141 The very straightforward considerations rehearsed here amount to a significant *prima facie* case for the idea that the level of generality of a moral judgment is not itself significant in a context in which one is deliberating about which judgments to accept. Nevertheless, some have claimed that considerations having to do with levels of generality have a kind of methodological significance that is missed by the more egalitarian picture favored by the reflective equilibrium theorist and endorsed here. As noted above, perhaps the most common claim of this kind is that the egalitarian picture encourages us to give excessive weight to intuitively plausible but relatively low-level moral judgments when we evaluate proposed moral principles that conflict with those lower level judgments. A classic statement of this concern is made by Peter Singer (1974), in the course of criticizing Rawls' (1971) view that some of our particular moral judgments count strongly against utilitarianism:

C2.P142 Why should we not make the opposite assumption, that all the particular moral judgments we intuitively make are likely to derive from discarded religious systems, from warped views of sex and bodily functions, or from customs necessary for the survival of the group in social and economic circumstances that now lie in the distant past? In which case, it would be best to forget all about our particular moral judgments, and start again from as near as we can get to self-evident moral axioms. (p. 516)

C2.P144 According to Singer, we have more reason to distrust our judgments about particular cases than our judgments about newly encountered moral principles, precisely because the former are more likely to be due to a kind of cultural conditioning than the latter:

C2.P145 We have all been making moral judgments about particular cases for many years before we begin moral philosophy. Particular views have been inculcated into us by parents, teachers and society from childhood. Many of them we act upon every day... These judgments sink deep, and become habitual. By contrast, when we read Sidgwick for the first time we are suddenly called upon to decide whether certain fundamental moral principles, which we

54 MORAL KNOWLEDGE

may never have explicitly thought about before, are self-evident. If it is then pointed out to us that this fundamental moral principle is incompatible with some of the particular moral judgments we are accustomed to making, and that therefore we must either reject the fundamental principle, or else abandon our particular judgments, surely the odds are stacked against the fundamental principle. Most of us are familiar with lingering guilt feelings that occur when we do something that we are quite certain is right, but which we once thought to be wrong. These feelings make us reluctant to abandon particular moral views we hold, but they in no way justify these views.³⁴

C2.P146

(pp. 516–17)

C2.P147 Others have similarly argued that our lower level moral judgments are less trustworthy than our more general judgments. For example, according to Michael Huemer (2008), “level of generality matters because intuitions of different levels of generality differ in their susceptibility to various kinds of error” (p. 383). He argues that we have reason to prefer “abstract theoretical intuitions” over both “mid-level” and “concrete intuitions”, largely because the former are significantly less susceptible to being influenced by emotional, cultural, and biologically-generated biases than the latter (pp. 383–4).³⁵

C2.P148 Of course, the reflective equilibrium theorist holds not that one should rely on just any particular moral judgment that one might make, but only judgments that meet certain further conditions. For example, as we have seen, for Scanlon, the judgments in question are ones that “seem clearly...correct and seem so under conditions that are conducive to making good judgments.” Suppose that we became convinced that even those of our particular moral judgments that meet these further conditions are so unreliable as a class (perhaps because they are often influenced by various kinds of bias), that we should not rely on them when engaging in moral inquiry.³⁶ Where exactly would that leave us?

³⁴ In his more recent (2005), Singer places greater emphasis on biological and evolutionary programming, which he takes to similarly contaminate and undermine the reliability of many of the case specific moral judgments by which we are inclined to judge general principles.

³⁵ Huemer distinguishes between intuitions (intellectual appearances that might or might not be endorsed) and corresponding judgments. But I take it that any reason to think that our lower level intuitions are less reliable as a class than our higher level intuitions will also be a reason to think that our lower level judgments are less reliable than our higher level judgments, given that (as on Huemer’s view) judgments are typically formed in direct response to intuitions with matching content.

³⁶ Above, I argued that even if a moral judgment seems clearly true to someone under good conditions, it does not follow that it is a reasonable thing for him to think. For this reason, I believe that there is room to argue, in a specific case in which someone makes a particular judgment that satisfies these conditions, that that judgment nevertheless should not be relied upon. What Singer seems to have in mind, however, is that we should, as a matter of policy, set aside or discriminate against an

C2.P149 Consider Singer's suggestion that it might make sense to set aside (or "forget all about") our particular moral judgments and simply decide directly which allegedly self-evident general principles are true. Now, there are two questions that we might ask in response to proposals of this kind. The first and most obvious is whether we actually have compelling evidence that our particular moral judgments are unreliable to the point that it would make sense to set them aside. Although I am skeptical that we have such evidence,³⁷ I want to pursue a different question about methodological proposals along these lines. Suppose that we *did* acquire compelling evidence that we are unreliable when it comes to making particular moral judgments. Would it follow that we should adopt the kind of top-down, egalitarian methodology advocated by Singer and others? I will conclude this chapter by raising some doubts about this suggestion.

C2.P150 In order to make things vivid, imagine that you are given the opportunity to consult an oracle who has information about how reliable you are when it comes to making moral judgments at different levels of generality. Prior to consulting the oracle, you confidently endorse many low level moral judgments (judgments that strike you as clearly correct when you are thinking as carefully as you can about the issue), as well as a number of general moral principles, of which the same is true. Unfortunately, you are only allowed to ask the oracle a single question. You decide to ask the oracle about the reliability of your lower level moral judgments. The oracle reports back as follows:

C2.P151 I have some bad news you. When it comes to making low level moral judgments, you're very unreliable. Even when a particular moral judgment seems clearly true to you when you're thinking about it as carefully as you can in good conditions, that's simply not any indication that it *is* true. But don't take it personally, it's not as though you're any worse at this than your fellow human beings are. The truth is, given all of the nonsense about morality that people hear from their parents, teachers, and society, most of them are pretty terrible when it comes to this sort of thing, just like you are. It is what it is.

entire class of judgments (a class unified by the level of generality of its members), and it is this suggestion which I am concerned to criticize.

For a more wide-ranging defense of the practice of relying on specific ethical claims, see Harman (2015).

³⁷ Singer (2005) appeals to the fMRI research conducted by Joshua Greene and his colleagues (Greene et al. 2001, Greene and Haidt 2002). According to Greene, this data supports the view that our "deontological intuitions" about cases such as the Trolley Problem are less reliable than judgments about those cases that are derived by "top-down" reasoning from more general consequentialist principles (Greene 2003, Greene et al. 2008). But for a compelling critique of this interpretation of the empirical data, see Berker (2009).

56 MORAL KNOWLEDGE

C2.P152 How would it be reasonable for you to respond to this information? On the view suggested by Singer, it seems as though you would still have a perfectly good way of conducting moral inquiry: namely, set aside your particular judgments (even those that seem obviously correct) and simply *decide directly* which general moral principles are true. The suggested picture seems to be one on which we are simultaneously unreliable when it comes to making particular moral judgments, but reliable enough in our judgments about general principles to have justified confidence in those judgments (a level of reliability that we manage to attain by studiously ignoring the particular moral judgments that we are naturally disposed to make, judgments that would otherwise corrupt our assessment of the principles). Although I think that this is a perfectly consistent epistemological picture, it is implausible, for at least two reasons.

C2.P153 First, the picture is implausible because it rests on an implausible picture of our psychology, one on which we might be in a position to confidently endorse a general moral principle, or choose one general principle over another, in a way that is completely uninfluenced by our sense of the intuitive plausibility of the concrete implications of the principles under consideration. Against this, it's natural to think that we would not even be in a position to fully appreciate the content of a general moral principle unless we grasped at least some of its concrete implications,³⁸ and that once these concrete implications are understood, our sense of their intuitive plausibility will inevitably influence our assessment of the general principle that implies them. Thus, the procedure "just decide directly which general principles to accept, without being influenced by your sense of the intuitive plausibility of what they imply about cases" seems at least somewhat psychologically unrealistic.

C2.P154 An ultimately more serious worry about the picture is epistemological. Consider the moment when the oracle informs you that you are unreliable even with respect to those particular moral judgments that strike you as obviously correct. In addition to undermining your confidence in your ability to make accurate particular moral judgments, receiving this information should also significantly undermine your confidence that you're reliable when it comes to determining which general moral principles are true on the basis of which principles seem true. Even if there is some kind of defeasible presumption that when a general moral principle seems true to you, you are entitled to endorse it as true, that defeasible presumption would not survive learning

³⁸ As William James once wrote: "No one sees further into a generalization than his own knowledge of the details extends" (1995: 65). But one need not go as far as James in order to accept the general substance of the criticism offered here.

that, when it comes to lower level moral claims, *what seems clearly to me is not a good guide to what is true.*

C2.P155 Here is another way of developing the epistemological worry about any proposed methodology that would have us bracket an entire class of judgments on the grounds that we are incompetent or unreliable when it comes to making judgments of that level of generality. Proceeding in this way seems tantamount to ignoring highly relevant evidence—or at least, tantamount to ignoring *what would be highly relevant evidence, if not for our own incompetence.* Imagine that in the scenario in which we are graced with the presence of the oracle, there are also beings much like us, who differ only in that they receive somewhat better news from him. In particular, the oracle informs them of the following:

C2.P156 I have some good news and some bad news for you. The bad news is that you're prone to making mistakes with respect to both general moral principles and lower level moral judgments. The good news is that you're at least good enough with respect to both types of judgments that it makes sense for you to check your judgments at different levels of generality against one another. It's not simply that when your judgments at different levels of generality turn out to be inconsistent with one another, you know that something has gone wrong somewhere. It's also true that when a judgment to the effect that a general principle is true turns out to cohere well with your lower level moral judgments, that's positive evidence in favor of the principle, and vice versa. So checking your judgments about principles and your judgments about cases against each other is definitely a good strategy.

C2.P157 Given that the beings in question resemble us in other ways, their lower level moral judgments would be among the very best evidence they would have for evaluating the truth or falsity of proposed moral principles. For them, setting their particular judgments aside would be a very bad methodological practice, one that amounts to deliberately ignoring large quantities of highly relevant evidence. Of course, if we are sufficiently incompetent when it comes to independently judging the truth of lower-level moral claims, then setting aside our lower level moral judgments when we assess general principles might be the best that we can do. (For us, setting aside our lower level moral judgments would have a clear rationale, while for them carrying on in exactly the same way would lack any rationale.) But it's not as though the procedure that we would then be left with—a procedure that would clearly be a terrible one to employ, and blatantly irrational, if used by beings who were at least minimally

58 MORAL KNOWLEDGE

competent in the relevant way—would be any better or moral reliable for that. That is, we should not take consolation in the fact that we are not proceeding irrationally, if what spares us from the charge of irrationality is our own incompetence and unreliability.

C2.P158 What all of this suggests is the following: the closest possible world (perhaps the actual world) in which we are unreliable even with respect to those lower level moral judgments that strike us as clearly true in optimal conditions for judging is *not* a world in which we are nevertheless entitled to have justified confidence that those general moral principles that seem clearly true to us *are* true. Rather, it is a possible world in which we should not have much confidence in the moral judgments that we find ourselves disposed to make at *any* level of generality. Skepticism about our ability to make particular moral judgments with at least a fair degree of accuracy in optimal conditions leads to skepticism about our capacity for accurate moral judgment more generally. And general skepticism about our capacity for making accurate moral judgments is not a view according to which we should privilege some moral judgments over others in virtue of their level of generality (cf. footnote 30): rather, it is a view according to which we should stop making (or relying upon) moral judgments altogether, to the extent that we are able to do so.

C2.P159 In this chapter, I've attempted to identify some of the strengths and limitations of the method of reflective equilibrium as an account of how we arrive at moral knowledge, or how we *should* arrive at our moral views. I distinguished between *modest* and *ambitious* interpretations of the method, and I argued that we should accept the method given a modest interpretation but reject it when it is understood ambitiously. On a modest interpretation of the method and its epistemological role, the capacity of reflective equilibrium reasoning to deliver moral knowledge typically depends on the reasoner's already having substantial moral knowledge that is not itself the product of such reasoning. What are the sources of this knowledge? The next two chapters address this question.

C3

3

Moral Knowledge from Others

C3.S1

3.1 The Moral Inheritance View

C3.P1 On an ambitious interpretation of the method of reflective equilibrium and its epistemological role, moral knowledge issues from a particular kind of reasoning or deliberation. In the last chapter, I argued that we should reject that picture: although trying to make one's overall moral view more coherent and systematic can sometimes lead to moral knowledge, it is not true that this is *the* way of arriving at moral knowledge, or even the primary way. Moreover, when reasoning of the relevant kind does deliver moral knowledge, the process typically draws upon moral knowledge that the reasoner already has. In this respect, reflective equilibrium reasoning is better regarded as a possible way of extending one's stock of moral knowledge, as opposed to a way of generating moral knowledge from a cognitive starting point that consists exclusively of moral views that fall short of knowledge, along with whatever relevant non-moral knowledge one has. On this picture, by the time the typical moral agent begins to engage in reasoning that recognizably approximates the method of reflective equilibrium, she already has substantial moral knowledge. This picture makes salient the following question: where does such prior knowledge come from, if not from earlier episodes of reflective equilibrium reasoning?

C3.P2 One obvious possibility is this: in the usual case, the fact that a person will have at least some substantial moral knowledge by the time she is in a position to engage in anything like reflective equilibrium reasoning is guaranteed by the fact that she will have absorbed some such knowledge from her social environment.¹ She finds that others follow certain norms and hold certain views, and she adopts these norms and views as her own, typically without much in the way of critical reflection. Perhaps she is told by her parents that a

¹ Notice that this suggestion is consistent with the possibility that typical human beings also have some innate moral knowledge, a possibility that I take to be a live one. At least some leading developmental psychologists argue that human infants as young as three months manifest moral beliefs in experimental settings. (See especially Bloom 2013 and the further references given there.) Of course, even if infants can be credited with full-fledged moral beliefs, it is a further question whether they can be credited with moral knowledge.

60 MORAL KNOWLEDGE

certain practice (e.g. hitting other children) is wrong; in other cases, she might simply find that certain norms are generally observed by those around her, and that occasional departures from them are met with disapproval. In a favorable case, a moral belief that she picks up in this way amounts to knowledge, because those members of her community from whom she acquired the belief (either via testimony or by the example of their own practice) already knew it to be true. This moral knowledge can then be exploited in her future reasoning and deliberation. On this view, it is not simply that other people can play a *heuristic* role in our arriving at moral knowledge, by (for example) calling our attention to the possibility that a given practice is wrong, or to reasons for thinking that it is, reasons that we then autonomously recognize as compelling for ourselves. (Although the current suggestion is consistent with the claim that other people can and frequently do play those roles.) Rather, the claim is that one way of gaining full-fledged moral knowledge is through the manifestation of our natural tendency to adopt beliefs that are held by those around us and that are presupposed by common practices. In favorable circumstances, the child counts as knowing in virtue of having acquired the views from reliable sources; her ability to *justify* moral views acquired in this way (either to herself or to others) by giving an account of why they are true, might come only later, if at all. In any case, being able to justify a moral view acquired in this way by citing the reasons in virtue of which it is true is not a necessary condition of knowing that it is.

C3.P3 Call this picture **The Moral Inheritance View**. In the rest of this section, I want to further explain The Moral Inheritance View (in part by contrasting it with an alternative view about the status of moral beliefs acquired in this way), consider some objections to it, and offer some reasons for thinking that it is true.

C3.P4 Let's begin by noting that, if in fact we arrive at moral knowledge in all of the ways that we arrive at non-moral knowledge, this provides some reason to think that the Moral Inheritance View is true. A major theme of recent epistemology is the extent to which everyone—regardless of age or intellectual maturity—is dependent upon others for most of what she knows about the world.² In general, we credit people with knowledge *via* testimony, even when they have little or no capacity to explain why the facts that they believe on the basis of the testimony obtain, and even when they have little or no

² Coady (1992) is a seminal work in bringing the topic of testimony to the fore within contemporary epistemology. Adler (2012) and Lackey (2010) are useful overviews of the literature. Audi (1997, ch.7) is a particularly sophisticated treatment. For a more general overview of recent developments in social epistemology, see Goldman and Blanchard (2016).

grasp of the evidence that leads those who report the relevant facts to believe as they do. Consider, for example, the relationship that the non-scientist stands in to most of the scientific knowledge with which she can plausibly be credited. In such cases, it is enough that the believer acquire her true beliefs from reliable sources, sources that she has reason to trust or, at least, no reason to distrust.

C3.P5 Given that much of what any fully mature, adult human being knows about the world consists in things that she believes because others believe them, the same is *a fortiori* true of children. It is safe to assume that most of a child's beliefs about geography are beliefs that she accepts because they are accepted by those around her. In the case of geography, it is natural to credit the child with *geographical knowledge*, so long as her geographical beliefs are true and those from whom she learned them (e.g. her parents) are reliable sources of information about the subject matter. In particular, the child need not engage in any reasoning, deliberation, or critical scrutiny of her geographical beliefs in order to know. In short, she need not do anything to *validate* her geographical beliefs, in order for those beliefs to count as knowledge.

C3.P6 Theoretical pressure to credit the child with full-fledged geographical knowledge in these circumstances arises from the realization that it is quite exceptional for fully mature adults to do anything to validate the vast majority of their geographical beliefs. Unless we are prepared to embrace a relatively sweeping skepticism about the extent of our geographical knowledge, we should allow that adults are in a position to know geographical facts when they rely on sources of information that are reliable and which they have no reason to distrust. By parity of reasoning, a child's geographical beliefs should count as knowledge as well, provided that the child arrives at those beliefs by relying on sources of information that are reliable and which she has no reason to distrust. We can then ask: are things otherwise with respect to the *moral* beliefs that the child picks up from her parents, or from her social environment?

C3.P7 On the Moral Inheritance View, the story about a child's inherited moral beliefs is the same as the natural story about her inherited geographical beliefs: she knows so long as the people from whom she picked up the beliefs know, and they are reliable sources of information whom she has no reason to distrust. The Moral Inheritance View is so-called in order to reflect the fact that, on the view in question, a child might acquire substantial moral knowledge from others even if she is an essentially *passive recipient* of the relevant beliefs. In particular, there is no assumption that the child has critically scrutinized or reflected upon these inherited beliefs in a way that would have

62 MORAL KNOWLEDGE

prevented her from arriving at false moral views in the counterfactual scenario in which those whom she treats as authoritative had told her false rather than true things about morality. Indeed, if that had been the case, she might have believed false moral claims just as firmly as she actually believes the true moral claims that (given her favorable circumstances) she counts as knowing.

C3.P8 Thus, it is a feature of the Moral Inheritance View that whether a child's inherited moral beliefs amount to knowledge is very much a matter of fortune. Compare: a person in favorable circumstances can acquire great wealth by inheriting it even though he has done nothing to deserve that wealth, or anything that it would be natural to count as an achievement, or to his credit: his relationship to the wealth that he inherits is simply that of passive recipient. Similarly, on the Moral Inheritance View, a child might inherit substantial moral knowledge even if she has done nothing that it would be natural to describe as a cognitive achievement on her part—nothing that we would think of as to her intellectual credit. The fortunate heir might have done nothing to deserve his wealth, as witnessed by the fact that, had he conducted his affairs in the same way but in less favorable circumstances, he would now be impoverished. Similarly, the child's cognitive life might be such that had she been born into a less morally enlightened environment, she would have relatively little or no moral knowledge. Nevertheless, on the Moral Inheritance View, this fact about the child's current cognitive capacities and dispositions is perfectly consistent with her inheriting substantial moral knowledge given her favorable circumstances, just as the fortunate heir's wealth is no less real despite its manifest contingency.

C3.P9 On an alternative picture, which I will call the *Critical Reflection Picture*, although a moral view which the child uncritically inherits from others might be true—and even known to be true by those from whom she inherits it—she is not yet in a position to know it herself. Rather, in order to know the relevant moral claim, she must understand why it is true; or be able to validate or justify it, perhaps by engaging in proto-forms of reflective equilibrium reasoning. She would at least need to be able to subject it to a level of critical scrutiny which would lead her to reject false moral claims if her social environment had presented such claims as true instead.³

³ Clearly, there is a range of possible views here, views that differ in what they require in order for an inherited moral belief to count as knowledge. (For this reason, I refer to the *Critical Reflection Picture*, which encompasses a range of views that are inconsistent with one another as well as with the Moral Inheritance View.) The less a view requires on the part of the child in order for her true moral beliefs to count as full-fledged knowledge, the more the view resembles the Moral Inheritance View; on the other hand, the more demanding a view is in what it requires of the child, the greater its divergence from the Moral Inheritance View.

C3.P10 Here is a challenge to the coherence of the Moral Inheritance View that, if cogent, would also suffice to establish the Critical Reflection Picture. In Chapter 1, I endorsed the common view that, if a person knows that something is the case, it follows that she is justified in holding the relevant belief. It is always incoherent to attribute knowledge to someone while denying that she is justified in believing what she knows. That claim is a claim about knowledge in general, one that covers both the non-moral and the moral cases: just as “she knows that it’s raining, but she isn’t justified in believing that it’s raining” is guaranteed to be false, so too “she knows that hitting other children is wrong, but she isn’t justified in believing that hitting other children is wrong” is guaranteed to be false. Now, according to the Moral Inheritance View, young children can have moral knowledge even if they are not in a position to justify the moral views that they know, in the sense of providing compelling reasons in support of those moral views. But (so the objection runs), if one is justified in believing anything that one knows, then a child who genuinely knows a moral claim *would* be in a position to justify that moral claim by providing reasons that support it. Hence, given the truth of the claim that knowing entails justified believing, a child who genuinely knows will be in a position to justify the moral views of which she has knowledge; thus, the Moral Inheritance View is false and the Critical Reflection Picture is true.⁴

C3.P11 However, this objection is not compelling. The mistake lies in the assumption that, if a person’s belief has a certain epistemic status (“justified,” in any sense in which it is plausible to think that justified believing is necessary for knowing), then it follows that the person can successfully engage in a certain activity (“justifying”) with respect to the content of that belief. But we should distinguish sharply between the *state of being justified* in believing p, and the *activity of justifying* one’s belief that p, in the sense either of providing reasons that speak in favor of p, or *showing* that one’s belief has a certain epistemic status. There are, I think, good reasons to reject the assumption that being in a position to successfully engage in the activity—or even having the capacity to successfully engage in the activity—is a necessary condition for being in the relevant state.⁵

⁴ In Chapter 1, I noted that Audi (2003: 235–9) defends the claim that some unjustified beliefs are knowledge. Notice that, if Audi is correct in holding that knowledge does *not* entail justified belief, the current objection to the Moral Inheritance View does not get off the ground. The response to the objection offered in the text above concedes to the objector that the relevant entailment holds.

⁵ The classic case for this was made by William Alston. See the essays collected in his (1989) book *Epistemic Justification*, especially essays 1, 3, and 9. Cf. Pryor (2000). The claim that an individual might be justified in holding a belief even if she is not in a position to successfully engage in the activity of providing reasons for that belief is more or less trivial on so-called “externalist” accounts of justified belief. But it remains compelling on plausible “internalist” accounts of justification, such as those espoused by Alston and Pryor. A particularly vigorous defence of the claim that knowing that p does

64 MORAL KNOWLEDGE

- C3.P12 To see this, consider the non-moral case. A pre-linguistic child might see a mouse run into a hole, and thus know that the mouse ran into the hole, in more or less exactly the same way that a fully mature adult who has witnessed the same scene knows the same fact. The pre-linguistic child's belief that the mouse ran into the hole is thus justified, in any sense of "justified" in which knowledge requires justified belief. (Contrast, for example, the status of some true belief that the child holds as a result of wishful thinking.) Nevertheless, this is consistent with her inability to justify her belief, in the sense of being able to produce compelling reasons for it, in the way that the adult can. More generally, the ability to justify a belief by providing compelling reasons for thinking that it is true is a relatively sophisticated intellectual ability, while knowledge is something that we credit to quite unsophisticated subjects, including young children and some non-human animals.
- C3.P13 Consider also the phenomenon of preservative memory. In the usual case, when a person genuinely remembers that p (where p is some non-moral proposition), she knows that p. On the assumption that she is justified in believing anything that she knows, it follows that she is justified in believing p. It does not follow that she is in a position to justify her belief that p in the sense of providing evidence for p, for her genuinely remembering p does not guarantee that she will also remember the evidence in its favor.⁶ In the non-moral case, we should reject the inference from "S knows that p" to "S can justify her belief that p by providing reasons for thinking that p is true." Given that we should not assume that the standards for moral knowledge are higher than the standards for non-moral knowledge (cf. section 1.2), we should reject the relevant inference in the moral case as well.
- C3.P14 Even if the preceding objection to the Moral Inheritance View is not compelling, the question remains: what reason is there for thinking that it is true? Again, the presumptive case for the Moral Inheritance View over the Critical Reflection Picture consists in this: the standard for moral knowledge that it employs is the same standard for knowledge that we ordinarily assume in other domains. Those of us who have learned from reliable sources that Fermat's Last Theorem is true are in a position to know that it is, even if we have little or no understanding of its proof, and even if we would believe it with equal conviction in a counterfactual scenario in which prominent

not require an ability to successfully engage in the activity of justifying one's belief that p can be found in Dretske (2000).

⁶ On the importance of "forgotten evidence" cases for epistemology, see especially Harman (1986a) and Goldman (2009).

mathematicians had falsely reported that it had been proved even though it had not been. Similarly, we unhesitatingly credit an adult with the knowledge that Cheyenne is the capital of Wyoming when she has learned this from reliable sources, even if she has done nothing to verify it, and even if she would hold a false belief about the capital of Wyoming in the counterfactual scenario in which she had received misleading testimony. In the absence of some compelling reason for thinking that things are otherwise, we should prefer a unified account of knowledge, on which the standards that we must meet in order to count as knowing a proposition do not vary from domain to domain. And this desideratum, together with the standards that we employ in attributing knowledge and evaluating knowledge claims in other domains, gives us a reason to prefer the Moral Inheritance View over the Critical Reflection Picture.

C3.P15 If the Moral Inheritance View is true, then that supplies part of an answer to the question of how it is possible for a typical moral agent to come to have moral knowledge in the first place, if not from reflective equilibrium reasoning. Of course, even if the Moral Inheritance View is true, it hardly provides a satisfying answer to the question of “where moral knowledge comes from,” in the sense in which philosophers have often pursued that question. For the picture immediately makes salient the question: how did those from whom the moral agent inherits this knowledge acquire it? Perhaps they too originally acquired the knowledge by inheriting it from others in the same way. But clearly, the mechanism of inheritance cannot explain how the relevant truths came to be known in every case.

C3.P16 Even once the modesty of the Moral Inheritance View is emphasized—in particular, even once it is emphasized that the view does not purport to provide a solution to the philosophical puzzle of “where moral knowledge comes from”—doubts about it, and about its alleged superiority to the Critical Reflection Picture, might remain. Consider again the comparison between moral knowledge and geographical knowledge. Above, I suggested that when a child inherits true geographical beliefs from a reliable source that she has no reason to distrust, there is the following compelling reason to credit her with full-fledged geographical knowledge: the child’s position with respect to her geographical views does not seem markedly different from the average *adult’s* position with respect to a great number of *his* views about geography. On pain of an implausible skepticism about the extent of our geographical knowledge, we credit the adult with geographical knowledge when he learns from reliable sources that he has no reason to distrust. We should thus also credit the child with geographical knowledge in the same circumstances. But, one might think, it is exactly at this point where a crucial disanalogy between the case of

66 MORAL KNOWLEDGE

morality and the case of geography arises. For although we think that it is natural, inevitable, and perfectly appropriate for a normal adult to rely on authority for most of what he believes about geography and many other subject matters, we do *not* similarly think that it is natural, inevitable and perfectly appropriate for a normal adult to rely on authority for his *moral* views.

C3.P17 In support of the alleged contrast, consider the following case:

C3.P18 *The Deferential Spouse.* Your friend Al informs you that he is passionately opposed to eating meat because doing so is immoral. You ask him why he thinks this: is there, for example, some particular argument against the practice that he finds convincing? He replies that there is no such argument. In fact, the arguments that are typically offered against eating meat strike him as weak and unconvincing: as far as his own moral sensibility is concerned, doing so seems morally unproblematic. For this reason, if he were to make up his own mind, he would believe that eating meat is perfectly permissible. Nevertheless, he firmly believes that doing so is wrong, and indeed, takes himself to know that it is. His staunch opposition to eating meat is due to the fact that his wife has told him that doing so is wrong, and his belief that, in general, her moral judgment is more reliable than his own. Indeed, because he believes that her moral judgment is more reliable than his, he has adopted a general practice of adopting her moral view in any case in which her expressed view conflicts with what he would judge to be the case if he made up his own mind. His opposition to eating meat, he reports, is simply a manifestation of intellectual humility within the moral domain.

C3.P19 Many people are struck by the sense that there is something peculiar, problematic, or “fishy” (Enoch 2014) about the possibility of simply outsourcing one’s moral convictions to another person in this way. In contrast, there does not seem to be anything similarly peculiar or problematic about outsourcing one’s opinions about any number of purely factual, non-moral subject matters.⁷

⁷ Indeed, R.M. Hare once claimed that one of the things that *distinguishes* “any serious moral problem” from purely factual questions is that:... a man who is faced with such a problem knows that it is his own problem, and that nobody can answer it for him. He may, it is true, ask the advice of other people; and he may also ascertain more facts about the circumstances and consequences of a proposed action, and other facts of this sort. But there will come a time when he does not hope to find anything else of relevance by factual inquiry, and when he knows that, whatever others may say about the answer to this problem, *he* has to answer it (1962: 1).

Of course, one might think that in any sense in which it is true that a person must ultimately make up his own mind about a serious moral question with which he is concerned, it’s equally true that he must ultimately make up his own mind about *any* question that he considers. But Hare explicitly denies that this is so: the sense in which each of us is ultimately responsible for arriving at his or her own moral convictions is stronger than any sense in which we are responsible for arriving at our purely factual, non-moral opinions about how the world is arranged (p. 2).

For example, suppose that your friend believes both that he has a poor sense of direction and that his wife's sense of direction is excellent. Given this, there does not seem to be anything peculiar or problematic about his confidently believing that (for example) *we should turn right at the next intersection* on the basis of his wife's say-so, even if it very much *seems to him* as though they should turn left. Nor does there seem to be anything peculiar or problematic about his adopting a general practice of deferring to his wife about the directions, given his background belief in the superiority of her sense of direction.

C3.P20 In the moral case, there are at least two different (although related) aspects of the deferential spouse's practice that might seem suspicious. First, he believes that eating meat is wrong solely as a result of accepting his wife's testimony. Second, it is not simply that the deferential spouse accepts his wife's testimony about this particular moral question, in the way that he might if he believed that she just happened to be (for one reason or another) in a better position to answer it. Rather, his accepting his wife's testimony is part of a more general practice of outsourcing his moral convictions to her. In this respect, he treats his wife as a *moral expert*. As we will see, many philosophers, including many philosophers who see nothing objectionable about relying on moral testimony *per se*, have denied that there are genuine moral experts in anything like this sense, or people it would be reasonable for us to treat as such.

C3.P21 In addition to their intrinsic interest, issues about moral testimony and moral expertise are directly relevant to the current study for at least two reasons. First, if, as some suppose, testimony or deference to expert opinion play (or should play) little or no role in the moral domain, then this itself would constitute a significant difference between the epistemology of morals and the epistemology of many other domains. Second, these issues bear directly on the credibility of the Moral Inheritance View. The Moral Inheritance View entails that it is possible for children to acquire full-fledged moral knowledge by accepting the testimony of morally competent members of their community. Moreover, on the Moral Inheritance View, the relationship that young children stand in to the morally competent adults in their communities is at least very similar to (if not simply the same as) the relationship that the moral non-expert would stand in to the moral expert. According to the Moral Inheritance View, in favorable cases the moral views that children acquire from their social environment are views that they know to be true, even if they lack the ability to justify those views. So too the deferential spouse credits himself with knowledge of moral claims that he cannot justify, and for which (he freely admits) he fails to grasp the reasons. If, as some would argue, there is something deeply confused about the idea that the deferential spouse can

68 MORAL KNOWLEDGE

arrive at full-fledged moral knowledge in this way, then that casts serious doubt on the Moral Inheritance View.

C3.P22 Consider again the working hypothesis endorsed in Chapter 1, according to which we can gain and lose moral knowledge in all of the ways that we gain and lose non-moral knowledge. From the perspective of that working hypothesis, perhaps the most straightforward and satisfying story about moral testimony and expertise would be this: there are *no* deep or philosophically significant differences between the moral and the non-moral cases—the deferential spouse’s practice of outsourcing his moral convictions to his wife is on all fours with his practice of deferring to her sense of direction, and any impression to the contrary is an illusion. Although that conclusion would be particularly congenial to the working hypothesis that I uphold, I don’t think that it is true. I think that there *are* philosophically significant differences between the moral and the non-moral cases, although not of a kind that provide good reasons to reject the hypothesis. Exploring these differences, and explaining how their existence is consistent with the hypothesis, are among the central tasks of the rest of this chapter.

C3.S2

3.2 Testimony, Deference, and Expertise

C3.P23 In this section, I’ll explore some of the relationships between moral testimony, moral deference and moral expertise. I’ll suggest that what are often regarded as questions about moral testimony are better regarded as questions about the broader phenomenon of moral deference. I’ll then consider a series of cases in which one person arrives at a moral view by deferring to another. These cases range from ones in which the practice seems maximally unproblematic to cases in which its propriety seems increasingly questionable. This exercise will allow us to zero in on the most plausible forms of skepticism about moral deference and moral expertise, which I will focus on in section 3.3.

C3.P24 Consider again the case of the deferential spouse introduced above. Similar cases are often used to motivate philosophical questions about the status of moral *testimony*: for example, questions about the propriety of forming moral views on the basis of testimony, or questions about the ability of testimony to deliver moral knowledge.⁸ However, I believe that the philosophically interesting issues in this vicinity are not about moral testimony per se, but about a broader phenomenon. After all, while testimony is one common route by

⁸ See, for example, Nickel (2001), Hills (2009), Hopkins (2007), Groll and Decker (2014), and Sliwa (2012).

which we learn the moral views of others, there are any number of others.⁹ Perhaps the deferential spouse learns that his wife believes that eating meat is wrong by hearing her testify to that effect. But we can also imagine that he learns of her view in some other way. Suppose, for example, that he's noticed both that she consistently refrains from eating meat and that an unmistakable (though involuntary) look of moral disapproval crosses her face whenever someone else orders or consumes meat in her presence. As before, he adopts her strongly negative attitude towards eating meat as his own, despite the fact that he would judge the practice morally unproblematic if he made up his mind autonomously. On the face of it, this seems no less (and no more) puzzling or problematic than a case in which he defers to his wife about the morality of eating meat on the basis of her testimony.¹⁰

C3.P25 Let us use the term *deference* for cases in which one holds a view solely because another person holds that view. For example, if your friend Al believes, solely on the basis of his wife's say-so, that they should turn left rather than right at the next light in order to reach their destination, then he defers to her about which way they should turn.¹¹ My claim then, is that despite what might be suggested by the terms in which they are often discussed, the philosophically interesting issues in this vicinity concern not moral testimony per se, but rather the broader phenomenon of moral deference.¹²

C3.P26 Does the fact that someone else holds a moral view—say, a view to the effect that a given practice is morally wrong—ever give us a good reason to adopt that view in cases in which we are not in a position to appreciate whatever reasons exist for thinking that it is true? There are at least some cases in which moral deference seems clearly warranted. Here is one:

C3.P27 **Case 1.** Initially, Al suspends judgment about whether ϕ -ing is wrong, because he knows that he is ignorant about various non-moral facts about

⁹ In fact, I suspect that moral testimony accounts for only a modest fraction of the extensive knowledge that we have of the moral opinions of other people. For example, consider all of the people of your acquaintance to whom you would unhesitatingly and confidently attribute the view that *slavery is unjust*. In how many cases is this based on the person's *testimony* (even in the relatively wide sense of "testimony" in which philosophers often use the term)?

¹⁰ Similarly, according to the Moral Inheritance View, young children can inherit full-fledged moral knowledge not only by accepting testimony from morally competent members of their community, but also by absorbing true moral views from their social environment in other ways, for example, by observing that certain norms of behavior are generally observed, or by participating in practices that naturally encourage certain beliefs as opposed to others.

¹¹ Notice that here, holding a view "solely because another person holds that view" does not imply or suggest that one lacks independent reason to think that the person is reliable. (Indeed, Al's decision to defer to his wife about which way they should turn might be due to his background knowledge that her sense of direction is superior to his.)

¹² I first offered this claim in McGrath (2010a). Enoch (2014) suggests that the phenomenon of philosophical interest is broader still, and is present in any case in which a person holds a moral view on the basis of what he calls "opaque evidence." On opaque evidence, see section 3.4 below.

70 MORAL KNOWLEDGE

ϕ -ing and its consequences that he takes to be bear on its morality. He then learns that his wife is morally opposed to ϕ -ing. Because he knows that he and his wife share the same basic moral outlook but that she is well informed about the relevant non-moral facts, he defers to her judgment.

C3.P28 And here is a more radical case of the same kind:

C3.P29 **Case 2.** The situation is the same as in Case 1, except for the fact that Al initially judges that ϕ -ing is morally permissible on the basis of the information available to him (as opposed to merely suspending judgment about its status). He reverses his judgment upon learning that his wife thinks otherwise, because he reasonably assumes that the fact that they arrived at different views is due to the fact that she has relevant non-moral information that he lacks.

C3.P30 In these cases, it's reasonable for Al to defer to his wife's judgment. In general, given that the truth of many moral claims depends on the non-moral facts, the claim that we can have good reasons to defer to someone else about a particular moral issue seems to follow more or less immediately from the fact that we often have good reasons to defer to others about non-moral facts. In deferring to his wife's judgment in these circumstances, Al need not assume that she is privy to some piece of purely moral information or moral insight which he is unable to appreciate. Rather, he reasonably assumes that she judges as he would judge, if he too were aware of the relevant non-moral facts. In cases such as these, moral deference does not seem to raise any additional philosophical issues, beyond those that arise for deference concerning the relevant non-moral facts. Rather, deference here seems to approximate deference to one's better-informed self, and the reasons that one might have to engage in such deference might at least in principle be just as decisive as the reasons that one would have to adopt the moral views of one's better informed self.¹³

¹³ Cf. Allan Gibbard's (1990) discussion of "contextual" authority, which "presupposes a context of shared norms." He writes:

Might I take someone else's accepting something normative as my own reason for accepting it? In some cases I might well. Suppose that you tell me it made no sense for Cleopatra to be angry at the messenger. I am ignorant of history, perhaps, and confident that you know your history, and that you and I share the same basic norms for anger. In that case, I can take your normative reasoning as proxy for my own. I think that you are reasoning just as I would if I knew the facts. I can let you draw normative conclusions for me, and so I take the fact that you draw a normative conclusion as reason for accepting it myself.

(p. 174)

(As this passage suggests, Gibbard is not specifically concerned with the case of morality but with the normative realm more generally.)

C3.P31 Are there cases in which moral deference is warranted, but where this is not due to differences in non-moral information? I think that there are. Consider, for example:

C3.P32 **Case 3.** Al knows that he has all of the non-moral information about ϕ -ing that his wife has, and that they share the same basic moral outlook. If he made up his own mind, Al would judge that ϕ -ing is morally permissible. Nevertheless, he defers to his wife's view that ϕ -ing is wrong, because he knows that, with respect to this particular issue, his judgment is compromised in a way that hers is not.

C3.P33 Perhaps Al knows that, when he thinks about this particular issue, his ability to apply the basic moral outlook that he and his wife share is compromised by brute self-interest. If he knows that his wife's judgment is not similarly compromised, then it seems as though he might reasonably view deference to her unimpaired judgment as deference to his true self, in much the same way that one might view deference to a person of similar moral outlook who possesses superior knowledge of the relevant non-moral facts as deference to one's better-informed self.

C3.P34 Consider next a more extreme case:

C3.P35 **Case 4.** Left to his own devices, Al would judge that ϕ -ing is morally permissible. Nevertheless, he defers to his wife's judgment that ϕ -ing is wrong, not because he believes that there is something about this issue in particular that puts her in a better position to judge, but as part of a general practice of deferring to her moral views in cases in which they are disposed to disagree.

C3.P36 Clearly, Case 4 represents a much more radical possibility than those depicted in Cases 1-3. Nevertheless, given that Al's practice in Cases 1-3 can be rational, it seems as though there are possible circumstances in which he would have at least a defeasible reason to engage in the practice described in Case 4. Consider again Cases 1 and 2, in which Al defers to his wife's moral judgment about some issue because he knows that she is better informed about the non-moral facts. At least in principle, it seems that he might believe, and even have good reason to believe, not just that she is better informed about the non-moral facts relevant to deciding this or that particular moral issue, but better informed about the kinds of non-moral facts that are relevant to deciding moral issues more generally. In that scenario, it seems that he would have a defeasible reason to privilege her moral judgments over the moral judgments

72 MORAL KNOWLEDGE

he is disposed to make on his own. (We can assume that, in this hypothetical scenario, his wife habitually shares her moral opinions with him but is not in the habit of sharing the non-moral information that informs those opinions.) If this is in fact what motivates Al's practice in Case 4, then it seems as though that practice has a clear rationale: it is once again tantamount to deferring to one's better informed self.

C3.P37 Similarly, we can imagine that Al's practice in Case 4 is motivated by factors that are continuous with those that are present in Case 3. That is, we can imagine that Al has good reason to believe that his moral judgment has been compromised or impaired, not just with respect to some specific moral question, but more generally. For example, perhaps he knows that his natural moral sensibility has been compromised by participation in a brutal war, or that depression has deadened his natural affective responses in a way that he takes to adversely affect his moral judgment. If he knows that he and his wife shared the same basic moral outlook prior to his entering into his current state, then it seems as though he might reasonably view a general practice of moral deference as the best way he has of approximating the moral judgments of his true self.

C3.P38 Consider again the scenario in which Al consistently defers to his wife about moral questions because he believes that his wife is generally better informed about the kinds of non-moral facts that are relevant to deciding such questions. Notice that, even in that scenario, Al might very well combine his practice of across-the-board moral deference with the thought that his wife would *not* be more reliable about moral questions if they were equally well-informed about the non-moral facts. That is, he might very well think that, conditional on their having the same non-moral information that bears on a given moral question, she would be no more likely than he is to arrive at the correct view. (Indeed, he might even coherently think that she would be *less* likely to arrive at the correct moral view than he, conditional on their having the same non-moral information.) In this respect, he need not think of his wife as someone who is privy to purely moral information or insight that he lacks.

C3.P39 Contrast the following case:

C3.P40 **Case 5.** Al has all of the non-moral information relevant to the morality of ϕ -ing that his wife has, and he has no reason to think that his own judgment is impaired or compromised, either with respect to this issue or more generally. Indeed, he regards himself as a perfectly competent (although fallible) judge of moral matters. On the basis of this information and his own careful consideration of the issue, it seems to him that ϕ -ing is perfectly permissible.

Nevertheless, he judges that ϕ -ing is wrong because he knows that this is what his wife thinks, and he has adopted a general practice of deferring to her moral views in cases in which he is disposed to judge differently. He has adopted this policy because he believes that his wife is a *moral expert*.

C3.P41 As noted above, that there are or could be moral experts in anything like this strong sense has frequently been denied, including by philosophers who do not deny that moral knowledge can be transmitted by testimony. If this is right, then it would amount to a potentially significant difference between our cognitive relationship to morality and our cognitive relationship to many other subject matters. In the next section, I want to look in some detail at skepticism about moral expertise. For purposes of that discussion, a significant lesson that has emerged from the current section is this: given a natural understanding of what would be involved in moral expertise, skepticism about such expertise is *compatible* with the view that even relatively sweeping deference to another person's moral views is sometimes in order.

C3.S3

3.3 Doubts About Moral Expertise

C3.P42 Let us begin by considering some expressions of skepticism concerning moral expertise. Bernard Williams once wrote:

C3.P43 There are, notoriously, no ethical experts... Anyone who is tempted to take up the idea of there being a theoretical science of ethics should be discouraged by reflecting on what would be involved in taking seriously the idea that there were experts in it. It would imply, for instance, that a student who had not followed the professor's reasoning but had understood his moral conclusion might have some reason, on the strength of his professional authority, to accept it... These Platonic implications are presumably not accepted by anyone. (1995: 205)

C3.P45 Williams' last claim is an overstatement: some do accept the "Platonic implications" that he disavows.¹⁴ Nevertheless, even if Williams' skepticism about

¹⁴ For example, Enoch (2014) explicitly casts his paper as a "refutation" of what Williams says in this passage (p. 229). Peter Singer is perhaps the most prominent contemporary advocate of the idea that moral philosophers are well suited to play the role of moral experts by dint of their academic training. See especially Singer (1972) and (1988). The passage from Williams is also discussed by Hills (2009).

moral expertise is not universally shared, it is far from idiosyncratic. For example, according to Mary Warnock, “there is no such thing as a moral expert” (1985: 95), a claim that is echoed by Mark Schroeder (2007:175). According to Nicholas Rescher, “the very idea of moral expertise...is totally unrealistic” (2005: 200). For his part, Michael Smith (1994: 5) holds that “our moral life seems to presuppose that [moral] facts are available to all, that no one in particular is better placed to discover them than anyone else” (1994: 5; Cf. Gert 2004: 90). Moreover, even some of those who disagree with Williams’ specific claims suspect that there is some significant truth in their vicinity.¹⁵

C3.P46 At a minimum, there is considerable intuitive pull to Williams’ thought that there is an interesting difference in this vicinity between ethics (in anything like its current form) and our paradigms of “theoretical sciences.” By way of illustration, consider the contrast with physics. When I lived in Cambridge, Massachusetts, I occasionally attended public lectures given by members of the MIT physics department, lectures designed for “the interested layperson.” At these events, the physicists explained the latest discoveries or results in their subfields. Often, the reported discoveries were extremely surprising and counterintuitive, the exact opposite of what one would have expected if one had formed one’s beliefs about the relevant parts of physical reality on the basis of folk physics, or even on the basis of the last physics class one took years ago. The physicists generally made at least some effort to gesture at why these counterintuitive views were now believed by those working at the cutting edge of the relevant subfield. But for the most part, that aspect of the presentation was quite cursory: there simply wasn’t time to go into all of the considerations that the scientists had for thinking what they did. Even if there had been the time, the typical audience member would not have understood those considerations in any depth. To the extent that a typical member of the audience changed her views in response to the presentation, she was more or less taking the physicist’s word for it. Nevertheless, people generally did believe what was reported, despite the counterintuitiveness of the claims, and despite their lack of any real grip on the evidence for those claims. Moreover, this was, I assume, a perfectly reasonable thing for them to do.

C3.P47 Leading physicists are *full-blooded experts* about their specialties, in the following sense: a perfectly normal, intelligent layperson has compelling reason to *blindly defer* to the physicist about physics, adopting the physicist’s

¹⁵ Thus, commenting on the passage quoted above, and immediately after announcing his intention to contest Williams’ picture, Enoch (2014) declares that “it cannot be denied that Williams is on to something here, indeed, something important about moral deference and expertise, and presumably something about moral testimony as well” (p. 229).

views as her own, even in the absence of an appreciation of the evidence for those views, and even if those views contradict the layperson's own sense of how things are. In the same way, leading biologists are full-blooded experts about biology, and so on. Williams clearly thinks that there are no full-blooded experts about morality: that is, there is no group of people who stand to morality as leading physicists stand to physics and leading biologists stand to biology.¹⁶

- C3.P48 If it is true that there are no full-blooded moral experts in the way that there are full-blooded experts about various other subject matters, then this is presumably not a brute fact; rather it reflects some *other* difference between morality and those subject matters. That is, if Williams and others who deny the existence of full-blooded moral experts are correct, then there is presumably some explanation of this fact. In the rest of this section, I want to consider some potential explanations that have been offered for the putative non-existence of moral experts.¹⁷

C3.S4

3.3.1 Theory vs. Anti-theory

- C3.P49 We can begin by considering an idea that is naturally suggested by the passage from Williams quoted above. There, Williams attempts to leverage a certain discomfort with the idea of moral expertise in order to cast doubt on the possibility of a “theoretical science of ethics.” On one natural reconstruction, the line of thought suggested there runs as follows:

- C3.P50 If there were a theoretical science of ethics, then we would expect to find experts in that science, who stand to it as physicists stand to physics and as biologists stand to biology. In particular, we would expect to find moral

¹⁶ Cf. David Estlund (2008): 106. Of course, the thought that there are no full-blooded moral experts is consistent with the existence of moral experts in any number of various weaker senses, a point which Williams himself is concerned to emphasize (1995: 206–8). Moreover, for reasons given in the previous section, the thought that there are no people who stand to morality as physicists stand to physics is also consistent with the view that there are possible circumstances in which even a relatively sweeping deference to another person's moral opinions would be in order. In what follows, “moral expert” will always mean *full-blooded moral expert*, in the sense given in the paragraph to which this footnote is attached.

¹⁷ In considering these potential explanations, we need not assume, with their proponents, that it's *true* that there are no moral experts. Indeed, because the existence of full-blooded moral experts is a contested matter, the quality of the explanations that have been offered for their putative non-existence is itself evidence that is potentially relevant to the existence question. If there are plausible explanations for why morality, unlike, say, physics, is not the kind of thing about which we would expect to find full-blooded experts, then this might increase our confidence that there are no such people. If, on the other hand, there is no good explanation, then this might give us reason to suspect that there are or might be full-blooded moral experts after all.

76 MORAL KNOWLEDGE

experts, whose status as such makes it reasonable for non-experts to blindly defer to their pronouncements about morality, in the same way that it is reasonable for non-specialists to blindly defer to physicists about physics and to biologists about biology. But in fact, there is no group of people who stand to morality as the physicist stands to physics and as the biologists stand to biology. Therefore, there is no theoretical science of ethics.¹⁸

C3.P51 The way that Williams frames the issue might encourage the thought that if we simply dispensed with the idea that ethics either is or could be a “theoretical science,” this would explain the apparent absence of full-blooded moral experts. In fact, the idea that there is an intimate connection between the questions of (i) whether there are or could be moral experts, and (ii) whether ethics is or could be a theoretical science is a longstanding one.¹⁹ However, there are reasons to think that the issue of whether ethics is or could be a theoretical science does not get to the heart of the matter.

C3.P52 In order to see this, suppose that some version of *moral particularism* is true.²⁰ Although a sufficiently virtuous person will recognize what she is morally required to do on particular occasions, she does not arrive at this knowledge by applying general moral principles to the circumstances in which she finds herself. Rather, knowing what to do on particular occasions typically requires the exercise of judgment, and the set of true moral judgments about particular cases does not admit of theoretical codification, even in principle. If this were the truth about morality, then I assume that it would be enough to preclude any possibility of a “theoretical science of ethics.” Still, this would

¹⁸ In this context, it’s worth reflecting on what we would think in the (perhaps unlikely) event that ethics were to emerge as a full-blown theoretical science, on a par with the most progressive theoretical sciences with which we are currently acquainted. In the concluding section of *Reasons and Persons* (1984), Derek Parfit claims that “Non Religious Ethics” is “the youngest and least advanced of the sciences,” one that has been systematically studied by a significant number of people only since the 1960s. In view of this, Parfit provocatively suggests, it is not irrational to hope that future progress will be greater in ethics than in any of the other sciences, and that ethicists, like mathematicians, will ultimately come to unanimously agree in their substantive conclusions (p. 454). Perhaps if anything like this actually occurred, the existence of full-blooded moral experts would become as uncontroversial as the existence of expert physicists or mathematicians. It might come to seem perfectly natural for professional ethicists to give public lectures for “interested laypeople” concerned to know more about the content of their moral obligations, in which the latest moral theorems were reported. And any sense there is some interesting or important difference here between morality and other subject matters might disappear completely.

I take it that Williams would not think that the relevant kind of convergence is a genuine possibility in the case of ethics. See especially chapter 8 of his (1985) “Knowledge, Science, Convergence.”

¹⁹ See, for example, Hastings Rashdall’s 1894 *Ethics* paper “The Limits of Casuistry.”

²⁰ On particularism, see especially Dancy’s recent (2013) survey and bibliography. Aristotle’s *Nicomachean Ethics* is often viewed as the historical source for key particularist themes; McDowell (1979) is a seminal work for the modern discussion.

not explain the putative non-existence of full-blooded moral experts, or account for any putative asymmetry between moral deference and deference about other subject matters. After all, in many non-moral contexts, we are prepared to sweepingly defer to others about what to do in particular cases, even when we do not suspect that they have some general theoretical knowledge that we lack.

C3.P53 As an example of this, consider again the case of the man who consistently defers to his wife about directions. Perhaps as they approach an intersection while driving together, *it seems to him* that they should turn right in order to reach their destination. However, his wife informs him that they should turn left, and he knows from past experience that her sense of direction is strong while his is weak. He thus unhesitatingly overrules his own phenomenology and concludes that they should turn left. Given that he regards his wife's sense of direction as superior, his proceeding in this way seems utterly natural and perfectly appropriate. Moreover, the same seems true of his adopting a general practice of deferring to his wife, when it comes to judgments of the relevant kind.

C3.P54 When a person with a reliable sense of direction arrives at knowledge of which way to turn on a particular occasion, this is typically not a matter of her applying some general principle or body of theoretical knowledge to her circumstances. Rather, it involves the successful exercise of an underlying capacity or competence.²¹ In at least this respect, knowing which way to turn on a particular occasion seems like a good model for the way in which the neo-Aristotelian moral particularist would have us understand moral knowledge of what to do on a particular occasion.

C3.P55 Compare the moral analogue, in which the deferential spouse outsources his moral judgments to his wife, even when doing so requires ignoring his own disposition to judge differently, and even in the absence of any appreciation of her reasons for judging as she does. We can distinguish two cases here. In the first case, the man defers to his wife because he believes that she knows the correct *theory* of morality, and that it is her application of this theoretical knowledge that allows her to judge correctly on particular occasions. In the second case, she has explicitly disavowed any such theoretical knowledge:

²¹ Indeed, notice that if the man asks his wife what reason there is for thinking that they should turn left as opposed to right, she might not be in a position to cite some particular consideration in response to his request. Perhaps she has simply recognized, now that they are upon the intersection, that turning left is the thing to do. (Contrast a case in which, for example, she can point out that they passed a certain gas station on their original journey, and now need to turn left in order to pass it again and accurately retrace their steps.)

she simply judges, in the manner of a would-be Aristotelian *phronimos*, that such-and-such an action is the thing to be done in the circumstances, and he adopts these judgments as his own. Is the man's practice any less puzzling or problematic in the second case than in the first? To my mind, the second case seems no less (and no more) puzzling or problematic than the case that Williams himself mentions, in which a student arrives at a moral conviction by blindly deferring to a professor whose reasoning he has not followed. (And to the extent that I can get myself in a state of mind in which there does *not* seem to be anything puzzling or problematic about outsourcing my moral convictions to someone whom I have come to regard as a *phronimos*, blindly deferring to someone whose putative moral expertise consists in their allegedly having superior theoretical knowledge of morality seems similarly unpuzzling and unproblematic.) Thus, the issue of whether ethics is or could be a “theoretical science” is something of a red herring in the present context. In particular, even if it's true that ethics is not the sort of thing about which there could be a theoretical science, that would not provide a good explanation of the absence of full-blooded moral experts, or of any asymmetry that might be thought to exist between moral deference and deference with respect to various other subject matters.

C3.S5

3.3.2 Realism and Its Rivals

C3.P56 Although Williams explicitly connects the topic of moral expertise with the question of whether ethics is or could be a “theoretical science,” he hints at another idea when he dismisses the possibility of full-blooded moral expertise as an unacceptable “Platonic implication.” The charge of “Platonism” is often brought against any broadly *realist* view of morality.²² And it's natural to think that, if moral realism is true, there would be no significant asymmetry between moral expertise and expertise about other subject matters. According to the moral realist, there is a domain of moral facts, and those facts obtain independently of what we believe, or our reasoning about them. The proper aim of moral inquiry is to discover the moral facts, and the proper aim of moral judgment is to depict or represent them accurately. Paradigmatic moral judgments are thus attempts at *property detection*. When one judges that (for example) “Eating meat is wrong,” one attributes a certain property to a certain

²² A classic statement of the charge is Mackie (1977), chapter 1.

type of action, and one's judgment is true just in case actions of that type have the property.

C3.P57 Now, it seems as though as soon as this realist picture is in place, it would be perfectly possible for me to come to believe that someone else is a more reliable detector of the presence of the relevant properties than I am, in which case it would be perfectly natural and normatively appropriate for me to defer to the other person on particular occasions, even if I'm strongly inclined to judge differently. Moreover, it seems to make no difference whether I believe that she is more reliable because she has superior theoretical knowledge of the Science of Ethics, or rather because she is an Aristotelian *phronimos* (or at least, a better approximation to an Aristotelian *phronimos* than I am), or for some other reason entirely.

C3.P58 In this context, it is notable that some metaethical views that explicitly reject the picture of moral judgment as property detection would seem to readily account for the non-existence of moral experts. Consider, for example, A.J. Ayer's classic (1936) emotivist account of moral judgment. Famously, Ayer held that it is a fundamental mistake to think that a paradigmatic moral judgment such as "Eating meat is wrong" involves attributing a property to a class of actions. Rather, in judging that eating meat is wrong, one is *expressing* one's disapproval of eating meat. Notice that this account of moral judgment has no difficulty in accounting for the putative non-existence of full-blooded moral experts, or for a putative asymmetry between moral and non-moral deference. For if judging that eating meat is wrong is really a matter of expressing one's own emotional disapproval towards eating meat, then so long as I currently feel no disapproval towards meat eating, the mere fact that someone else judges that meat eating is wrong seems to be neither here nor there.²³

C3.P59 Similarly, the metaethical *constructivist* might seem to be in a strong position to account for the putative non-existence of full-blooded moral experts. For the constructivist, like the noncognitivist, will insist that it is a deep confusion to suppose that there is some independent realm of moral facts or truths that we are attempting to bring into view and to which our moral judgments are answerable. For example, a Kantian constructivist might hold that, whenever we engage in moral deference, we are guilty of misconstruing the nature of the moral domain: we are in effect treating the moral domain as a

²³ Cf. McNaughton (1988: 10) who motivates non-cognitivism as an at least initially plausible alternative to cognitivism by emphasizing its ability to vindicate the "common belief that there are no moral experts."

80 MORAL KNOWLEDGE

repository of antecedently existing facts, as opposed to something that we must construct by the autonomous exercise of our own reason.²⁴

C3.P60 In short, if moral realism were true, then it would be at least *prima facie* surprising—something that calls out for explanation—if there were no full-blooded moral experts, or if there were some significant asymmetry between moral and non-moral deference. Conversely, it is natural to think that a philosopher who rejects moral realism is in a promising position to account for the alleged non-existence of full-blooded moral experts.²⁵

C3.P61 However, despite the initial plausibility of this line of thought, we should resist it: plausibly, neither the issues in this area or their solution ultimately depend on the choice of metaethical framework.²⁶ Although at least some non-realist views would, if true, readily explain the non-existence of full-blooded moral experts, it is not the fact that they are non-realist views that allows them to do this. It is another feature of the views in question, a feature that I will refer to as their *egalitarianism*. Significantly, this is a feature that they share with versions of realism that are sometimes defended, and these egalitarian versions of realism would, if true, account for the non-existence of moral experts just as well. For this reason, I do not think that the choice of a general metaethical framework is ultimately of crucial significance here. I will briefly illustrate how this plays out in the case of the two metaethical views mentioned earlier, expressivism and constructivism.

C3.P62 Consider first views in the broadly expressivist tradition that descends from Ayer to contemporary philosophers such as Blackburn (1993) and Gibbard (1990). One possibility that we should note immediately is the following: even if Ayer's original emotivist view accommodates and explains the non-existence of full-blooded moral experts, it does not follow that this feature is shared by its more sophisticated successors. Consider again Ayer's account, on which making a moral judgment is an expression of disapproval. Perhaps you judge that eating meat is wrong because you disapprove of the practice, while I judge that it is good because I approve of it. The account of moral judgment offered in *Language, Truth, and Logic* offers no resources by which either of our stances might be privileged over the other. Given this, the

²⁴ For clear statements of the constructivist rejection of the idea that the aim of moral judgment is to depict antecedently existing moral facts, see Korsgaard (2008), Rawls (1980), and Street (2010).

²⁵ David Brink (1989) is among those who see a connection between the existence of moral experts and the credibility of moral realism. There, Brink considers the possibility that someone might reject moral realism on the grounds that it has “the implausible implication that there are moral experts or authorities”; in response, he offers a qualified defense of moral expertise, and thus attempts to turn the point to the moral realist's advantage (pp. 95–8).

²⁶ On this point I am in agreement with Enoch (2014): 234–5.

idea that I should defer to you about the morality of eating meat seems completely out of place, for the theory offers no resources by which we might make sense of the idea that you are in a better position to make a judgment than I am, or vice versa.

C3.P63 This suggests that the real reason why Ayer's emotivism can accommodate the non-existence of full-blooded moral experts is not the fact that it is a non-realist view, but the fact that it is a strongly egalitarian view. In contrast, Blackburn (1993) repeatedly emphasizes that his account can do full justice to the thought that some moral sensibilities are superior to others, so that, for example, I can recognize that my current moral sensibility is not the end of the story, and might be improved. However, once inegalitarianism about moral sensibilities is in play, the way seems clear for the return of full-blooded moral experts. For once we allow that it makes sense for me to think that my current sensibility might undergo improvements, it seems that there is in principle no obstacle to my believing that others already have the superior moral sensibility to which I aspire. In which case, it would seem that the reasonable thing for me to do would be to defer to their moral judgments.²⁷

C3.P64 A similar point holds in the case of the metaethical constructivist. Recall the way in which the constructivist seemed to be in a strong position to account for the putative absence of full-blooded moral experts. Again, according to the constructivist, the proper aim of moral inquiry is not the discovery of some set of already existing moral facts; in this respect, it is not analogous to physics on a realistic construal, which aims to discover certain facts about the nature of physical reality. Rather, moral facts are *constructions*, which emerge out of procedurally impeccable practical reasoning, or procedurally impeccable practical reasoning subject to certain constraints. But if there are no antecedently existing moral facts, then there is no possibility that some individuals might occupy a privileged position for detecting such facts. Thus, it might appear that the constructivist, unlike the moral realist, is in a good position to account for the putative non-existence of full-blooded moral experts.

C3.P65 However, this appearance is misleading. After all, it is not as though constructivists are crude relativists about morality, who think that the moral facts for each person are simply whatever conclusions she happens to arrive at when she engages in moral reasoning. Rather, constructivists think that the moral facts are constructions from the *correct* constructive procedure, when that procedure is impeccably carried out. On many constructivist accounts, it

²⁷ The point that the notion of a moral expert seems to make perfect sense within a quasi-realist framework is noted by Fricker (2006: 238).

82 MORAL KNOWLEDGE

is this idealization that underlies the objectivity of morality, and the claim that fully rational agents would tend to converge in their moral views, at least to some extent. However, once this idealization is introduced into the account, the prospect of egalitarianism among moral inquirers immediately arises. On the face of it, there does not seem to be any reason to expect that every person will be equally good at correctly carrying out the constructive procedure (however that constructive procedure is characterized); indeed, it would be extremely surprising if this turned out to be so. Suppose that I come to believe that another person is better than I am at the relevant kind of reasoning. Given this belief, it would seem that the reasonable course for me is to defer to her moral judgments in any case in which I find myself uncertain, or in any case which I am disposed to disagree—that is, to treat her as a full-blooded moral expert. And this would seem to the case even if what she is better at figuring out are not antecedently existing moral facts, but rather the conclusions that both of us would reach if we engaged in the constructive procedure impeccably.

C3.P66 Of course, a constructivist might proceed to tell a story about the reasoning procedure according to which any normal adult human being is equally well-equipped to engage in such reasoning. Interestingly, Kant explicitly held that in the practical sphere, unlike the theoretical sphere, “ordinary understanding” (*der gemeine Verstand*) is at no disadvantage compared to the kind of reasoning of which the philosopher is capable.²⁸ There are, then, some constructivist views which would seem to account for the putative absence of full-blooded moral experts, just as there are some expressivist views which would do the same. However, by the same token, a moral realist might consistently tell a story on which the objective, mind independent moral facts are equally accessible to any normal adult human being, a story on which moral deference would be otiose, and which would leave no room for full-blooded moral experts. In each case, it is the egalitarian character of the view in question, and not the fact that it belongs to a certain metaethical genus, which is crucial. Given that it is this feature that is potentially explanatory, let us consider it directly.

²⁸ “And the most extraordinary thing is that ordinary understanding in this practical case may have just as good a hope of hitting the mark as that which any philosopher may promise himself. Indeed it is almost more certain in this than even a philosopher is, because he can have no principle other than what ordinary understanding has, but he may easily confuse his judgment by a multitude of foreign and irrelevant considerations and thereby cause it to swerve from the right way” (1785/1981: 404).

Commenting on this passage,

Rawls (2000: 148) offers the following gloss: “moral philosophy is not needed to teach us our duties and obligations—to tell us what they are—for these we already know.”

C3.S6

3.3.3 Egalitarianism

C3.P67 Physicists and biologists know truths of their respective sciences that the non-scientist is simply not in a position to know. For this reason, it makes sense for the layperson to blindly defer to them when it comes to their areas of specialization. If there are no full-blooded moral experts, then there is no individual who stands to morality as the physicists and biologists stand to their respective sciences. One straightforward way in which this could turn out to be true is the following: perhaps when it comes to the truths of morality, no one is in a privileged position compared to anyone else, in the way that some are in a privileged position compared to others with respect to the truths of physics and biology. (Or at least: when it comes to the truths of morality, no *normal adult human being* is in a privileged position compared to any other normal adult human being.) That is, perhaps the kind of clear informational asymmetries and differences in strength of epistemic position that make it reasonable for some to defer to others with respect to many scientific questions simply do not exist in the moral domain.

C3.P68 Here is one possible way in which this general line of thought might be developed. A venerable and still popular idea is that at least the core truths of morality are knowable *a priori*, i.e. from the armchair, or without the benefit of any specific experiences or empirical observations. An obvious suggestion, then, is that the *a priori* knowability of morality leaves no room for full-blooded moral experts, because it guarantees that every normal adult human being is already in a position to know the core moral truths, regardless of his or her experience of the world.

C3.P69 However, an appeal to the putative *a priori* knowability of morality does not seem like the most promising way to develop the general egalitarian thought. Even if a given domain is *a priori*, and so no one is precluded from grasping that truths of that domain in virtue of not having had specific courses of experiences, it hardly follows that there are no significant disparities in how well positioned people are to know the relevant truths. And where such disparities exist, it will seem perfectly natural for some to treat others as full-blooded moral experts. Consider, for example, the discipline of mathematics, which I assume is *a priori* if any discipline is. Surely the vast majority of those who believe Fermat's Last Theorem do so not because they understand its proof, but on the basis of deference to others. Believing Fermat's Last Theorem as a result of deferring to the mathematician seems on a par with believing the theory of special relativity as a result of deferring to the physicist. Of course, there are many mathematical propositions that are both knowable

84 MORAL KNOWLEDGE

a priori and about which it would be bizarre to defer to another person, at least in any ordinary circumstance. (Consider, for example, elementary truths of arithmetic.) But in such cases, it is the *obviousness* of the claims, as opposed to their a priori knowability, which seems to explain why deference would be out of place. After all, it would seem just as strange for one of us to defer to the other about paradigmatically a posteriori truths that are sufficiently obvious (consider, for example, facts about our environment that are immediately available to both of us on the basis of perception) as about comparably obvious a priori claims. Thus, we have good reason to think that the allegedly a priori character of morality is not what is crucial here.

C3.P70 Could the feature that relevantly distinguishes morality from physics and biology be its obviousness or potential obviousness? In fact, a distinguished tradition in moral philosophy insists that everyone (or least, every non-defective adult human being) is in a position to know the essentials truths of morality, independently of the assistance of others.²⁹ Here, for example, is Thomas Reid:

C3.P71 From self-evident first principles the whole system of moral conduct follows so easily, and with so little aid of reasoning, that every man of common understanding, who wishes to know his duty, may know it. The path of duty is a plain path, which the upright in heart can rarely mistake. Such it must be, since every man is bound to walk in it. There are some intricate cases in morals which admit of disputation; but these seldom occur in practice; and when they do, the learned disputant has no great advantage...in order to know what is right and wrong in human conduct, we need only listen to the dictates of our conscience, when the mind is calm and unruffled, or attend to the judgment that we form of others in like circumstances.

C3.P72

(1788 [2003]: 645)

C3.P73 Similar ideas have been endorsed by a number of contemporary philosophers. For example, Bernard Gert (2004: 90) claims that “morality does not require beliefs that are not known to all moral agents” (2004: 90).³⁰ He opens his *Common Morality* with the following declaration:

²⁹ A main theme of Schneewind (1988) is the extent to which the development of modern moral philosophy was driven by the desire for an account of morality that would make an awareness of its requirements universally available.

³⁰ Cf. p. 128 “Because morality must be understood by all of those to whom it applies, the content of the moral system cannot be determined by beliefs that are not shared by all other rational persons.”

C3.P74 This book...contains no new information about what kinds of actions morality prohibits, requires, discourages, encourages, or allows. Anyone who is intelligent enough to read this book already has all of this information.

C3.P75 (p. 3)

C3.P76 Rescher (2005: 200) explicitly cites the putative triviality of morality as that which explains the nonexistence of moral expertise:

C3.P77 Anything that requires extensive knowledge or deep cogitation is *ipso facto* ruled out as a moral precept or principle. The very idea of moral expertise...is

C3.P78 for this reason totally unrealistic. (2005: 200)

C3.P79 In a similar vein, Smith (1994: 5) writes, “our moral life seems to presuppose that [moral] facts are available to all, that no one in particular is better placed to discover them than anyone else.” Others have suggested that there are innate “moral universals” akin to the linguistic universals posited by contemporary cognitive science, and that, more generally, there are deep similarities between morality and linguistics.³¹ If so, then perhaps an ordinary agent has an underlying competence for making moral judgments about particular cases, just as any ordinary speaker has an underlying competence for autonomously judging the grammaticality of sentences in his or her natural language.

C3.P80 Could the alleged universal availability of morality account for the non-existence of full-blooded moral experts? As I see it, this general approach is undermined by a dilemma. Consider two different interpretations of the thought that the truths of morality are universally available. On the first interpretation, the truths of morality are, like the simplest truths of arithmetic, literally trivial: they are obvious or potentially obvious to anyone who is genuinely concerned to know them. (Recall again Reid’s view, according to which “the whole system of moral conduct follows easily and with little reasoning” from self-evident first principles, so that anyone who sincerely desires to know what is morally required will know what is morally required, and that moral questions that “admit of disputation” seldom arise in practice.) If the truths of morality were universally available in anything like this sense, then this would effectively guarantee the non-existence of moral expertise. Similarly, if Gert were correct in holding that that the only beliefs that morality requires are already known by all moral agents, then deference about moral matters would

³¹ See especially Mikhail (2000, 2007). For discussion and further references, see Roedder and Harman (2010).

86 MORAL KNOWLEDGE

surely be otiose. However, such a characterization of our cognitive relationship to morality seems implausibly optimistic. Notably, the diversity of opinion that we find with respect to many moral issues belies the idea that the truths of morality are obvious or even potentially obvious; *pace* Reid, genuine moral disagreement is neither an uncommon nor a marginal phenomenon.³² Moreover, there seems to be no good reason to deny that, among those who disagree about issues such as the morality of abortion or capital punishment, at least some on each side of the debate are people of good will who sincerely desire to know what morality requires. Even apart from interpersonal considerations, genuine moral uncertainty seems far more common than would be the case on the suggested picture.

C3.P81 Consider then a second interpretation of the thought that the truths of morality are universally available. On this interpretation, even though some moral truths are unobvious, those truths are nonetheless *in principle* accessible to all normal adult human beings.³³ Perhaps, for example, all normal adult human beings have a level of conceptual sophistication and general intelligence that is more than sufficient to grasp the truths of morality, provided that that intelligence and sophistication is exercised well. Perhaps it is even the case that the great bulk of moral disagreement is due to certain common biases, false non-moral beliefs, and simple performance errors that distort our moral reasoning, but that we would converge on the same moral views given suitably idealized conditions of reflection and deliberation. In short, perhaps the sense in which moral truths are universally accessible is just this: none of us lacks the *wherewithal* to autonomously grasp those truths, in a way that not all of us have the wherewithal to autonomously grasp the truths of quantum electrodynamics or algebraic topology.

C3.P82 This view about morality is at least somewhat more plausible than the view according to which morality consists of a body of trivialities. However, notice that even if it were true, it would not provide a good explanation of the putative non-existence of moral experts. For it often makes perfect sense to treat someone as an expert with respect to a range of questions, even when one could *in principle* figure out the answers to those questions on one's own. For example, I treat my accountant as an expert with respect to questions about my taxes, even though I believe that I am in principle capable of figuring out the answers to the questions about which I defer, given sufficient time and

³² Schroeder (2007: 175) also emphasizes the potential tension between the idea that there are no moral experts and the extent and depth of moral disagreement that we actually find.

³³ Of course, there will be a number of different views here, of varying degrees of plausibility, corresponding to different senses of “*in principle*.”

attention to the task. However, in contrast to the possibility of outsourcing my moral views to another person, this seems like a paradigmatic case in which outsourcing my beliefs to another person is perfectly natural and utterly unproblematic.

C3.P83 Thus, the dilemma for such “universalist” approaches to accounting for the perceived asymmetry between morality and other subject matters is the following. On the one hand, one might hold a view according to which the truths about morality are “universally available” in the sense that they are obvious or potentially obvious to everyone. A view of this sort would account for the non-existence of moral experts if it were true, but it seems implausibly optimistic concerning our cognitive relationship to morality. On the other hand, the more a view allows for the unobviousness of moral facts by idealizing the sense in which those facts are universally available (e.g. “available to all normal adult human beings given sufficiently idealized reflection and deliberation”) the more plausible it becomes, but the less it is able to explain why there would be any asymmetry between the moral and the non-moral cases.³⁴

C3.P84 Although in formulating this dilemma I have employed the characteristic idioms of the realist, I believe that the same dilemma arises for any non-realist attempt to appeal to the alleged universal availability of morality in this context. Let me briefly illustrate the point with respect to constructivism. As noted above, Kant held that “ordinary understanding” (*der gemeine Verstand*) is perfectly sufficient for the kind of reasoning required for morality, and that therefore, the relevant kind of reasoning is something that anyone of ordinary understanding is in a position to carry out for herself. However, any plausible constructivist view will be consistent with the thought that one of the ways in which one’s moral reasoning might be corrupted is through the influence of self-interest or bias. But in practice, not all human beings will be equally prone to succumbing to such influences. Therefore, we should expect that, as a matter of contingent psychological fact, not all human beings will be equally

³⁴ In this connection, it is interesting to note changing conceptions of the epistemic role of *conscience*. For the scholastics Aquinas and Suarez, the primary role of the conscience is to deliver knowledge of first principles (e.g. “Good is to be sought and evil avoided”), at least some grasp of which is possessed by everyone. However, not everyone has the same ability to become fully aware of what these most basic principles entail in particular cases. Because of this, Aquinas holds that “the truth is the same for all, but it is not equally known to all” (*Summa Theologiae* Ia.IIae.94.4,93.5). (According to Aquinas, at least part of the difficulty in reasoning correctly about particular cases is due to original sin (Ia.IIae.92; cf.94.6); compare modern secular views according to which it is the influences of bias and ideology which distort one’s moral reasoning about particular cases.) In the early modern period, a significant number of philosophers reinterpreted conscience as the ability to grasp immediately what morality requires in particular circumstances. (On the point, see Schneewind 2003: 27). This leads to views such as Reid’s, according to which in order to know what to do we need only “listen to the dictates of our conscience, when the mind is calm and unruffled.”

88 MORAL KNOWLEDGE

good at the relevant kind of reasoning, however exactly it is characterized. But if I know that someone else is better than I am at the relevant kind of reasoning, then it seems that as though I would have a compelling reason to defer to her moral conclusions when she and I arrive at different answers: that is, to treat her as a full-blooded moral expert. And this is so even if I am *in principle* just as good as she is at the relevant kind of reasoning.³⁵ On the other hand, a constructivist view according to which every normal adult human being is not just *in principle* equally able to carry out the constructive procedure but is *in practice* equally good at carrying out the procedure does not seem plausible.

C3.P85 In this section, I've criticized a number of attempts to account for the (alleged) non-existence of full-blooded moral expertise. One possible response to the inadequacy of these explanations is that there is actually nothing to be explained in this vicinity. Perhaps the deferential spouse's practice of outsourcing his moral convictions to his wife is on a par with his practice of deferring to her sense of direction; and perhaps there is no philosophically significant difference between the situation of Williams' imagined student (who understands her professor's moral conclusion but not the reasoning behind it) and the situation she would be in if she found herself in a structurally similar position at the end of a physics lecture. However, that response would be premature. There *are* philosophically significant differences between the moral and non-moral cases. The next two sections are devoted to exploring some of these differences.

C3.S7

3.4 Opacity, Justification, and Understanding

C3.P86 A full-blooded expert about X is not merely someone who knows more about X than the non-expert, but someone to whom it is rational for the non-expert to defer in arriving at her views about X. Even in a case in which you would in fact arrive at more accurate views about X by outsourcing your opinions to another person, it does not follow that it would be rational to arrive at your views in this way. First, you might not be in a position to know that you would

³⁵ Compare: even if you and I are both perfectly competent when it comes to adding together three digit numbers, I might know that you are in practice significantly more reliable than I am, in virtue of committing significantly fewer performance errors than I do (suppose that I'm habitually more careless). If so, then it makes sense for me to favor your answers over mine in cases in which we arrive at different answers, even though we are both perfectly competent when it comes to the relevant kind of reasoning.

end up with more accurate views in this way. I will consider this issue as it arises in the moral domain in section 3.5. Second, even if you knew that you would end up with more accurate views about X by outsourcing your views, there might be costs to proceeding in this way, costs that tend to offset the expected gains in accuracy.

C3.P87 What, if anything, would be lost by outsourcing your moral views to another person if you knew that by doing so you would end up with views that are at least as accurate as those which you would otherwise hold? A satisfying answer to this question should identify some feature of morality (or our relationship to morality) that distinguishes it from those subject matters where deference seems comparatively natural and unproblematic. In this section, I offer two complementary proposals, each of which appeals to a deep feature of distinctively moral judgments. The first proposal appeals to the interpersonal aspects of moral judgment and the norms that govern moral discourse; the second appeals to the intimate link between moral judgment and action.

C3.P88 As a first step towards developing both proposals, let us borrow a useful distinction from David Enoch (2014), between *transparent* and *opaque* evidence.³⁶ Intuitively, transparent evidence for a proposition not only suggests that the proposition is true, but also provides insight into *why* it is true, or why it would be true. In contrast, opaque evidence for a proposition suggests that the proposition is true without providing any indication as to why it is or would be true. Consider, for example, two different kinds of evidence you might have that some mathematical proposition is true. First, while taking a mathematics exam, you might encounter a question that asks you to prove the proposition. Since you know that it is unlikely that you would be asked to prove a proposition that is not true, the fact that the exam asks you to prove this particular proposition is evidence that it is true. This evidence is opaque evidence, since it provides no insight into why the mathematical proposition is true. Suppose that you subsequently succeed in proving the proposition in some canonical way. The proof itself is excellent (indeed, conclusive) evidence that the proposition is true; in addition, it might very well provide insight into why the proposition is true. Insofar as it does afford such insight, it is transparent evidence.³⁷

³⁶ Although both the terminology and the general idea behind the distinction are due to Enoch, the characterization that I offer here differs from his, in subtle but significant ways.

³⁷ I do not assume that every genuine mathematical proof affords insight or understanding into why the proposition of which it is a proof is true, only that this is a feature of at least some mathematical proofs.

90 MORAL KNOWLEDGE

C3.P89 Here are two observations about the distinction between opaque and transparent evidence.

- C3.P90 (1) Whether evidence is opaque or transparent does not in general correlate with its probative force. In particular, despite the fact that opaque evidence for p does not provide insight into why p is or would be true, it might nevertheless be highly probative and license a high degree of certainty that p is true. For example, given that you know that you would not be asked to prove a false proposition on the exam, the fact that you have been asked to prove a certain proposition licenses a high degree of certainty that it is true. A corollary of the fact that opaque evidence can have a great deal of probative force is that it can underwrite knowledge. One might know on the basis of evidence that *drinking saltwater causes dehydration* even if one's evidence does not provide any understanding of why saltwater causes dehydration. On the other hand, *transparent* evidence for p might be relatively *weak* evidence that p is true, even though it is potentially explanatory of p .³⁸
- C3.P91 (2) *The fact that another person judges that p is true is (at best) opaque evidence that p is true.* Suppose that I learn that you believe that drinking saltwater causes dehydration. Even if this amounts to an excellent reason for me to believe the same proposition—indeed, even if puts me in a position to *know* that proposition—I might still have no idea why drinking saltwater causes dehydration. Similarly, when the deferential spouse outsources his moral convictions to his wife, he bases his moral beliefs on what is at best opaque evidence for their truth. Even in a favorable case in which the moral view at which he arrives is true, his grounds for holding that view provide no insight into why it is. More generally, when you hold a moral view because you defer to a person whom you regard as a full-blooded moral expert, you typically lack understanding of why that view is true.³⁹

³⁸ An example: suppose that you know that Smith (who is on his way to your party but has yet to arrive) has a poor sense of direction. That *Smith has a poor sense of direction* is both evidence that *Smith is lost* (it makes sense for you to be more confident that Smith is lost given your knowledge that he has a poor sense of direction) and also potentially explanatory of his being lost. Nevertheless, the fact that Smith has a poor sense of direction might be relatively weak evidence that Smith is lost. (Imagine that the party has started only recently and that Smith is one of several guests who has yet to arrive.)

Although he does not address the point explicitly, Enoch's (2014) various formulations of the distinction imply that transparent evidence for p entails that p is true, i.e. that transparent evidence for p is always *conclusive* evidence for p . But I believe that the intuitive distinction between transparent and opaque evidence crosscuts the distinction between conclusive and non-conclusive evidence.

³⁹ The fact that moral testimony does not typically deliver moral understanding is noted by Nickel (2001), Hopkins (2007), and emphasized by Hills (2009).

C3.P92 Even if we assume that it is better to understand why a moral belief that you hold is true, as opposed to merely having evidence that it is, this does not constitute an asymmetry between the moral and the non-moral cases. After all, it is equally true that when you treat a physicist as a full-blooded expert, your evidence for your scientific beliefs is similarly opaque and fails to afford insight or understanding as to why they are true. And plausibly, it would be even better if you had genuine understanding of why the relevant scientific claims are true, as opposed to mere evidence that they are.

C3.P93 One respect in which holding a moral belief on the basis of opaque evidence is suboptimal that does *not* carry over to the non-moral case is this: when you hold a moral belief on the basis of opaque evidence, you are not in a position to offer the kind of interpersonal justification that you might be expected to offer, when your moral belief manifests itself in judgment in characteristic ways. Suppose I tell you that a practice you habitually engage in is wrong, and that you should refrain from engaging in it in the future. Assuming that this is not something you already believed, it seems both fair and natural for you to respond by asking me to provide some reason for thinking that the practice is wrong. Here, what you expect me to provide is some fact about the practice in virtue of which it counts as wrong. More generally, when one criticizes the behavior of others as falling short of some standard, one is expected to be able to explain or give some indication of the way in which the behavior falls short of the relevant standard. When one's belief that the practice is wrong is based on opaque evidence, this burden cannot be discharged.

C3.P94 In section 3.1, I noted the distinction between a token belief's *being justified* (i.e. its having a certain epistemic status) and the *activity of justifying* it. I noted there that there are principled reasons for thinking that a belief can be justified (in any sense in which it is plausible to think that being justified is a necessary condition for knowledge) even if the person who holds the belief is not in a position to successfully engage in the activity of justifying it. Nevertheless, even if this is not a necessary condition for knowing, there are certainly contexts in which being in a position to successfully engage in the activity is important. A natural thought is that an ability to successfully engage in the activity is particularly important in interpersonal contexts, in which you are expected to justify a view that you hold to another person. A paradigmatic context in which this expectation is present is a case in which you have criticized the behavior of another person as wrong, or as falling short of some relevant standard.

C3.P95 Indeed, we can draw a further distinction here, between two distinct but related activities that might be described as “justifying a belief.” First, there is

92 MORAL KNOWLEDGE

the activity the aim of which is to show that your token belief has a certain epistemic status—for example, that it is a reasonable thing to think, given your evidence. Second, there is the activity that aims at providing interpersonally compelling reasons for thinking that *what you believe* (i.e. the proposition that is the content of your token belief) is true. You can successfully perform the first task without successfully performing the second. For example, in order to show that a token belief that I hold is a reasonable thing for me to think, it is enough for me to show that it is a reasonable thing for me to think given my evidence (perhaps it was told to me by someone whom I have good reason to trust). Even if I successfully demonstrate to my audience that my belief meets this standard, it does not follow that I have given them compelling reason to think that what I believe is true, for their evidence might differ from mine in significant ways. (For example, perhaps members of my audience do not have reason to trust, or even have reason to distrust, my source.) With this distinction in mind, consider again a context in which I have criticized a certain behavior of yours as wrong, and you have called on me to justify this judgment. In order to successfully meet your demand, it is not enough for me to show that my moral belief is a reasonable thing for me to think given my evidence. Rather, what is expected is that I offer considerations that bear on *why* the practice is wrong. When my basis for believing that the practice is wrong is a matter of my having opaque evidence, I will not be in a position to offer such considerations.

C3.P96 A second way in which holding a moral belief on the basis of opaque evidence is suboptimal that does not carry over to the non-moral case concerns the intimate connection between moral judgment and action. In particular, when you hold a moral view on the basis of opaque evidence, you are not in a position to fulfill an important ideal associated with moral agency: the ideal of doing the right thing for the reasons that make it right.⁴⁰

C3.P97 In order to appreciate this point, we should distinguish the following three ideals associated with moral agency:

C3.P98 (1) The ideal of doing the right thing.

C3.P99 (2) The ideal of doing the right thing because it is the right thing to do.

C3.P100 (3) The ideal of doing the right thing for the reasons that make it the right thing to do.

⁴⁰ The point that an agent who relies on testimony for a moral belief is not in a position to do the right thing for the right reasons is noted independently by Hills (2009) and McGrath (2010a). Here I'll defend this observation and its significance against an important recent line of criticism.

C3.P101 It is a familiar point that an agent can do the right thing even if he does not do the right thing because it is right, or for the reasons which make it the right thing to do. Kant's ([1785] 1981) ultimately self-interested shopkeeper, who consistently gives his customers the correct change because doing so is the most efficient way to maintain an advantageous reputation for honesty, is an example of an agent who fulfills the first ideal without fulfilling either the second or the third.

C3.P102 Consider then the two more demanding ideals, (2) and (3). Suppose that on a given occasion the deferential spouse believes that Φ -ing is the right thing to do in his circumstances because he has been informed that this is so by his wife, who knows that Φ -ing is the right thing to do. Here, the deferential spouse is in a position to do the right thing because it is right: that is, his concern for morality might motivate him to Φ . However, even on the assumption that the deferential spouse genuinely knows that Φ -ing is the right thing to do (i.e. that his wife's knowledgeable testimony is enough to underwrite his knowing), he is *not* in a position to do the right thing for the reasons that make it right. For his knowledge that it is the right thing to do is not grounded in an appreciation of or sensitivity to the features that make Φ -ing the right thing to do, but rather on the opaque evidence provided by his wife's testimony. Thus, even in the best case of a moral view held on the basis of deference—one in which it is natural to credit the agent with genuine knowledge of the action's rightness—the agent is still not in a position to fulfill the ideal of doing the right thing for the reasons that make it right, for fulfilling that ideal requires a kind of insight that the deferential spouse lacks. Because moral deference does not deliver understanding, moral deference does not put one in a position to do the right thing for the right reasons even if it delivers moral knowledge.⁴¹ By contrast, when a person defers to another about the weather, or about geography, or about some esoteric scientific matter, her doing so does not similarly frustrate the achievement of any ideal associated with agency.⁴²

C3.P103 Consider an objection to this line of thought. I have assumed that there is a robust distinction between (i) the reasons that make an action the right thing

⁴¹ A parallel point holds with respect to the ideal of refraining from wrongdoing for the reasons that make an action or behavior wrong: if a child refrains from lying on a given occasion because he knows via testimony that lying is wrong but has no grasp on why it is, then his refraining from lying on that occasion is not based on the reasons that there are not to lie.

⁴² The ideal of doing the right thing for the reasons that make it right, and the idea that fulfilling this ideal requires genuine understanding on the part of the agent, ultimately descend from Aristotle. See especially his classic discussion of what is required in order for a "just action to be done justly" *Nicomachean Ethics* II.4.

94 MORAL KNOWLEDGE

for an agent to do, and (ii) the reasons that justify the agent in believing that the action is the right thing for her to do. Some contemporary accounts of right-making reasons collapse this putative distinction.⁴³ On views of this kind, whenever an agent has compelling reasons to believe that Φ -ing is the right thing for her to do, those reasons also *make it the case* that Φ -ing is the right thing for her to do in the circumstances. There is thus no genuine possibility of the kind to which I've appealed, in which an agent has access to compelling reasons to believe that she ought to Φ but does not have access to the reasons that make her Φ -ing right.

C3.P104 Markovits (2012) applies a view of this kind to the issue of moral deference and argues that it undermines the kind of agency-based account on offer here.⁴⁴ She writes:

C3.P105 I *don't* agree that deference prevents an agent from acting on right-making reasons.

C3.P106 This is because the moral reasons for us to perform some action are *subjective*...if expert testimony gives us most reason to believe some act would be best then that testimony *is the reason* we ought to perform that act...in those cases in which the advice of a (recognizable) moral expert provides our agent with sufficient evidence for the belief that a particular act would be best, our agent's acting so may be *made* right by the fact that the moral expert advises him to do it...And his act is right *because* it follows the expert's advice: this is the reason *why* he ought to perform it...So the deferential agent may, it seems to me, act for right-making reasons after all.

C3.P107 (pp. 306–7)

C3.P108 However, we should reject the general view about right-making reasons on which these conclusions depend, according to which an agent's reasons to believe that an action is right are also the reasons that tend to make that action the right thing for her to do. Even setting aside contested examples involving moral testimony or deference, we can describe cases in which an agent has compelling reasons to believe that she should Φ , while being ignorant of the reasons that make it the case she should Φ . The clearest cases of this kind will be ones in which the facts that provide an agent with compelling reasons to believe that a given action is the right thing *for her to do are also* reasons for her *not* to perform the action. Imagine a possible world run by a

⁴³ See especially Kearns and Star (2009) and Markovits (2010).

⁴⁴ She is responding to earlier statements of the idea in Hills (2009) and McGrath (2010a).

demon who has comprehensive knowledge of right and wrong, but who desires to maximize the amount of wrong-doing, and minimize the amount of right-doing, that the world's inhabitants freely choose to engage in. As a means towards these ends, the demon arranges things so that whenever a person contemplates or is in the process of performing an act that is in fact morally right, the person experiences certain painful sensations; and whenever a person contemplates or is in the process of performing an act that is morally wrong, the person experiences certain pleasurable sensations. Suppose, however, that at some point I catch on to these correlations. In these circumstances, if I'm initially uncertain whether some action would be right or wrong, but then find myself experiencing pain upon beginning to perform it, this would be a compelling reason to believe that the action is right. However, the fact that performing the action would involve a certain kind of pain is *not* a plausible candidate for *what makes it the case* that the action is the right thing for me to do in the circumstances. On the contrary, the fact that the action would be painful is a reason *not* to perform it. A parallel point holds for wrong-making features. Even if the evidence that alerts me to the fact that what I'm doing is wrong is a certain characteristic feeling of pleasure or enjoyment, the fact that I'm experiencing that pleasure or enjoyment is not a plausible candidate for being the feature that *makes* the action wrong. In general, evidence or reasons to believe that a certain action is right (or wrong) are not the same as reasons that make that action right (or wrong).

C3.P109 We have identified two significant respects in which holding a moral belief on the basis of opaque evidence is suboptimal that do not carry over to the non-moral case. First, when one holds a moral belief on the basis of opaque evidence, one is not in a position to provide the kind of interpersonal justification that one is expected to be able to provide in contexts in which the moral belief manifests itself in characteristic ways. Second, when one holds a moral belief on the basis of opaque evidence, one is not in a position to fulfill an important ideal associated with moral agency, that of doing the right thing for the reasons that make it right. Taken together, these two considerations might go a substantial way toward vindicating the sense that relying on another person for one's moral beliefs (even another person whom one has good reason to think reliable) is not on all fours with relying on another person for one's beliefs about many other subject matters.⁴⁵ Notice, however, that

⁴⁵ Of course, these considerations do not exclude the possibility that there are other significant respects in which moral deference is not on a par with non-moral deference. For example, Howell (2014) develops the suggestion that the primary cost of moral deference is "the crippling effect such deference can have on the moral character of the deferring agents" (p. 412).

96 MORAL KNOWLEDGE

neither of these considerations gives us any reason to think that moral beliefs held on the basis of deference cannot amount to full-fledged moral knowledge.

Indeed, because both of these considerations depend on the opacity of the evidence, we should expect that such beliefs can amount to knowledge, since it is uncontroversial that, in the general (non-moral) case, there is no obstacle to knowing propositions on the basis of opaque evidence. As noted above, whether evidence is opaque or transparent does not generally correlate with its probative force, or its ability to ground knowledge; opaque evidence can be extremely strong evidence, notwithstanding the fact that it does not deliver understanding.

C3.P110 Similarly, from the fact that deference leaves one unable to fulfill certain ideals associated with moral agency or moral discourse, it does not follow that one ought to eschew moral deference in cases in which one has good reason to believe that one is more likely to act correctly by deferring.⁴⁶

C3.P111 An important theme in what has been said thus far is that *understanding why p is true* is often more intellectually demanding than *knowing that p is true*. I take this point to be uncontroversial outside of the moral domain. (Again, one can know, and even know with certainty, that drinking saltwater causes dehydration without understanding the process by which drinking saltwater causes dehydration.) In general, it is not true that one's reasons for believing p must be among the reasons why p is true in order for one's belief to count as knowledge. That is why testimony can deliver knowledge of empirical matters even in the absence of any explanation as to why things are as the testifier describes them as being. On the view on offer here, what holds true generally holds also for the specific case of morality: one might know that some moral claim is true even if one is not in a position to understand why it is, and genuine moral understanding is often a more demanding intellectual achievement than moral knowledge.

C3.P112 Let us bring this last claim together with the discussion of the method of reflective equilibrium offered in Chapter 2. In section 2.2, I argued that placing the method of reflective equilibrium at the center of moral epistemology risks overintellectualizing our cognitive relationship to morality. Given the descriptions of the method offered by its proponents, it is implausible to think

⁴⁶ That there can be specific circumstances in which one is not only morally permitted but required to defer is the main claim of Enoch (2014), who argues that deferring to an another person is sometimes the only morally acceptable response to moral uncertainty. As Enoch notes (p. 234), this might be true even if moral beliefs acquired via deference inevitably fall short of knowledge. The case for the claim is presumably even stronger if, as I believe, moral deference can deliver full-fledged moral knowledge.

that acquiring moral knowledge requires engaging in reflective equilibrium reasoning, for the amount of moral knowledge that can plausibly be credited to ordinary people is out of proportion to the frequency with which they engage in such reasoning, or the role that doing so plays in their lives. *A fortiori*, employing the method is not required in order to have moral beliefs that are justified, in any sense of “justified” in which being justified is a necessary condition for knowing. However, the fact that understanding is generally a more intellectually demanding state than either knowing or justified believing makes salient the following possibility: the method of reflective equilibrium is more promising when taken as an account of moral understanding (both what such understanding would involve and how we should go about pursuing it) than as an account of either how we acquire moral knowledge or the conditions under which we are justified in holding moral beliefs.

C3.P113 Indeed, there are aspects of the method and the way that it is often presented that strongly suggest this interpretation. For example, when proponents emphasize the importance of seeking general principles and lower-level judgments that fit well with one another, what we are enjoined to seek are not merely principles and judgments that are consistent with one another, or even principles and judgments such that the principles *entail* the judgments.⁴⁷ Rather, what we are supposed to seek is a set of plausible principles and judgments such that principles would explain or “account for” the judgments (Rawls 1999; Scanlon 2014: 77). Thus, what seems to be called for is a kind of *explanatory coherence* among one’s moral views; and what (correct) explanations deliver is understanding. We might, then, think of the method of reflective equilibrium as an account of how we should pursue moral understanding, given one model of what such understanding would consist in: viz., the successful subsumption of lower level judgments under true principles that account for their truth.⁴⁸

C3.P114 Of course, just as there is a risk of overintellectualizing moral knowledge, so too there is the risk of overintellectualizing moral understanding. Insofar as ordinary people know that *it would be wrong to encourage children to smoke*, it is also plausible that they have some understanding of why it would

⁴⁷ After all, notice that if a moral judgment about a hypothetical case is necessarily true if true at all (as many will suppose), then any true moral judgment about a hypothetical case will be entailed by any general principle whatsoever.

⁴⁸ Clearly, there are other, rival models of moral understanding. It is obvious, for example, that a moral particularist would regard any account of moral understanding according to which such understanding consists in the subsumption of lower level judgments under general moral principles as a complete non-starter. I will not explore alternative models here. For more on moral understanding, see especially Hills (2009).

98 MORAL KNOWLEDGE

be wrong. (It is not as though the relevant moral claim is something that they know on the basis of opaque evidence.) One might worry that treating the method of reflective equilibrium as an account of how we should pursue moral understanding would encourage a hyper-intellectualized picture of such understanding. This concern might seem especially pressing when we recall that many proponents of the method emphasize that the *state* of reflective equilibrium is a kind of ideal, and not a state that we can realistically hope to attain in practice.⁴⁹

C3.P115 Here, however, we can exploit the fact that understanding, unlike propositional knowledge, is something that comes in degrees, and that our understanding of a domain can be more or less comprehensive. We might then see the idealized state of reflective equilibrium as a state that would provide comprehensive understanding of the moral domain as a whole, by showing how certain moral claims account for the truth of others. Of course, in order to arrive at a genuine understanding of the moral domain, the moral views that have been brought into reflective equilibrium must presumably be true. Following Hempel (1965) and others (see esp. Lipton 1991: 59–61), we can distinguish between *actual* and *potential* explanations, where, to a first approximation, a potential explanation of some fact is something that would actually explain the fact if it were true.⁵⁰ If we think of understanding as the goal of explanation, we might then say this: one actually understands why something is the case when one knows its actual explanation, while one *potentially understands* why something is the case when one has a potential explanation for it and one regards that potential explanation as actual. (What is needed in order for one's potential understanding to amount to genuine understanding is for the potential explanation to turn out to be the actual explanation.) If we think of the method of reflective equilibrium as a method for pursuing moral understanding, we might then say this: someone who has succeeded in achieving the ideal of a reflective equilibrium among a comprehensive set of moral views has arrived at potential understanding of the moral domain as a whole. She *actually understands* the moral domain as a whole just in case the comprehensive set of moral views that she holds in reflective equilibrium is true.⁵¹

⁴⁹ On this point, see especially Rawls (1993: 97) and Scanlon (2014: 77).

⁵⁰ As Lipton notes (p. 60) this somewhat oversimplifies the relationship between actual explanations and potential explanations, in ways that need not concern us here.

⁵¹ Because the picture of moral explanation and understanding naturally suggested by the reflective equilibrium picture gives a central place to the subsumption of less general moral claims under broader moral principles, the general account of explanation that it most resembles is Hempel's (1965) classic "covering law" model of explanation, which sees subsumption under law-like generalizations as the key to explanation. This might give one pause, given the well-known problems that have led

C3.S8

3.5 The Epistemology of Moral Expertise

C3.P116 Consider again the case of the deferential spouse, who treats his wife as a full-blooded moral expert. As noted above, it is useful to compare this relatively exotic case with some less exotic ones: for example, a case in which the deferential spouse consistently defers to his wife's sense of direction, because he is convinced both that he has a poor sense of direction and that his wife's is significantly more reliable. In thinking about the latter case, it is natural to picture him as having a certain kind of evidence for his convictions—for example, a past history of frequently having gone wrong when he relied on his own sense of direction, and significantly better results on those occasions when he relied upon his wife's. In general, having an unreliable sense of direction is not a significant obstacle to knowing that one does: people who have a weak sense of direction often have access to relatively clear and compelling evidence that they do, and thus are in a position to know this about themselves. (And of course, this is so even if considerations of ego or other distorting factors prevent them from responding to that evidence appropriately.) In this respect, having a poor sense of direction is similar to being near-sighted or having a bad memory. One can easily imagine having the kind of compelling evidence that would make it perfectly reasonable, indeed rationally mandatory, to defer to another's sense of direction (or eyesight, or memory), even when one would judge differently if one relied on one's own sense of direction, eyesight, or memory. Indeed, I take it that many of us actually have evidence of the relevant kind.

C3.P117 Contrast this to the case in which the deferential spouse says, "It certainly seems to me that ϕ -ing is morally permissible, but I'm sure that it isn't; after all, I have relatively unreliable moral judgment—my wife's is much better." Even if what the deferential spouse says is true, it's natural to wonder what evidence he has for such an assessment. What grounds might he have, for treating his wife as full-blooded moral expert? In general, it is much easier to acquire clear and unequivocal evidence that someone else has a superior sense of direction (or superior memory, or eyesight) than it is to acquire clear

philosophers of science to reject Hempel's model. However, there are reasons to think that many of the most debilitating problems for Hempel's account of explanation would not arise for a parallel account of moral explanation. In particular, many of the classic counterexamples to Hempel's account depend on its failure to track underlying *causal asymmetries* that are not reflected in true law like generalizations that the model wrongly counts as explanatory. (Although we can typically explain why an effect occurred by citing the fact that its cause occurred, we cannot explain why the cause occurred by citing the occurrence of the effect.) Insofar as moral explanations are not causal explanations, we should expect that analogues to this class of counterexamples will not arise in the moral case.

and compelling evidence that someone else has superior moral judgment, for reasons that are worth elaborating.

C3.P118 If you and I are inclined to disagree about which way we should turn in order to reach our destination—it seems to you that we should turn left, while it seems to me that we should turn right—then, assuming that we ultimately do one or the other, we will know soon enough who was correct on this particular occasion. Of course, even if you were correct on this occasion, it does not follow that you have a better sense of direction. But it is a piece of evidence that bears directly on that question, and crucially, it is a piece of evidence that each of us is in a position to appreciate as such. Indeed, it's natural to think that such objective "track record" evidence is in practice among the very best kinds of evidence that one could have of someone's superior reliability in a given domain. As we accumulate more such evidence, we are in an increasingly good position to know whose sense of direction is more accurate.

C3.P119 Similarly, if a genuine question arose as to whose long distance vision was more accurate, it would be easy enough to determine this: we could, for example, record our perceptual judgments about various objects at a distance from us and then check these judgments by viewing the same objects up close. Crucially, we can check our long-distance perceptual judgments for accuracy in a way that does not require us to rely on our long-distance vision, for we have independent access to facts in the target domain. Because of this, we have a way of *calibrating* our long distance vision for accuracy.⁵² By contrast, there does not seem to be a parallel way to calibrate the accuracy or reliability of one's moral judgment. If one attempted to rank others with respect to the reliability of their moral judgments by checking how often they answered difficult and controversial moral questions correctly, it seems as though one would inevitably be led to engage in first-order moral reasoning and deliberation of one's own. In practice, this has the following important consequence: those to whom we would attribute an unusually accurate sense of right and wrong will tend to be people whose sense of right and wrong overlaps strongly with our own, as measured by agreement with respect to specific moral issues. On the other hand, we generally will not judge that someone has a superior sense of right and wrong, when it leads her to make judgments that conflict with those that we are disposed to make, on the basis of our own sense of right and wrong.⁵³

⁵² Compare the way in which a scientist might calibrate a new instrument for accuracy by checking the reliability of its verdicts about some range of facts to which he has independent access. For discussion of calibration in science and elsewhere, see Cummins (1998) and Franklin (1997).

⁵³ Again, it is very far from trivial that this is so. It is decidedly *not* the case, for example, that I would only attribute a superior sense of direction to someone who makes more or less the same

C3.P120 In discussions of moral testimony, philosophers sometimes downplay the significance of the kinds of epistemic concerns raised here by emphasizing just how little is required, in the general case, in order to be justified in relying on the testimony of another person. For example, Sliwa (2012) emphasizes that, in order to be justified in relying on another person's testimony in some domain, one does not need to establish that she can correctly answer controversial questions in that domain; nor does one need to have independent evidence that she is reliable about that domain (p. 191). Howell (2014) points out that something can be a legitimate source of knowledge even if it is not known to be reliable in advance (p. 395); that we do not require independent and prior knowledge of the reliability of our testimonial sources (p. 395); and that beliefs formed on the basis of moral testimony or deference can satisfy the same standards for knowledge that apply in other domains (p. 396). All of these points seem to me true, and significant. However, they concern what is required (or what is *not* required) in order to justifiably rely on another person's moral testimony in particular circumstances, as opposed to what is required (or not required) in order to justifiably conclude that another person is a full-blooded moral expert, in the sense at issue here. Again, a full-blooded moral expert would be someone to whom the non-expert has a compelling reason to defer *even in cases in which the non-expert judges differently* (or would judge differently if he made up his own mind). Even if there is a default entitlement to rely on the testimony of others, there is obviously no default entitlement to assume that another person is a full-blooded moral expert. In order to be in a position to recognize another person as a full-blooded moral expert, one would need substantial evidence that she is.⁵⁴

C3.P121 The idea that the recognition of genuine moral expertise presents formidable challenges—challenges beyond those that arise for the recognition of expertise in many other domains—is a venerable one. Notably, it is clearly

directional judgments that I am disposed to make on the basis of my own sense of direction, in difficult cases. (In fact, I would regard more or less consistent agreement with my own unaided directional judgments in difficult cases as evidence that a person who made them does *not* have an excellent sense of direction.) In my own case, parallel points hold with respect to memory and long-distance vision.

⁵⁴ It is significant here that even on liberal views about the epistemology of testimony (see, for example, Burge 1993 and Coady 1992), according to which one is entitled to trust testimony even in the absence of any positive reason for thinking that the testifier is reliable, one's entitlement is defeated as soon as one has some reason to doubt the veracity of the testimony. But in a case in which a moral judgment made by another person conflicts with one's own judgment, one *does* have at least some reason to doubt the veracity of the testimony. (Contrast the situation one would be in if one had never considered the question before, or if one were disposed to make the same judgment as the other person.)

To be clear, I do not mean to suggest that the points raised by Howell and Sliwa are inadequate to support the more limited conclusions at issue in the contexts in which those points are made.

present in a number of Plato's early dialogues, where a general concern with moral expertise—what such expertise might consist in, whether anyone actually possesses it, and if so, how it might be recognized—is a recurrent theme.⁵⁵ Nehamas (1998) summarizes the theme as follows:

- C3.P122 Experts in *arête* therefore present a very complex problem. Do such experts exist at all? And if they do, how are they to be recognized? In the case of shoemakers or doctors, we can tell whether the shoe fits or the fever has gone: we have relatively clear ways of recognizing them. But in the case of ethical experts, it is not clear that we can recognize the experts independently of the fact that we find their views and their reasons for them—their C3.P123 reasons for living as they do—convincing. (p. 81)

- C3.P124 This is a clear statement of two ideas to which we have appealed: (i) that in the moral domain, we typically lack an independent basis for attributing genuine expertise that is comparable to the independent bases that we frequently possess in many other domains; and (ii) that someone who was in a position to recognize another's genuine moral expertise would have relatively little need to avail himself of that expertise, compared to those who are *not* in a position to recognize it.

- C3.P125 You have your moral outlook; I have mine. In some places our outlooks overlap; in other places, they conflict. Are there any possible routes by which I could come to know (or at least, acquire evidence that) your moral outlook is more likely to be correct than mine in places where the two differ? I have emphasized that assessments of the accuracy of rival moral outlooks are made from the standpoint of one's own current moral outlook, and that might seem to rule out this possibility. But consider the following route by which I might acquire inductive evidence that your current moral outlook is more likely to be right than mine when the two conflict. It might be that your past selves have consistently outperformed my past selves with respect to accuracy of moral outlook, *as judged from the perspective of my current moral outlook*. Suppose, for example, that when you and I have disagreed about some moral issue in the past, more often than not I have come around to your point of view after further reflection, experience, or discussion. Perhaps the moral judgments that I once contested are now incorporated into my own, current moral outlook. If this happened often enough, then it might become reasonable for me to think that, when it comes to those moral issues about which

⁵⁵ See especially *Protagoras* 313a–314c and *Laches* 185a.

we currently disagree, you are more likely to be right than I am.⁵⁶ In that case, I will have a reason to reverse my moral judgment when it conflicts with yours: that is, to treat you as a full-blooded moral expert. Thus, even if it is true that I will inevitably rely on my current moral outlook in appraising moral outlooks, it does not follow that it would never be reasonable for me to conclude that your current moral outlook should be trusted over mine. Still, if this is the kind of route by which one might acquire good reasons to treat another as a full-blooded moral expert, it is unsurprising that it is a relatively marginal phenomenon compared to what we find in other domains.

C3.P126 Significantly, however, there is at least one familiar situation in which it is perfectly natural for a human being to treat others as full-blooded moral experts: specifically, the situation of a child, whose own moral outlook is as yet largely undeveloped, who absorbs moral views from people who are presented to her as authority figures. I have emphasized that judgments about the quality of another person's moral judgment are typically made from the perspective of one's current moral outlook, and that in practice this tends to limit the opportunities for acquiring evidence that their moral judgment is superior to one's own. However, when a young child inherits moral views from more mature members of her community, it is not as though she typically has evidence that those from whom she inherits those views are reliable moral guides. Rather, the naturalness of her adopting their views as her own is largely due to the very fact that her own current moral outlook and capacity to engage in autonomous moral reasoning are undeveloped. Although our best empirical evidence suggests that even very young children are quite far from moral blank slates,⁵⁷ they approximate that condition to a significantly greater degree than they will at any subsequent time.

C3.P127 Thus, the fact that it is, in the usual case, both natural and reasonable for an immature member of a moral community to take on moral views that have currency among the mature members of her community does not derive from any evidence that she may have that they are reliable. Instead, it is natural and reasonable for her to do this because the views in question are presented as true by people who she in practice has little choice but to trust and no discernible reason to distrust. Of course, because this is her situation, she is to a large extent at the mercy of fortune. For even if she has no *discernible* reason to distrust those whom it is natural for her to regard as moral

⁵⁶ I noted this possibility in McGrath (2009). It is the mechanism to which Enoch (2014) appeals in the driving example of that paper.

⁵⁷ For an engaging survey of the psychological evidence on this front, see Bloom (2013).

authorities, it does not follow that they are in fact trustworthy moral guides. In the unfortunate case in which she is in fact embedded in a corrupt or morally benighted environment, she is likely to arrive at many false views about morality. On the Moral Inheritance View, this is perfectly consistent with the fact that deference of the same kind is an important way that moral knowledge can be transmitted to a child with the same cognitive capacities who is embedded in a more morally enlightened environment. After all, in empirical case, the mere fact that a community might manage to pass on radically false views about what the world is like to its younger members (who, quite reasonably, believe what they are told) supplies no reason to doubt that the same mechanism might deliver genuine knowledge when the sources of information are reliable. Here again, there does not seem to be any reason to think that the moral case differs in any significant way from the non-moral case.

C3.P128 In this chapter, I've explored some of the social aspects of moral epistemology. According to the working hypothesis endorsed in Chapter 1, any source of ordinary empirical knowledge is also a source of moral knowledge. When conjoined with the obvious fact that one's social environment is a source of ordinary empirical knowledge (via testimony and other mechanisms of transmission), this entails that the social environment is similarly a source of moral knowledge. The conclusions reached in this chapter are in accord with that consequence.

C3.P129 In the previous chapter, I argued that, by the time a typical moral agent begins to engage in reasoning that recognizably approximates the method of reflective equilibrium, she already has substantial moral knowledge to draw upon. More generally, arriving at moral knowledge by pursuing coherence typically takes place against a background of already existing moral knowledge. If the Moral Inheritance View is true, then this supplies part of an answer to the question of how it is possible for a typical moral agent to come to have moral knowledge in the first place, given that not all of her moral knowledge is the result of reflective equilibrium reasoning. Of course, as emphasized in section 3.1, this could not be the *whole* story of how such knowledge is acquired. For the hypothesized acquisition of moral knowledge via social mechanisms immediately makes salient the following question: given that the agent inherits this knowledge from other people, how did they acquire the knowledge in question? Even if they too first acquired the knowledge from their social environment, this just pushes the relevant question back to an earlier stage of the process. Generally speaking, testimony and other social mechanisms by which individuals acquire knowledge are

mechanisms of knowledge *transmission* as opposed to knowledge *generation*.⁵⁸ Just as acquiring knowledge by way of pursuing coherence typically takes place against a background of already existing knowledge that is not itself arrived at in this way, so too the acquisition of knowledge by testimony is typically⁵⁹ underwritten by knowledge that at some point in the past was arrived at in some other way.

C3.P130 The sources of moral knowledge are not exhausted by the method of reflective equilibrium and the mechanisms of social transmission. In the next chapter, I consider some additional sources.

⁵⁸ But see Lackey 1999.

⁵⁹ In section 2.6, I argued that in some exceptional cases, the pursuit of coherence can lead to the acquisition of new knowledge even in the absence of any prior knowledge. Interestingly, an analogous point seems to hold (and for parallel reasons) with respect to testimony: in some cases, a person can come to know a claim on the basis of testimony even if (none of) the testifier(s) knows the claim in question. (Imagine, for example, that a person receives two pieces of independent, mutually corroborating testimonial reports to the effect that a given claim is true, and thus comes to know that the claim is true on that basis, even though the beliefs of the two testifiers fall just short of qualifying as knowledge. Cf. Coady 1992.) As in the case of knowledge through coherence, I believe that such cases are epistemologically interesting but relatively uncommon.

C4

4

Observation and Experience

C4.S1

4.1 Introduction

C4.P1 According to the working hypothesis endorsed in Chapter 1, any way by which we arrive at ordinary empirical knowledge is also a way by which we can arrive at moral knowledge. A pressure point for this claim concerns the role of observation and sense experience.¹ On the one hand, it is obvious that observation is a crucial source of information about the world. Indeed, within the broadly empiricist tradition, it has been regarded as the preeminent, and perhaps the sole, source of such knowledge. In contrast, even philosophers who are not skeptical about moral knowledge tend to agree that we do not actually observe moral properties or their instantiations in the world around us, and that our moral views are not inductively confirmed by observations. On these points, the conventional wisdom is well-summarized by Sayre-McCord (1996: 139):

C4.P2 We don't seem to see, taste, hear, smell, or touch moral properties, nor do we seem to rely on common methods of empirical investigation and confirmation to discover them... Despite the apparent dependence of moral properties on non-moral properties, the results of empirical investigation appear to be altogether irrelevant to the justification of our moral views (although not to their application). Our moral beliefs have, at best, it seems, only a tenuous connection to experience, a connection evidently established more by the moral convictions we bring to bear on that experience than by the untainted input of experience.

¹ In what follows, I will write of empirical *observation* when what is at issue is perceptual knowledge. Generally speaking, when one observes that p (e.g. that the wall is red) one knows that p via perception. In contrast, I assume that one can have an *experience*, including an experience with propositional content, without believing or judging (and therefore, without knowing) that that content is true. For example, one can have a visual experience that represents the wall as being red, even if one refrains from believing or judging that the wall is red (as when one has reason to believe that the wall is actually white but is being bathed in red light).

C4.P3 Compare Shafer-Landau (2003:112):

C4.P4 I think we must admit that ethical evidence is different in kind from the sort we find in the natural sciences. Provided we are entitled to trust our senses, scientists can rely on them to supply evidence to test a wide array of hypotheses. Ethics cannot rely on sense evidence in the same way, for any moral theory is perfectly compatible with such evidence.

C4.P5 Indeed, from a historical perspective, it is striking how little enthusiasm there has been for empiricism about morality in the Western philosophical tradition, even among those who are otherwise firmly committed to empiricist epistemologies. Classical ethical rationalists such as Clark, Butler, and Price regularly analogized moral knowledge to mathematical knowledge in their writings, with the latter understood as the paradigm of a priori knowledge.² However, it was not only the canonical rationalists who made this comparison, but also Locke in his *Essay Concerning Human Understanding*, the founding text of modern empiricism.³ In the twentieth century, the logical positivists were committed empiricists who explicitly saw themselves as heirs to the tradition of Hume, Mill, and Mach; nevertheless, they took it as obvious that normative ethics is not a matter for empirical investigation. Unlike Locke, they refused to compromise their empiricism for the sake of ethical knowledge, and instead tended to embrace non-cognitivism.⁴ In general, those who have held that morality is a possible object of full-fledged knowledge have also held, for the most part and with some noteworthy exceptions, that such knowledge is a priori.⁵

C4.P6 We should, however, distinguish the view that moral knowledge is a priori, in the sense that it *can* be acquired by a priori means (i.e. without the benefit of empirical observation), from the much stronger view that moral knowledge can be acquired by a priori means *and only* in that way. In general, to say that a given subject matter is a priori is to say that knowledge of that subject matter is potentially available from the armchair. This view can be combined with the view that truths about the subject matter can also be known in other

² On the centrality of this comparison for the rationalist tradition, see Gill (2007), especially section 2.

³ See, for example, Locke (1979/1685: 643–5). On Locke's quasi-rationalism about ethics in the *Essay*, see Ayers (1993: 187–8).

⁴ On this aspect of the positivist tradition, see especially Ayer (1959: 21–3).

⁵ Among canonical figures writing after the distinction between empiricism and rationalism assumed center stage, Mill is a plausible exception. Among contemporary philosophers, prominent exceptions include Boyd (1988) and Railton (2003).

ways.⁶ For example, even if the proposition that $2+2=4$ is knowable a priori, it does not follow that it cannot also be known empirically. Arguably, a child just learning arithmetic can confirm her belief in that $2+2=4$ by performing a series of mini-experiments with her fingers. Similarly, even if a given moral truth can be known from the armchair, it does not follow that it cannot be known on the basis of empirical observation. It is true that, as a historical matter, many traditional rationalists have maintained, against their empiricist opponents, not only that mathematical and/or moral truths are knowable a priori, but also that mere empirical observation could never deliver knowledge of those truths. (Compare the related claim that empirical observation could never deliver full-fledged modal knowledge, e.g. the knowledge that $2+2=4$ is necessarily true.) But this is an additional bit of rationalist doctrine, and it should not be confused with the more modest (though still substantive) claim that the knowledge in question can be acquired a priori. For our purposes, the upshot is: the view that moral knowledge can be acquired a priori is consistent with the consequence of the working hypothesis with which we are currently concerned, namely that experience and observation are sources of moral knowledge.

C4.P7 Even once this point is taken into account, the claim that experience and observation are sources of moral knowledge remains contentious, as witnessed by the way in which it conflicts with the views articulated by Sayre-McCord and Shafer-Landau in the passages quoted above.

C4.P8 In this chapter, I will defend the claim that sense experience and empirical observation are sources of moral knowledge. This claim is consistent with the common view that at least the core truths of morality are knowable a priori. I will *not* argue against that view—what follows is not a defense of “empiricism” about moral knowledge.⁷ Indeed, in the final section of the chapter, I will take up the question of how some moral knowledge is available from the armchair. Nevertheless, on the picture developed here, observation and experience contribute to moral knowledge in ways that go beyond those acknowledged by many rival pictures. I will argue that experience and observation can contribute to moral knowledge in any of the ways in which they contribute to our ordinary, non-moral knowledge of the world around us.

C4.P9 On reflection, it is clear that there is not just one way in which experience and observation contribute to ordinary empirical knowledge. We can distinguish at least four ways in which experience and observation contribute

⁶ A point forcefully emphasized by Kripke (1972/1980: 35), among others.

⁷ At least, it is not a defense of empiricism about moral knowledge if one means by “empiricism” (what typically seems to be meant) that empirical observation is *the way* by which we know about the subject matter under consideration (where “the way” can signify either uniqueness or primacy).

to ordinary empirical knowledge. None of these seems to be wholly reducible to any of the others, either singly or in combination:

- C4.P10 (i) **Enabling.** Sense experience might contribute to one's knowledge that x is F by enabling one to judge that x is F in the first place: either by enabling one to grasp the concept of *Fness*, or by enabling one to refer to the object x of which F is being predicated, or both.
- C4.P11 (ii) **Triggering.** Even if one is already able to entertain the content that x is F , and even if one's epistemic position with respect to that proposition is already strong enough to know it, one might nevertheless fail to know it in virtue of failing to believe it. (Perhaps the thought that x is F has simply never occurred to one, or one doesn't realize that one has compelling reason to believe that x is F .) In these circumstances, a particular experience or observation might be responsible for one's coming to know that x is F by prompting one to judge that x is F for the first time. (For example, by calling one's attention to the possibility that x is F , or by prompting a course of reflection that ultimately leads to the realization that there are compelling reasons to believe that x is F .) In such cases, although one comes to know the proposition as a result of the experience or observation, the role of the latter in bringing about the former is not that of providing further evidence for the truth of the proposition.
- C4.P12 (iii) **Empirical confirmation and disconfirmation.** Observation might put one in a position to know that x is F by providing confirming evidence for the claim that x is F , even if the content of the claim that x is F outstrips the collective content of the observations that confirm it. Alternatively, one might learn that x is F is false by making observations that disconfirm it.
- C4.P13 (iv) **Conditioning.** The fact that one is currently in a position to know that x is F might be due to the role that past experience has played in refining one's judgment, to the point that it is now sufficiently reliable to deliver knowledge. Here, the role of prior experience in underwriting one's later reliability (and hence, one's later knowledge) is not that of supplying evidence that one retains in memory and then exploits at later times. (Thus, the phenomenon is distinct from (iii).) Rather, the role played by experience is that of endowing one with a capacity for discrimination or skill in judging (perhaps by providing feedback on earlier efforts), the successful exercise of which allows one to know at later times.

C4.P14 Again, my thesis is that experience and observation can contribute to moral knowledge in any of the ways in which they contribute to empirical knowledge, including the four ways listed here. For reasons that I will explain below, the claim that *enabling* and *triggering* contribute to moral knowledge should be relatively uncontroversial among theorists who are not skeptical about the existence of moral knowledge. Accordingly, I offer a relatively brief discussion of enabling and triggering in the next section, and then devote sections 4.3 and 4.4 to more detailed discussions of possibilities (iii) and (iv). Finally, I will conclude the chapter with a discussion of “armchair” moral knowledge in section 4.5.

C4.S2

4.2 Enabling and Triggering

C4.P15 *Enabling*. It is highly plausible that experience is needed in order to acquire various moral concepts. For example, it is highly plausible that sense experience played an essential role in my acquiring the concept *murder*. And of course, if I did not have that concept, I would not now be in a position to believe that murder is wrong. Given that knowing entails believing, I would not know that murder is wrong in the absence of the requisite experience. More generally, one seemingly pervasive way in which experience contributes to moral knowledge is by enabling us to entertain the relevant contents.

C4.P16 The idea that experience contributes to moral knowledge in this way can and should be acknowledged by anyone who is not skeptical about moral knowledge in general. In particular, there is nothing here that need be denied by those who hold that the core moral truths are knowable *a priori*. Indeed, the claim that experience contributes to moral knowledge in this way is consistent with the classical rationalist claim noted earlier, namely that our *only* access to moral truths is via *a priori* means, on a philosophically and historically well-motivated interpretation of that claim. Even if my ability to entertain the thought that murder is wrong depends on my having had certain experiences in virtue of which I grasp the concept *murder*, this has no tendency to show that my knowledge that murder is wrong is not *a priori*. For as proponents of the *a priori* have long emphasized, we should distinguish between the *enabling* role that experience plays in allowing one to grasp a certain content, and the *evidential* role that experience might or might not play in justifying one in thinking that the relevant content is true. The status of a given piece of knowledge as *a priori* or *a posteriori* concerns the latter, and not the former.⁸

⁸ The point goes back at least to Leibniz and is a commonplace in contemporary discussion of the *a priori*. See, for example, Audi (1997: 100), BonJour (1998: 9–10), and Boghossian and Peacocke (2001: 2).

C4.P17 *Triggering.* An obvious way in which someone might fall short of knowing a proposition is by failing to believe it. Given this, if there is any moral knowledge at all, then it seems like experience and observation can contribute to its attainment by prompting us to take up beliefs that we would not have held in the absence of those experiences and observations. For example, prior to the rise of Nazism, Einstein was an absolute pacifist who maintained that the use of violent force was unjustified in any circumstances. In response to the rise of Nazism, Einstein changed his mind and became convinced that there are at least some possible circumstances in which the use of such force is justified.⁹ Let's assume for the sake of argument that Einstein's earlier absolutism was mistaken, and that his later view that there are at least some possible circumstances in which the use of force is permissible is both true and something that he knew. Perhaps this is something that he never would have come to believe if not for his having lived through the Nazi era; if so, then his experience of that historical event played a crucial role in his acquiring this piece of knowledge.

C4.P18 Notice, however, that even if Einstein's experience of actual historical events is what prompted him to take up the relevant belief, it does not follow that the resulting knowledge should be classified as empirical as opposed to a priori. This is because not every case in which experience prompts a person to take up a belief that qualifies as knowledge should be understood as a case in which the experience provides *evidence* for that belief. Suppose, for example, that (i) any nearby possible world in which Einstein reflects on the possibility of a "powerful opponent whose unconditional goal is to destroy me and my people" (cf. fn. 118) is a possible world in which he recognizes (and thus, comes to know) that there are some possible circumstances in which the use of force is morally permissible, but that (ii) in our world he would as a matter of contingent fact never have engaged in the relevant kind of reflection if not for the actual experience of having lived through the Nazi era. If so, then his knowledge that there are some possible circumstances in which the use of force is justified might count as a priori knowledge, for much the same reason that a mathematician's intellectual recognition of certain truths might count

Notice that the fact that our moral knowledge is dependent on sense experience in this respect does not do anything to distinguish it from much mathematical knowledge arrived at via proof, which I take it is a priori knowledge if anything is. For presumably, our possession of many of the concepts that we employ in mathematical reasoning is dependent on sense experience in the same way.

⁹ On Einstein's early absolute pacifism and the subsequent evolution of his view, see Rowe and Schulmann (2007). As he clarified his new position in a later letter: "there are circumstances in which in my opinion it is necessary to use force...such a case would be when I face a powerful opponent whose unconditional aim is to destroy me and my people" (quoted in Rowe and Schulman 2007: 211).

as a priori even if she would not have arrived at that recognition but for having perceived certain written symbols on the page.¹⁰

C4.P19 Compare the following case. Many people believe that the practice of involuntary female genital mutilation is morally wrong. Let's assume, for purposes of illustration, that at least some of the many people who believe this proposition also know that it is true. It is plausible that few if any of the people who know the relevant proposition would believe it, if female genital mutilation had never actually been practiced. However, it would be strange to think that their knowledge that the practice is wrong is a matter of their having empirical evidence that it is, empirical evidence that they would lack in any possible world in which the practice never actually occurred. A more plausible account is the following. For most of those who know that the practice is wrong, the role that experience plays in their knowledge is that of prompting them to take up a moral view that they would otherwise have lacked, but which they have the capacity to recognize as true upon considering it. The capacity to recognize the relevant moral view as true upon consideration is one that they also have in nearby possible worlds in which the practice does not actually occur. It is just that usually—but not invariably¹¹—the capacity remains unexercised in those worlds.

C4.S3

4.3 Confirmation and Disconfirmation

C4.S4

4.3.1 Clarifying the Issue

C4.P20 Even if observation sometimes leads to moral knowledge by calling our attention to possibilities that would otherwise have escaped notice, it does not follow that our moral views are ever *confirmed* by observation. In general, an observation confirms a view just in case it is evidence that the view is true; it disconfirms a view just in case it is evidence that the view is false. Although talk of empirical confirmation and disconfirmation naturally calls to mind the testing of scientific hypotheses in laboratory settings, the general phenomenon is pervasive in everyday life. For example, a professor's belief that a particular student is a good student might be either confirmed or disconfirmed by the student's performance on a final exam, because it makes sense

¹⁰ For further discussion of the mathematical case, see Kim (1981).

¹¹ The same capacity might be exercised, for example, in response to a philosophy professor's presentation of a purely hypothetical example in an ethics seminar, as an example of something that it would be morally wrong to do.

for the professor to adjust his or her credence in the relevant proposition in the light of that performance. Are our moral views similarly susceptible to being confirmed or disconfirmed by observation?¹² It might seem obvious that the answer to this question is “no.” After all, what would count as observational evidence either for or against the view that the death penalty is intrinsically immoral, or for or against the view that the agent in the Trolley Case is morally required to flip the switch?¹³

C4.P21 Of course, it is clear enough that at least *some* moral claims are susceptible to empirical confirmation and disconfirmation. For example, the credibility of the claim that it is wrong to encourage children to smoke depends on empirical evidence about the effects of smoking. (Similarly, the credibility of the moral claim that Smith wronged Jones by hitting him is highly sensitive to empirical evidence that bears on the question of whether Smith actually hit Jones.) Our knowledge that it is wrong to encourage children to smoke is not a plausible candidate for *a priori* knowledge. But that is because, on a natural reconstruction of the case, the empirical evidence puts us in a position to know some *non-moral* proposition or propositions (e.g. a non-moral proposition to the effect that smoking tends to have such-and-such effects). We then combine that non-moral, empirical knowledge with a piece of moral knowledge that is *not* known on the basis of empirical evidence (perhaps: the proposition that it is wrong to encourage children to do things that are likely to harm their health) in order to arrive at the knowledge that it is wrong to encourage children to smoke.

C4.P22 According to a picture that I will call **Generic Rationalism**, this natural story about how we know that *it is wrong to encourage children to smoke* generalizes. That is, according to Generic Rationalism, the *core* or *fundamental* truths of morality are known (when they are known at all) *a priori*; empirical evidence comes into play only in the application of this core moral knowledge. On the Generic Rationalist picture, the fact that it is wrong to encourage children to smoke is *partially grounded* in some more fundamental moral fact(s) that we know *a priori*, and *partially grounded* in some non-moral fact(s) that we know on the basis of empirical evidence.¹⁴ On this picture, although the contribution that empirical evidence makes to moral knowledge might be

¹² Harman (1977, chapter 1) is perhaps the classic statement of this question.

¹³ Recall Sayre-McCord's claim that “the results of empirical investigation appear to be altogether irrelevant to the justification of our moral views (although not to their application)” and Shafer-Landau's claim that “Ethics cannot rely on sense evidence... for any moral theory is perfectly compatible with such evidence.”

¹⁴ The recent literature on grounding is massive, but for an illuminating account, see Rosen (2010).

substantial and important (someone who was relatively ignorant of empirical facts would do poorly as a moral agent), there is an obvious respect in which it is of secondary interest to moral epistemology. For according to Generic Rationalism, the *distinctively moral* content of our views is not justified by empirical evidence; rather, empirical evidence becomes relevant only when we apply moral truths that we grasp a priori.¹⁵

C4.P23 An alternative to Generic Rationalism is a view that I will call **Empirical Reductionism**. According to the Empirical Reductionist, even relatively fundamental moral truths require empirical evidence in order to be known, for even relatively fundamental moral truths are *reducible* to sets of non-moral, empirical truths. When we are in a position to know a moral truth of this kind, it is because we know the non-moral, empirical truths that ground it, and empirical evidence is required in order to know the non-moral truths that make up the reduction base. Because Empirical Reductionism holds that empirical evidence is required even in order to know relatively fundamental moral truths, it assigns a greater role to empirical evidence than Generic Rationalism. Notice, however, that the Empirical Reductionist and the Generic Rationalist agree that the significance of empirical evidence for moral epistemology is that it puts us in a position to know *non-moral* truths that ground the truth of various moral claims.

C4.P24 When empirical evidence confirms a moral claim because it confirms some non-moral claim that either grounds or potentially grounds the moral claim, let's say the empirical evidence **indirectly confirms** the moral claim. Thus, when empirical evidence increases the credibility of the moral claim *it is wrong to encourage children to smoke* because it confirms the non-moral claim *smoking tends to have such-and-such effects*, it indirectly confirms the moral claim. Similarly, if empirical evidence increases the credibility of the moral claim that *Smith wronged Jones by hitting him* because it confirms the non-moral claim that *Smith hit Jones* (a claim that helps to ground or potentially ground the moral claim), that empirical evidence indirectly confirms the moral claim. Let's say that empirical evidence **directly confirms** a moral claim just in case it confirms that moral claim but not because it confirms the claim indirectly.

¹⁵ I label this view “generic” rationalism in order to emphasize that it is neutral among competing accounts of how we arrive at a priori knowledge of the fundamental or core moral truths. For example, it is neutral between the view that the core moral truths are conceptual or analytic truths that we come to know by exercising our conceptual or linguistic competence and the view that the core moral truths are synthetic a priori claims that we come to know by acts of rational intuition. For a critical discussion of some of the available options here, see section 4.5 below.

C4.P25 Neither the Generic Rationalist nor the Empirical Reductionist holds that observations confirm moral claims directly. Rather, both hold that when empirical evidence confirms a moral claim, it does so by way of confirming some non-moral claim on which the truth of the moral claim depends.

C4.P26 In what follows, I argue that our moral beliefs are susceptible to being confirmed and disconfirmed by observational evidence in ways that go beyond those recognized by either Generic Rationalism or Empirical Reductionism. Against Generic Rationalism, I defend the view that even the most fundamental moral truths are susceptible to being confirmed and disconfirmed by observational evidence. Suppose the most fundamental moral truths are necessary, knowable *a priori*, and irreducibly normative. I argue that still, non-moral observations that deliver contingent information about how things stand in *this* world can supply us with evidence that bears on the credibility of those truths. Against both Generic Rationalism and Empirical Reductionism, I argue that the role of non-moral observation is not limited to that of confirming non-moral, empirical claims that ground (either wholly or partially) moral claims. Rather, non-moral observations sometimes confirm moral claims *directly*. Non-moral observation is thus of greater interest and importance for moral epistemology than either Generic Rationalism or Empirical Reductionism suggests.

C4.P27 I will begin my defense of this alternative picture by arguing against some reasons for thinking that non-moral observations cannot directly confirm moral claims. I will then offer a positive case for thinking that the relevant kind of confirmation sometimes occurs.^{16,17}

¹⁶ Elsewhere I have defended the view that some observations are *moral observations*. (See McGrath 2018, which attempts to improve on the arguments of McGrath 2004.) But I believe that there is a compelling case to be made for the view that moral beliefs are sometimes observationally confirmed or disconfirmed even when we restrict our attention to the uncontroversial phenomenon of non-moral observation. I will focus on this issue in what follows. From this point on, “observation” can be read as “non-moral observation.”

¹⁷ The possibility of moral confirmation even in the absence of a full-blown reduction of the moral to the non-moral is resourcefully defended by Sturgeon (1985, 1986) in the context of his classic debate with Harman (1977, 1986b). Sturgeon’s approach has some important affinities, but also some important differences, with the approach developed here. Significantly, the primary focus of the Harman-Sturgeon debate is not the relation of confirmation but rather that of *explanation*, and in particular, the question of whether we should ever regard moral facts as relevant to the explanation of non-moral facts, especially psychological facts about why we make the moral judgments that we do. (Thus, the key example that provides the starting point for the debate is whether it’s reasonable to suppose that a certain “moral observation”, namely the demonstrative judgment that “That’s wrong!”, made in response to witnessing a group of children set a cat on fire, might be best explained by the putative moral fact that *it’s wrong to set the cat on fire*.) I want to distinguish sharply between the question of whether moral beliefs or observations are ever best explained by moral facts and the question of whether non-moral observations sometimes confirm moral beliefs, in part because the view that there is a close connection between these two questions depends on a contentious view about the relationship between explanation and confirmation that I do not wish to presuppose. In general, for

C4.S5 4.3.2 What the Possibility of Confirming Moral Claims by Observation Does Not Presuppose

C4.P28 What would it take for a person to make an observation that confirms one of her moral views? To a first approximation:

C4.P29 (i) she makes an observation;

C4.P30 (ii) she treats the content of that observation as a reason to increase her confidence in the moral view; and

C4.P31 (iii) it is rational for her to treat the content of the observation in that way.

C4.P32 Presumably, the rationality proviso is essential: even if my psychology is such that I am hard-wired to increase my confidence that *the death penalty is intrinsically immoral* in response to observing that *there is a mouse in my kitchen*, actually observing a mouse in my kitchen would not confirm my belief about the death penalty.

C4.P33 Why might one think that our moral views are not susceptible to being confirmed and disconfirmed by non-moral observations? A natural and familiar thought is that there is a kind of “logical gap” between the contents of our non-moral observations and the contents of our moral views, and that the existence of this gap effectively guarantees that the former have no evidential bearing on the latter. Our observations deliver information about how things are, while our moral beliefs concern how things ought to be. In this respect, moral beliefs and non-moral observations seem to have distinct subject matters. Perhaps then, the obvious explanation of why my observing a mouse in my kitchen fails to either confirm or disconfirm my belief about the morality of the death penalty *generalizes*: in general, non-moral observations do not confirm or disconfirm moral beliefs because their contents concern distinct subject matters.¹⁸

both Harman and Sturgeon, there *is* a close connection between explanation and confirmation, because both philosophers seem to accept a broadly explanationist model of confirmation, according to which E’s confirming H involves H’s standing in an explanatory relation to E. (For this aspect of Harman’s epistemological views, see especially his early (1965, 1967, 1968) papers and Harman 1973/2015.) In what follows, I will not assume that confirmation is the converse of explanation, or (more generally) that E must stand in some kind of explanatory relation to H in order to confirm it. That having been noted, the brand of “non-reductive moral naturalism” developed by Sturgeon in the course of his exchange with Harman and in other work harmonizes very well with (although it is not entailed by) what I say both in this section and elsewhere in the book.

¹⁸ Notice that this line of thought goes significantly beyond “Hume’s principle,” the thesis that—in the words of one of its most recent proponents—“No amount of empirical discovery about the state of

C4.P34 This line of thought has some intuitive appeal. But we should be wary of accepting it too quickly. The general trajectory of confirmation theory as it has developed over the course of its history gives us reason to be suspicious of claims that statements of one type cannot confirm statements of another type due to a gap in their subject matters. In fact, a great deal of the development of confirmation theory in the twentieth century was driven by the issue of how observational evidence can bear on hypotheses that are not themselves about observations. For example, how can theoretical hypotheses about electrons and protons be confirmed or disconfirmed by observational claims that make no reference to electrons or protons at all, but which instead are concerned with what would seem to be an entirely different subject matter—for example, the readings of scientific instruments in a laboratory?¹⁹

C4.P35 The diverse attempts to provide an account of confirmation that would allow for such possibilities typically share a common idea: specifically, the idea that confirmation should be understood not as a two place relation that holds between observational evidence and theoretical hypotheses, but rather as a three place relation that holds between observational evidence, theoretical hypotheses, and some third parameter. On one way of spelling out this general idea, the third parameter consists of a body of putative *analytic* or *definitional* equivalences between claims couched in the observational vocabulary and claims couched in the theoretical vocabulary.²⁰ On a different but related approach, the auxiliary hypotheses that connect the observational evidence to the theoretical hypotheses are not assumed to be analytic or definitional, but are assumed to be further empirical claims. On the more ambitious versions of either of these approaches, the possibility of confirming a theoretical claim by observation in effect required a *reduction* of the theoretical to the observational. In the case of the first approach, the reduction would be analytic or

the world can entail any conclusion about what ought to be without a further premise about what ought to be" (Dworkin 2011: 17). In general, evidence can *confirm* a hypothesis even if it does not *entail* that hypothesis. (For example, the observation that the suspect's fingerprints are on the murder weapon might confirm the hypothesis that the suspect committed the murder, even though it does not entail that he did.) Therefore, even if some version of Hume's principle is true, it does not follow that empirical discoveries never confirm moral conclusions. Still, the general picture that leads some to endorse Hume's principle might also lead one to endorse the stronger claim that empirical discoveries never *confirm* moral claims, and it is this line of thought that I want to consider.

For further discussion of Hume's principle, see section 5.7.3.

¹⁹ For an overview of this type of concern as a driving force behind the development of confirmation theory in its early decades, see Glymour 1980 especially pp. 10–62.

²⁰ This route was pursued with particular vigor by Carnap. For a retrospective overview, see Carnap (1963).

118 MORAL KNOWLEDGE

definitional. In the case of the second approach, the reduction would be empirical or synthetic.

C4.P36 If it were true that the confirmation of one type of statement by statements of another type required that statements of the first type be reducible to statements of the second type, this would provide a principled motivation for denying that non-moral observations can confirm moral claims. In particular, if such reducibility were required for empirical confirmation, then any philosopher who is not an Empirical Reductionist would have reason to deny that moral claims can be empirically confirmed and disconfirmed. Moreover, even a philosopher who holds that moral claims are *in principle* reducible to observational claims might very well hold that we are not in a position to observationally confirm or disconfirm any moral claims, provided that we are ignorant of *how* the moral reduces to the observational.²¹

C4.P37 However, it is now generally agreed that relations of confirmation can obtain between statements about different subject matters even when statements of the one type are not in any sense reducible to statements of the other type. Here are some examples:

- C4.P38 (1) Claims about the outcomes of experiments can confirm theoretical claims about subatomic particles, despite the fact that theoretical claims about subatomic particles are not reducible to claims about the outcomes of experiments.
- C4.P39 (2) Claims about how things are now can confirm claims about how things were in the past, despite the fact that claims about how things were in the past are not reducible to claims about how things are now.
- C4.P40 (3) Claims about how things are now can confirm claims about how things will be in the future, despite the fact that claims about how things will be in the future are not reducible to claims about how things are now.
- C4.P41 (4) Claims about how things appear can confirm claims about how they are, despite the fact that claims about how things are cannot be reduced to claims about how they appear.
- C4.P42 (5) Claims about bodily movements of other people can confirm claims about their mental states, even though claims about mental states are not reducible to claims about bodily movements.

²¹ The point that the reducibility of moral claims to claims to which our cognitive access is relatively unproblematic might be of limited epistemological significance given the apparent widespread ignorance of the details of any such reduction is made by Schroeder (2007: 170–5), who offers an interesting critical discussion.

- C4.P43 The analogous claim in the moral case would be:
- C4.P44 (6) Claims about how things are can confirm moral claims about how they ought to be, despite the fact that moral claims about how they ought to be are not reducible to claims about how they are.
- C4.P45 Thus far, I have not offered any positive reason for thinking that (6) is true. The point is just that even if moral claims are irreducible to non-moral claims, that does not show observational claims cannot confirm moral claims. Examples (1)–(5) show that relations of confirmation can obtain between statements about distinct subject matters, even though statements of one type are not in any sense reducible to statements of the other type.
- C4.P46 Before turning to the positive case for (6), we should note that the possibility of confirming moral claims by observation is also consistent with the truth of *moral error theory*. According to moral error theorists, putative moral properties such as *being morally wrong* do not exist, or are never instantiated. Thus, ordinary moral claims like “It is right to keep your promises” or “It was wrong for the 9/11 hijackers to fly planes into the Twin Towers” are false, because they attribute to actions properties that nothing ever has. In this respect, talk of right and wrong and other moral properties is on a par with talk about phlogiston.²²
- C4.P47 Suppose that it turns out that moral error theory is true: all claims of the form “x is wrong” are false, and so no such claims are ever known by anyone. It does not follow that such claims are never confirmed by observation. False claims—including claims that are false because they postulate properties or relations that are never instantiated—sometimes *are* confirmed by observation. The history of science is filled with theories that postulate entities that never existed, and properties and relations that have never been instantiated. It does not follow, and indeed, it is not true, that such theories were never confirmed by observation. For example, although there is no such thing as phlogiston, there was in fact observational, confirming evidence in favor of the phlogiston theory of combustion.²³ The same is true of countless other now discarded theories. Therefore, the possibility of moral confirmation does not entail the falsity of moral error theory. (At best, the truth in the vicinity is: if you *know that a claim is false*, that claim can’t be confirmed for you. Because

²² On moral error theory, see especially Mackie (1977) (for the classic statement of the view) and Joyce (2001, 2016b). The comparison with phlogiston is used by Joyce in order to explicate the view in both of these works.

²³ For a useful overview, see especially Chang (2010).

if you know that it is false, any observations that would have served to confirm it will instead confirm some alternative claim(s) that are not yet ruled out by what you know.) I will return to the relationship between moral confirmation and moral error theory after I present the positive case for moral confirmation.

C4.S6

4.3.3 An Initial Argument in Favor of the Possibility of Confirming a Moral Claim by Non-Moral Observation

C4.P48 Let's now turn to the positive case for thinking that moral confirmation is possible. First, consider the way in which the currently most widely accepted account of empirical confirmation, Bayesianism, accommodates facts like (1) and (2):²⁴

- C4.P49 (1) Claims about the outcomes of experiments can confirm theoretical claims about subatomic particles, despite the fact that theoretical claims about subatomic particles are not reducible to claims about the outcomes of experiments.
- C4.P50 (2) Claims about how things are now can confirm claims about how things were in the past, despite the fact that claims about how things were in the past are not reducible to claims about how things are now.

C4.P51 The Bayesian retains the idea that confirmation is fundamentally a three-place relation rather than a two-place relation. According to the Bayesian, in typical cases of confirmation, the confirmation relation holds between an observation, a hypothesis, and the observer's *prior probability distribution*. Within this framework, a hypothesis H would be confirmed by observational evidence E just in case the prior probability that the observer assigns to H is lower than the prior probability that she assigns to H on the condition that E (i.e. $\text{Pr}(H) < \text{Pr}(H/E)$). The hypothesis H is confirmed for the observer when the relevant observation is actually made. On this account, the fact that a scientist can be in a position to confirm a theoretical hypothesis about subatomic particles by observing a certain experimental outcome is grounded in the fact that she assigns a higher probability to the hypothesis conditional on that experimental outcome than she does to the hypothesis unconditionally. Similarly, the fact that we are sometimes in a position to confirm a historical claim by observing that things are *currently* thus and so is grounded in our

²⁴ Useful primers for Bayesian confirmation theory include Horwich (1982) and Easwaran (2011).

assigning a higher probability to the historical claim conditional on things being currently arranged thus and so than we do to the historical claim unconditionally.

C4.P52 Within this general framework, the possibility of confirming a *moral* belief by a non-moral observation would amount to this: the believer rationally assigns a higher prior probability to the moral claim conditional on a certain non-moral observation than she does to the moral claim unconditionally, and then she makes the relevant observation. As a purely psychological matter, it seems clearly *possible* that a believer could distribute her prior probabilities in the relevant way. A theorist who denies that moral confirmation is possible should deny that it would ever be *rational* for a believer to distribute her credences in this way. That is, the theorist should deny that it would ever be rational for a believer to assign a higher probability to a moral view conditional on a non-moral observation that is higher than the probability that the believer assigns to the same moral view unconditionally.²⁵

C4.P53 But, in fact, it can be perfectly reasonable to distribute your probabilities in the relevant way. Suppose that you initially find yourself uncertain about whether some moral claim is true; you thus assign a probability of n to the hypotheses, where n is some value significantly less than 1 but significantly more than 0.²⁶ Suppose that you decide to ask a friend, whom you reasonably

²⁵ According to *Subjective* Bayesians, *any* prior probability distribution is rational, so long as it is probabilistically coherent (i.e. obeys the laws of the probability calculus). Given the truth of Subjective Bayesianism, the thesis that non-moral observations can confirm moral claims follows immediately, for it is clearly possible for a believer to have a coherent prior probability on which the conditional probability of some moral claim conditional on a certain non-moral observation is higher than the unconditional probability of the moral claim itself. However, many hold that Subjective Bayesianism is too liberal in allowing for confirmation, and that not every prior probability is rational. As White (2015) notes, the subjective Bayesian will hold that it can be perfectly rational for me, in advance of any observational evidence, to be extremely confident that when I look to the heavens I will see that the stars are aligned in such a way as to constitute a perfect representation of my face. Similarly, applied to the moral case, Subjective Bayesianism entails that that the observation that there is a mouse in my kitchen *does* confirm or provide evidence for me *the death penalty is immoral*, provided that (i) my psychology is eccentric in the right way, and (ii) my degrees of belief are probabilistic coherent. I agree with the common thought that this approach makes empirical confirmation and rationality too easy to come by, in both the moral and non-moral domains. According to *Objective* Bayesians, not every coherent prior probability distribution is rational; rather, there are substantive constraints on prior probability distribution beyond the demand for purely formal coherence. Having noted that if Subjective Bayesianism is true, then the conclusion that this section aims to establish follows straightforwardly, I will assume that Subjective Bayesianism is false. The burden of the paragraphs that follow is to show that, even on a more objective picture of confirmation, we still have compelling reasons to accept that conclusion.

²⁶ Here I assume, in the standard way, a scale on which probability 1 represents perfect certainty that a proposition is true and a probability of 0 represents perfect certainty that it is false.

I also assume that there are circumstances in which it would be reasonable to invest some level of credence in a moral view that is between 0 and 1, even if we restrict our attention to fundamental moral claims and assume that any such claim is metaphysically necessary if it is true at all. (Compare: even if the claim that *Water is H₂O* is a metaphysically necessary truth, there are many possible circumstances in which it is reasonable to be far from certain that it is true.)

regard as having good moral judgment. Because you regard your friend as having good moral judgment, you assign probability $n+m$ to the moral claim conditional on your friend's judging that it is true (where $m>0$). Your friend then tells you that she thinks the moral claim is true. In response, you increase your confidence in the claim. In doing so, you rationally treat a certain non-moral observation as evidence in favor of a moral claim.

C4.P54 We should note that the assumptions about moral testimony that are required in order for this to be a rational response on your part are extremely weak, and that they are consistent with a relatively skeptical attitude towards the epistemic force of moral testimony. In particular, a theorist might admit that a modest increase in confidence is reasonable in the circumstances even if she denies that moral testimony can ever deliver full-fledged knowledge, and even if she denies that it would ever be reasonable to adopt a moral view on the basis of testimony that conflicts with the view that one would have held if one had made up one's mind autonomously. Again, all that is required is the modest claim that it can be reasonable to treat your friend's opinion as a reason for being at least somewhat more confident than you were before you found out what she thought. Moreover, the plausibility of this modest claim does not seem to depend in any way on the modal status of the moral view of which you are uncertain, whether it could *in principle* be known a priori, or on whether it is irreducibly normative.

C4.P55 When you learn that your friend believes as she does, this is a piece of non-moral, empirical knowledge about how things stand in the actual world. The content of your observation is not itself a moral claim; rather, it is a non-moral proposition to the effect that your friend has endorsed the moral view of which you were uncertain.²⁷ The upshot, I take it, is this. *If there are possible*

²⁷ This description of the case is not completely uncontroversial. On some views of the epistemology of testimony (see especially McDowell 1980), in paradigmatic cases one does not move (either consciously or unconsciously) from a belief of the form "S said that p" to a belief with the content p; rather one directly takes up the content which is presented as true by the testifier. (For example, in a conversation in which you report that you were born in England, in the usual case I will simply add the proposition that you were born in England to my stock of beliefs, without inferring this from the belief that *you asserted that you were born in England*, or something similar.) If this model is applied to the moral case, then increasing one's confidence in a moral view on the basis of someone else's testimony would *not* involve a non-moral observation that has as its content the fact that someone engaged in a certain speech act; rather one, would simply adjust one's attitude toward the moral content directly. However, even if McDowell is right that this is the usual case of belief revision via testimony, it is obviously not the only case. For example, you might learn from the testimony of a third person that your trusted friend endorses the moral view of which you are uncertain. In that case, even on McDowell's view of testimony, the empirical evidence that you gain via testimony consists in a non-moral proposition. If McDowell's account is correct, then the case described in the main text can be amended in the relevant way, so that we once again have a (now more complicated) case in which a non-moral observation is rationally treated as evidence in favor of a moral belief.

circumstances in which testimony can make it reasonable to be at least marginally more confident that some moral belief is true, then there are possible circumstances in which a moral belief is confirmed by a non-moral observation, because the former is just a special case of the latter. Similarly, if there are possible circumstances in which receiving testimony to the effect that a given moral view is false can make it reasonable to be at least marginally less confident of that moral view, then there are possible circumstances in which a moral belief can be *disconfirmed* by a non-moral observation. Again, the former is a special case of the latter. But for reasons rehearsed in the previous chapter, we *should* think that there are circumstances in which it would be reasonable to become more confident of a moral belief in response to learning what someone else thinks about it, as well as circumstances in which it would reasonable to become less confident of a moral belief in response to learning someone else's opinion. Therefore, we should believe that it is possible for non-moral observations to confirm and disconfirm moral beliefs.

C4.P56 Moreover, notice that, even if the moral claim that your friend endorses is true, it will not be plausible to think that its truth is even partially *grounded* in the fact that your friend thinks that it is. A case in which a moral belief is confirmed by testimony is a case in which a non-moral observation confirms a moral claim but *not* by way of confirming some non-moral claim that grounds the truth of the moral claim. It is a case in which a non-moral observation confirms a moral belief *directly*, in the sense given above.²⁸

A very different challenge to my description of the case as involving a non-moral observation is offered by an anonymous referee, who suggests that the content of the observation in the example might be better regarded as a *morally competent judge asserts that p* (or something similar). I am skeptical that this *could* be the content of an observation, as opposed to the deliverance of an inference from an observation like *Ted asserts that p* and a standing background belief that *Ted is a morally competent judge*. But in any event, even if it is possible to observe contents like *A morally competent judge asserts that p*, I assume that it is also possible to observe the more minimal content that *Ted asserts that p*, and this is the kind of case at issue in the text. Of course, the fact that you increase your credence in in the moral proposition that p in response to learning via observation that Ted judges that p might very well depend on the presence of a standing background belief that Ted is a morally competent judge. If so, does this undermine the claim that this is a case in which a moral claim is confirmed by a non-moral observation? No, for the same dependence on standing background beliefs is present in paradigmatic cases of confirmation outside the moral domain. Compare: observation about the outcomes of experiments can confirm theoretical hypotheses about subatomic particles, even if this depends on the scientists having certain standing background beliefs about the connections between the experimental outcomes and the theoretical hypotheses in question. Sturgeon (1985) emphasizes this parallel dependence on background beliefs, although he is not concerned with the case of testimony.

²⁸ In the last section, we noted that the possibility of moral confirmation does not entail the falsity of moral error theory. Given the conclusion of this section, we can now say the following: the fact that moral confirmation might occur even in a world in which moral error theory is true is witnessed by the fact that, in a given case, someone in that world who is ignorant of the truth of moral error theory might rationally raise her credence in a given moral claim upon learning that that claim is endorsed by another person.

C4.S7

4.3.4 Beyond the Case of Testimony

C4.P57 I have just argued that, given modest assumptions about the way in which it can be rational to respond to empirical discoveries about the moral opinions of others, the following question should be answered affirmatively:

C4.P58 (i) Are moral beliefs ever confirmed or disconfirmed by non-moral observations?

C4.P59 Still, that leaves open the following question, which might be thought to be the question of greatest philosophical interest in the neighborhood:

C4.P60 (ii) Are moral beliefs ever confirmed by non-moral observations, where the non-moral observations are not evidence about the moral opinions of others?

C4.P61 I think that this question should also be answered affirmatively. Before turning to my main argument for that conclusion, let me quickly offer a reason for thinking we should *expect* question (ii) to be answered in the affirmative, given our answer to question (i). The example of testimony shows that an observation to the effect that a certain non-moral, contingent state of affairs obtains can count as evidence for a moral claim. The fact that *our* possible world turns out to be a certain way, as a matter of contingent fact, can legitimately be treated as indicating that a particular moral view is true, even if that view is taken to be irreducibly normative, knowable *a priori*, and necessarily true if true at all. Now, given that there is at least one type of empirically discoverable non-moral fact that can legitimately be treated as evidence for and against such moral claims (namely that facts about the moral opinions of others), this should increase the plausibility of the idea that there are *other* observable, contingently obtaining states of affairs that can legitimately be treated as such evidence. Indeed, it would be surprising (and something that calls out for explanation) if non-moral, empirical facts about what other people believe were the *only* non-moral, empirical facts that could ever play this role.

C4.P62 As a first step towards a more substantial argument in favor of an affirmative answer to question (ii), consider the following closely related question:

C4.P63 (iii) Are moral beliefs ever *disconfirmed* by non-moral observations?²⁹

C4.P64 Here is a hypothetical, but realistic case in which it's plausible to think that such disconfirmation occurs:

²⁹ Again, setting aside the case in which the non-moral observations are evidence for the moral opinions of others. From this point on, I will treat this qualification as understood.

C4.P65 **The Story of Ted.** At time t_0 , Ted believes that same-sex marriage is morally objectionable. His opposition to same-sex marriage involves both a belief that people of the same sex should not enter into such marriages given the opportunity to do so, as well as the belief that it is wrong for society to recognize such marriages, or to treat them as having equal standing to marriages between a man and a woman. Ted also believes that the recognition of same-sex marriages, and an increase in their number, would have numerous negative consequences. Some of these consequences he would characterize in explicitly moral terms, while others he would characterize in non-moral terms (e.g. “a significant increase in the divorce rate”). However, although Ted believes both that same-sex marriage is morally objectionable and that it would lead to consequences that he regards as bad, he does not believe that it is morally objectionable *because* it would lead to bad consequences, or that its (alleged) wrongness is in any way reducible to its having those consequences. Because of this, if he were asked at time t_0 whether he thinks that same-sex marriage would still be morally objectionable even if it did not have the bad consequences, he would truthfully report his opinion as follows: “Yes, the wrongness of same-sex marriage does not consist in its having such-and-such consequences: rather, it is intrinsically immoral.” On the other hand, Ted does believe that there is the following connection between the (alleged) wrongness of same-sex marriage and the (alleged) fact that it would lead to negative outcomes for society as a whole: its moral wrongness is part of why it can be expected to lead to negative outcomes.

C4.P66 At this point, let me offer two remarks about the story of Ted. First, there is nothing formally incoherent about Ted’s views. Ted’s views towards same-sex marriage are very much like the views that many people hold toward *lying*. Many people hold the following beliefs about lying: (1) lying is wrong; (2) the wrongness of lying does not consist in, nor is it any way reducible to, facts about its consequences; (3) nevertheless, lying *does* tend to lead to bad consequences; and (4) the fact that lying tends to lead to bad consequences is not independent of the fact that it is wrong. (Perhaps the last belief is in some cases due to a more general belief to the effect that actions that are morally wrong tend to give rise to bad consequences.) Second, this combination of views toward same-sex marriage does not seem in any way unrealistic or even uncommon among people who oppose same-sex marriage. Indeed, most people who oppose same-sex marriage are not consequentialists, and thus would deny that its (alleged) wrongness in any way consists in or is reducible to its consequences. Nevertheless, they expect that same-sex marriage will lead to bad consequences, and that its having such consequences is intimately bound up with what they

126 MORAL KNOWLEDGE

take to be its wrongness. (That is, I take it that a significant number of actual people would accept the analogues of (1)–(4) in the case of same-sex marriage.)

C4.P67 Now consider:

C4.P68 **The story of Ted, continued.** The society in which Ted lives grants full legal recognition to same-sex marriage and begins to treat those marriages as having equal standing with marriages between a man and a woman. The number of such marriages steadily increases. Ted retains his expectation that over time these developments will lead to (among other things) a significant increase in the overall rate of divorce within society. However, no increase in the divorce rate occurs, an observed fact that Ted acknowledges to be true. In response to the falsification of his prediction, Ted decreases his confidence in the views that led him to make that prediction, including the view that same-sex marriage is morally wrong.

C4.P69 I take it that it would be reasonable for Ted to revise his view about same-sex marriage in this way, just as it is often reasonable to decrease one's confidence in a non-moral view that leads one (typically in conjunction with other views) to make a prediction that is ultimately falsified. Indeed, given the description of the case, Ted's responding this way seems at least as reasonable, if not more reasonable, than the alternatives (e.g. adopting a new sociological hypothesis that "explains away" the fact that the expected increase in the rate of divorce failed to materialize).

C4.P70 If this is the right verdict about the The Story of Ted, then it follows that moral beliefs can sometimes be *disconfirmed* by non-moral observations. From there, it is a short step to the conclusion that moral beliefs can sometimes be confirmed by non-moral observations as well. In a standard Bayesian framework, it is generally true that any belief that can in principle be disconfirmed by an observation can also in principle be confirmed by some (other) observation.³⁰ In general, the possibility that reality could turn out in such a way as to provide observational evidence in favor of a view that one holds is

³⁰ This is because a possible observation E disconfirms a hypothesis H just in case one's probability for H conditional on E is lower than one's unconditional probability for H (i.e. $\text{Pr}(H/E) < \text{Pr}(H)$); given that one's unconditional probability for H is the weighted average of one's probability for H on E ($\text{Pr}(H/E)$) and one's probability for H on not-E ($\text{Pr}(H/\text{not-}E)$), it follows that $\text{Pr}(H/E) < \text{Pr}(H)$ only if $\text{Pr}(H) < \text{Pr}(H/\text{not-}E)$, i.e. only if the probability of H conditional on not-E is higher than the unconditional probability of H. But then, it follows that observing not-E would *confirm* H, since in a Bayesian framework, confirmation consists in probability-raising.

Although it is generally true that any belief that can in principle be disconfirmed by an observation can also in principle be confirmed by an observation, there are deviant cases where this fails to hold. For example, the hypothesis "Nothing exists" is falsified by any possible observation, and is thus a

simply the flipside of the possibility that reality could turn out to provide observational evidence against that view.

C4.P71 Thus, the argument that there are possible circumstances in which a moral belief is confirmed by observational evidence runs as follows. If there are possible cases in which people receive observational evidence that disconfirms their moral beliefs, then there will also be possible circumstances in which people receive observational evidence that confirms their moral beliefs. But there are possible cases in which people receive evidence that disconfirms their moral beliefs (as witnessed by the The Story of Ted). Therefore, there are possible cases in which people receive observational evidence that confirms their moral views.

C4.P72 Even a relatively fundamental moral conviction (of a kind whose truth or falsity seems maximally independent of any contingent matters of fact) can give rise to expectations about how the world will turn out to be when it is conjoined with other views. When those expectations are disappointed, it can make sense to treat that outcome as evidence against the views that gave rise to them. When those expectations are fulfilled, it can make sense to treat that outcome as evidence in favor of those views.³¹

C4.P73 Next, I want to look at this phenomenon as it arises in a context that brings together the themes of coherence, confirmation, and other people.

C4.S8 4.3.5 When Facts About How People Actually Behave Provide Evidence for Claims About How We Should Behave

C4.P74 Consider the following case:

C4.P75 **Surprising Discovery.** Initially, Ted believes that ϕ -ing is seriously morally wrong. He also believes that Fred, with whom he has worked closely on a professional basis for years, is an ordinary, decent person who would not habitually engage in behavior that is seriously morally wrong. He then discovers that Fred habitually ϕ -s.

claim that can be disconfirmed even though it cannot be confirmed. Thanks to Yoaav Isaacs, via Tom Kelly, for this example.

³¹ Cf. the holistic picture of confirmation famously sketched in Quine (1951, sec.6)—although it seems clear that Quine himself had no sympathy for extending that picture to the moral domain. (On this last point, see especially Quine 1979.) I want to emphasize, however, that one can accept the kind of holism suggested here without endorsing some of Quine's more extreme claims, e.g. that particular observations test “the totality of our so-called knowledge or beliefs” (1961: 6).

C4.P76 There are, of course, different ways in which Ted could revise his beliefs in order to make them coherent. Here are three:

C4.P77 **Option 1:** Ted might retain his belief that ϕ -ing is seriously morally wrong while abandoning his original assessment of Fred. Responding in this way involves treating the surprising discovery as evidence that disconfirms his original assessment of Fred.

C4.P78 **Option 2:** Ted might abandon his belief that ϕ -ing is seriously morally wrong while retaining his original assessment of Fred. Responding in this way involves treating the surprising discovery as evidence that disconfirms his original moral belief about ϕ -ing.

C4.P79 **Option 3:** Ted might reduce his confidence in both the proposition that ϕ -ing is seriously morally wrong and his original assessment of Fred, in a way that renders his overall view coherent. Responding in this way involves treating the surprising discovery as evidence that disconfirms both his original assessment of Fred and his original moral belief about ϕ -ing.

C4.P80 If there is any way of filling in the details of Surprising Discovery so as to yield a version of the case in which it is rational for Ted to adopt either Option 2 or 3, then that version of the case will be one in which he gets empirical evidence that disconfirms his belief that ϕ -ing is seriously morally wrong. Moreover, any such version of the case will also be one in which Ted gets *confirming* evidence in favor of a moral proposition. For any evidence that disconfirms the proposition that ϕ -ing is seriously morally wrong for Ted will at the same time be evidence that confirms for him the moral proposition that *it is not seriously morally wrong to ϕ* .

C4.P81 Given only the facts stipulated in Surprising Discovery, there is, I think, no answer to the question of which of these options is the most reasonable; different answers will be called for by different ways of filling in the details of the case. In a version of the case in which ϕ -ing is *torturing children in order to relieve one's boredom*, Option 1 is clearly the way to go. Assuming that Ted is a more or less normal person, the view that *it is wrong to torture children in order to relieve boredom* is one that he will hold with a relatively high degree of rational certainty. Given this, it will be reasonable for him to maintain that belief in the face of surprising empirical discoveries, and to instead alter his non-moral beliefs (e.g. about which agents do or would engage in which behaviors) in response to such discoveries. But in other versions of the case, it might make sense for Ted to treat the surprising discovery as evidence against his original moral belief. Consider, for example:

C4.P82 **Surprising Discovery***. Initially Ted believes that participating in a romantic relationship with someone of the same sex is seriously morally wrong. He strongly believes that Fred, with whom he has worked closely on a professional basis for many years, is an ordinary decent person who would not repeatedly engage in behavior that is seriously morally wrong. Ted then discovers that Fred has been in a same-sex relationship for many years. In response to this discovery, Ted becomes less confident of his belief that it is seriously morally wrong to take part in a same-sex relationship.

C4.P83 If this could be a rational response on Ted's part, then Surprising Discovery* is a case in which an empirical discovery of a non-moral fact provides disconfirming evidence against a moral view. Moreover, in that case, Surprising Discovery* is also an example in which an empirical discovery of a non-moral fact confirms a moral view, for any evidence that *disconfirms* Ted's belief that *it is seriously morally wrong to participate in a same-sex relationship* will also be evidence that confirms for him the proposition that it is *not* seriously morally wrong to participate in such a relationship.

C4.P84 Of course, Ted is only in a position to treat the surprising discovery as evidence that bears on his moral views because of his standing background belief that Fred would not do something that is seriously morally wrong. It is this belief that allows the non-moral discovery about Fred's behavior to stand in evidential relationships to Ted's moral views. But this feature of the case does not cast doubt on my description of it as one in which a moral view is (dis)confirmed by a *non-moral* observation. For an analogous line of thought could be used to incorrectly cast doubt on the claim that theoretical hypotheses about subatomic particles are sometimes (dis)confirmed by observations about experimental outcomes (Cf. footnote 140, regarding the case of testimony). In general, it is a mistake to take this characteristic role of background beliefs (or prior probability distributions, etc.)—one that we consistently find across many, quite different domains—as providing grounds for doubting the original claims of confirmation.

C4.P85 I have claimed that Surprising Discovery* is a case in which an empirical discovery of a non-moral matter of fact provides disconfirming evidence against a moral view, and confirming evidence for the opposite moral view. Notice that the plausibility of these claims is independent of whether the truth about the morality of participating in a same-sex relationship is knowable *a priori*. Perhaps an ideally rational person would know *a priori*, and with perfect certainty, whether it is permissible to participate in a same-sex relationship. Maybe many actual people know *a priori*, and with perfect certainty, this same

truth, on the basis of the very reasons that would convince an ideally rational person. And maybe for any such person, the epistemic position that she occupies with respect to the relevant proposition, and her rational certainty that it is true, is such as to preclude her belief in its truth from receiving any (further) confirmation or disconfirmation from empirical sources. But the position that she occupies is obviously not the position occupied by Ted. Nor is it the position that we occupy, with respect to any moral proposition of which we are currently uncertain (even in cases in which the moral proposition in question is one that could *in principle* be known to be true or false on a purely a priori basis). Where there is moral uncertainty—even if it is not uncertainty that would be shared by a perfectly rational person—there will be room for empirical discoveries to make a difference.

C4.P86 Notice, moreover, that even on the assumption that *it is morally permissible to participate in a same-sex relationship*, it is not plausible that this truth is *grounded* (either wholly or in part) in the fact that Fred (or any other individual) participates in a same sex relationship. Therefore, Surprising Discovery* seems to be a case in someone gets empirical evidence that confirms a moral claim, but where this is not a matter of his getting evidence that confirms some non-moral claim that helps to ground the truth of the moral claim that is confirmed. (That is, it is not like a case in which empirical evidence confirms the moral claim that *it is wrong to encourage children to smoke* by way of confirming some non-moral claim about the effects of smoking.) In this respect, it is a case in which non-moral, empirical evidence confirms a moral claim *directly*, i.e. not by way of confirming some non-moral claim that potentially grounds (either wholly or partially) the truth of the moral claim.

C4.P87 Finally, notice that the verdict that Ted receives disconfirming evidence against his moral belief in Surprising Discovery* is one that can be accepted even by philosophers who are themselves of the opinion that participating in a same sex relationship is seriously morally wrong, i.e. that Ted's original belief is true. Such philosophers should, I think, regard the case as one in which Ted receives genuine but misleading evidence against his belief: the fact that Fred, a person who as far as Ted can tell is *in all other respects* a morally upstanding person, participates in a practice that is seriously morally wrong, is evidence that misleadingly suggests that the practice is not seriously morally wrong, given Ted's uncertainty about its morality.

C4.P88 The kind of empirical confirmation and disconfirmation that occurs in Surprising Discovery* seems to be not just a theoretical possibility, but something that actually happens with some frequency. Consider the rapid shift in

American public opinion about the morality of same-sex relationships.³² The extent of this shift in public opinion over a relatively short period of time cannot be accounted for exclusively or even primarily in terms of generational displacement: a large number of Americans seem to have *changed their minds* about the morality of same-sex relationships.³³ Although no doubt many diverse factors played a role in such changes of mind, many people report that the single most important factor was learning that someone who they knew personally (perhaps a family member, or neighbor, or co-worker) and generally held in high regard participated in same-sex relationships.³⁴

C4.P89 Consider those cases in which getting this kind of information did play a role in the relevant change of mind. One model for understanding such cases is that of a purely psychological process that is either irrational or arational: while it is relatively easy to think of strangers as immoral people who regularly engage in serious wrongdoing, it is too uncomfortable and unpleasant to think the same of close family members (etc.). People change their moral views in order to avoid discomfort or unpleasantness. On another model of the relevant process, the transition is, whenever rational, a case of *triggering*: becoming aware that a salient person is a participant in a same-sex relationship serves as a kind of causal cue, one that prompts a reconsideration of the quality of one's reasons for thinking that same sex relationships are morally wrong in the first place. Having been prompted to undertake further reflection by the experience, one judges that one does not have good reasons for this belief after all, and thus abandons it as a result of a priori reflection.

C4.P90 I don't doubt that at least some actual cases are best captured in terms of these models. But that should not lead us to neglect another type of case: acquaintance with a person who participates in an activity that one has considered wrong, but who in every other respect seems very far from immoral, is treated as some evidence (or reason to suspect that) that the activity in which she engages is not immoral, after all.³⁵

³² According to a May 2015 Gallup poll, 6 percent of Americans believe that same sex relationships are morally acceptable while 34 percent of Americans believe that they are morally wrong. Remarkably, these numbers are more or less the mirror images of what Gallup found when it asked the same questions in a poll conducted ten years earlier. See <http://www.gallup.com/poll/1651/gay-lesbian-rights.aspx>.

³³ On this point, see, for example, the 2013 Pew Research Center Report, "Growing Support for Gay Marriage: Changed Minds and Changing Demographics" (which addresses not just attitudes towards same-sex marriage but also towards same-sex relationships more generally).

³⁴ Ibid.

³⁵ In fact, I suspect that many actual cases of changes of mind about moral questions are psychologically far messier matters than what any simple model would suggest, and involve a combination of

C4.P91 Structurally, Surprising Discovery is an extremely simple case: the incoherence that creates rational pressure for Ted to revise his overall view arises among a small number of his beliefs, the beliefs in question concern a single individual, and the incoherence among them is a matter of outright logical inconsistency. But the same basic epistemic mechanism can be present in cases that differ from Surprising Discovery in any or all of these ways. You might believe, of a particular person, not that she *would not* engage in behavior that is seriously wrong, but that she is *unlikely* to engage in such behavior, or even (more weakly still) that she is less likely to engage in behavior given that it is morally wrong (a course of action's being morally wrong is at least negatively correlated with the likelihood that the agent will engage in it). A belief of this kind would, in a context in which you are confident that a certain type of behavior is morally wrong, play a role in shaping your expectations about what the agent is likely to do in the future, or what she has done in the past. But in a context in which you are uncertain whether a given type of behavior is wrong, the fact that the agent engages in it might count as evidence against the claim that the behavior *is* wrong, given your belief that the way in which the agent acts is not completely insensitive to moral factors.

C4.P92 You might treat as evidentially relevant to the morality of some behavior, not the fact that some *individual* engages in it or abstains from engaging in it, but rather the fact that the members of some group do, or even the fact that the behavior is common (or not) in your society. In such cases, there will be both normative and psychological pressure to maintain and improve the coherence of your overall view, and your overall view will contain both moral and non-moral beliefs. For example, if you hold, as many people do, views about the general moral condition of your own society (e.g. that your society is morally enlightened, or morally corrupt, or somewhere in between), then those views will bear rational connections to your normative views about the morality of certain practices, given non-moral views about the frequency of those practices within your society.

C4.P93 Although normative views about the morality of a given practice and non-normative views about the frequency with which people engage in that practice generally do not bear on one another when considered in isolation, they can bear on each other in the context of views that forge connections between the moral and the non-moral. The reason why a non-moral observation can

elements. For example, learning that a person who one generally holds in high regard participates in a practice that one previously considered wrong might both give one a reason to lower one's confidence in one's moral conviction as well as prompt further reflection on that conviction, a process that might result in a further diminution of one's confidence.

sometimes confirm or disconfirm a moral claim (even when this is *not* a matter of confirming or disconfirming some non-moral truth that grounds the moral claim) is that the non-moral observation can make an overall view less coherent, and in a given case the most reasonable way to respond to this incoherence might involve adjusting confidence in a moral view as opposed to (or in addition to) adjusting confidence in a non-moral view.

C4.P94 As noted above, the general mechanism under consideration here can generate misleading evidence, i.e. genuine evidence for a false moral view. On the current account, a potentially important source of misleading evidence in the moral domain is the possibility that people who either are or seem to be morally upstanding people *in other respects* will engage in some activity or practice that is in fact morally wrong, in a context where its moral wrongness is not generally recognized. Perhaps one reason why the seemingly evident moral atrociousness of slavery proved so difficult to grasp in many societies (even among those who did not personally profit from the institution of slavery) is that a good number of slaveholders seemed to be *morally upstanding people in other respects*—the kind of people who treated their families and friends well and did not cheat their business partners—in short, the kinds of people who seemed quite unlikely to participate in practices that are morally atrocious. Of course, from our perspective, such considerations might seem underwhelming to the point that there is a temptation to conclude that they never could have amounted to evidence for anyone at all. But from a perspective of genuine moral uncertainty, the same considerations might have seemed formidable. On this view, it is not simply that it would be hard to believe that people who seemed (and perhaps were) morally admirable in other ways were constantly engaged in morally atrocious behavior, but that their behavior in other contexts supplied evidence that made it more difficult to recognize their morally atrocious behavior as such.³⁶

C4.P95 The same mechanism can pose epistemological obstacles for would-be moral reformers. For example, some maintain that it is seriously wrong to eat meat. An obvious practical challenge to making this view plausible is that many seemingly good people regularly eat meat. On one view, the practical

³⁶ To be clear, nothing that I say here entails that that it was ever all things considered reasonable for anyone to believe that slavery is morally acceptable. For even if the behavior of some in a given historical context supplied misleading evidence about its wrongness to others, what it is rational for the latter individuals to believe about the morality of slavery presumably depends on their total evidence, as opposed to any proper part of their total evidence. If, as some hold, we *always* have sufficient reason to believe that slavery is morally wrong, then it will be reasonable for any individual in the relevant circumstances to believe this, even if she also possesses some evidence that suggests a different conclusion.

obstacle that faces would-be moral reformers is *purely psychological*: for non-rational reasons, it is simply difficult to convince people that anyone who regularly eats meat thereby regularly engages in behavior that is seriously wrong. But I am suggesting that there is often an additional, *epistemological* obstacle, one that contributes to the psychological difficulty of converting others. The epistemological obstacle is that from the perspective of a person who is currently uncertain whether eating meat is wrong, the fact that many seemingly good people regularly eat meat might be (defeasible) evidence that it is *not* morally wrong.

C4.P96 Of course, the would-be moral reformer can point to historical cases in which a behavior that is now generally condemned was once widely practiced and taken to be morally unproblematic, in order to make salient the possibility that the case at hand is among these cases. And she can explain why good people mistakenly engage in the practice in a way that reconciles its wrongness with their general goodness. In many cases, I suggest, we should view these as rational attempts to defuse, or blunt the force of, considerations that would otherwise amount to evidence that the practice is permissible.

C4.S9

4.4 Conditioning

C4.P97 In this section, I defend the idea that experience can contribute to moral knowledge in a way that is distinct from any of the ways that have been considered thus far. I call this the *conditioning* role of experience. To a first approximation: experience might refine or hone one's judgment to the point that subsequent exercises of that judgment are sufficiently reliable for its deliverances to count as knowledge. In both the moral and non-moral domains, the conditioning role of experience can be difficult to disentangle from the other ways in which experience contributes to knowledge. In order to isolate the conditioning role, I'll first describe a relatively clear-cut non-moral case. I'll then argue that the same mechanism can contribute to moral knowledge.

C4.P98 Consider the game of baseball. Some baseball managers and scouts have an unusual skill: they are able to make relatively precise estimates about the velocity with which a pitcher is throwing the baseball. Because the effectiveness of a pitcher can vary greatly with relatively small differences in velocity, the ability to recognize that (for example) a pitcher who is usually capable of throwing a fastball at 95 miles per hour is now throwing in the lower 90s can make a crucial difference with respect to questions about which strategy to

employ. This ability is not, of course, something that those who have it are born with, nor is there some specific technique that they have learned to apply. Rather, acquiring the skill typically involves, not only having observed a great deal of baseball, but also having had the opportunity to calibrate one's own estimates of velocity against the readings of a radar gun. By comparing one's initial estimates against the readings of the radar gun, one receives feedback about their accuracy, and this provides an opportunity to make suitable adjustments. In this way, one can over time increasingly refine one's judgment for accuracy and attain a greater degree of precision.

C4.P99 Suppose that someone who has acquired the skill in this way correctly recognizes that a pitch travelled to home plate at a velocity between 90 and 93 miles per hour. I assume that, given that she is generally reliable when it comes to making such judgments, and her arriving at a true belief on this particular occasion is a matter of his deploying the competence or skill in judging that explains her reliability, the belief she arrives at is something that she *knows*, at least if circumstances are otherwise favorable for the acquisition of knowledge. If she were insufficiently reliable at making such judgments, then this belief would not count as knowledge even if it were true. (Consider another observer who lacks the ability and whose estimates of the velocity are almost invariably mistaken by a significant margin. Even if on this occasion he luckily arrives on the same estimate of the pitch's velocity, this is not something that he knows, despite the fact that he arrived at his estimate in response to observing the same pitch—and even if he observed that pitch from the same distance and vantage point, etc.). The fact that she is now in a position to recognize that the pitch travelled at a certain velocity is due to her having had the relevant past experience. However, it is *not* a matter of her past experience having provided her with evidence that she exploits in arriving at her judgment. After all, it is not as though she has somehow retained memories of particular past pitches of known velocities that she can now call up and compare in her mind's eye with this most recent pitch. Nor is it a matter of applying to the present case some general technique for estimating velocities that she has learned. Rather, her past experience has instilled in her a certain competence or skill in judging, a competence or skill that she can now bring to bear in the present case. Experience has honed her "eyeball" estimates to the point that such estimates are now sufficiently reliable to qualify as knowledge, provided that other conditions are met.³⁷

³⁷ Williamson (2007) argues that past experience often plays a similar role in our "imagination-based knowledge of counterfactuals." The idea of knowledge as the typical output of competence in judgment is central to the work of Ernest Sosa. See, for example, his (2007) book *A Virtue Epistemology*.

C4.P100 The account of the case that I offered assumes that if someone is sufficiently unreliable, then she doesn't know, even if she bases a true belief on grounds that would be sufficient for knowledge were they possessed by a more reliable person. Call this the *minimal reliability condition* on knowledge. It is plausible that the minimal reliability condition holds for knowledge in general, regardless of subject matter, and irrespective of whether a judgment is arrived at on the basis of evidence, or by application of a theory, or in some other way. Consider for example, a moral philosopher who arrives at her views about contested cases in applied ethics by explicitly applying a badly mistaken normative theory, one that frequently delivers mistaken verdicts about cases. Because her way of arriving at her views is so unreliable, it is natural to think that she lacks knowledge even on those relatively infrequent occasions when she does arrive at a correct view by applying the theory.

C4.P101 One can endorse the minimal reliability condition on knowledge even if one thinks that reliability is neither necessary nor sufficient for having justified beliefs about some subject matter. Moreover, one can endorse the minimal reliability condition even if one has no enthusiasm for the project of analyzing the concept of knowledge in terms of reliability, or rejects standard reliabilist accounts of knowledge.³⁸

C4.P102 The baseball example illustrates that whether someone satisfies the minimal reliability condition with respect to a case might depend in part on whether she has had suitable past experiences, where this is *not* a matter of those past experiences providing her with evidence on which she can now draw. *Prima facie*, it would seem that experience can play the same role in the moral domain. For example, recognizing that a given action would be morally right or wrong is often mediated by recognition that the action would fall under a certain thick concept (e.g. that the action would be *cruel*). But one's dexterity with such concepts is not independent of one's past experience. And quite apart from providing evidence, past experiences influence the range of cases to which one is willing to apply the concept by conditioning one's judgment.

³⁸ Among other things, standard reliabilist accounts of knowledge (see especially Goldman 1979 and 1986) are *externalist* accounts. But a theorist might accept the minimal reliability condition even if she requires that a subject base his beliefs on internally available justifying grounds in order to know. (For example, such a theorist might claim that in the baseball case, the internally available justifying grounds are the visual experience(s) of the ball's flight.) On the other hand, a theorist who accepts the minimal reliability condition is committed to the claim that appeals to the concept of reliability in epistemology are not fatally undermined by "the generality problem" (see especially Conee and Feldman 1998). But I think that this claim is safe: as Williamson (2000: 100) notes, surely historians can sensibly ask which of their sources are reliable. Of course, any theorist who is a reliabilist will presumably accept the minimal reliability condition. Shafer-Landau (2003: 272–302) defends moral reliabilism, at least concerning our "verdictive" moral beliefs; he discusses the generality problem in the context of moral epistemology on pp. 280–5.

It thus influences how reliable one is with the concept. In this way, one's past experience can determine whether one is in a position to recognize that the concept applies to a particular case, and thus, whether a correct verdict constitutes knowledge.

C4.P103 Here is an objection to the idea that experience can play this conditioning role in the moral domain. Even if in the baseball case the role played by past experience is not that of enabling, triggering, or providing evidence, there is a crucial difference between that kind of case and cases involving moral judgment. In the baseball case, there is an instrument (the radar gun) against which a person can calibrate her judgment. She can compare her initial guesses with the readings of the instrument, which she treats as an authoritative standard. By contrast, it might be claimed, there is simply no analogous possibility in the moral case. Of course, how a person applies a given moral concept such as *cruel* depends in part on past experience: that one of us but not the other is willing to apply it to a borderline case might very well depend on subtle differences in our respective histories. But crucially, when a person reaches a moral verdict about a particular case, the world typically does not provide feedback about whether she has applied the relevant moral concept correctly. There is thus no genuine opportunity for a person to calibrate her moral judgment.³⁹

C4.P104 However, we should not underestimate the extent to which feedback is available in the moral domain.⁴⁰ Here again, the fact that we are members of a moral community is significant. Just as you might treat the radar gun as providing a standard against which to measure your estimates of velocity, you can treat the verdicts of trusted others as providing data that bear on the accuracy of your moral opinions. Of course, someone who calibrates her estimates of velocity against the readings of a radar gun treats the radar gun as an authoritative source in a very strong sense. In part for reasons discussed in the last chapter, it is perhaps most natural to treat other people as moral authorities in a comparably strong sense early in life. But even once you stop treating others as moral authorities, the possibility of treating the moral opinions of others as data against which to assess your own opinions remains. The point is not just that what others believe about some issue might influence what you believe about that very issue. Rather, the point is that experience

³⁹ Daniel Jacobson presses an objection of this kind (2005: 400) in the course of his argument that the neo-Aristotelian “skill model of virtue” that he attributes to John McDowell (1979) and others is subject to limitations. The position that Jacobson criticizes is more ambitious than the one I offer here.

⁴⁰ On this possibility, see also Railton (1986).

with others can influence the capacity for moral judgment in ways that make a difference to the verdicts that you reach in future cases.

C4.P105 Of course, the possibility of treating the moral opinions of others as feedback against which to calibrate one's own judgment immediately raises questions about the quality of that feedback. When someone brings her estimates of velocity into line with the readings of a well-functioning radar gun, she can be justifiably confident that she is *improving* her judgment by doing so. But when you allow your own moral judgment to be shaped by the feedback provided by the opinions of others, how can you be confident that this is likely to constitute an improvement? Indeed, how can you know that such influence isn't making things worse?

C4.P106 While there are a number of deep issues here, none of them threatens to undermine the line of thought that is currently under consideration. I have claimed that experience can contribute to moral knowledge in a way that is not a matter of enabling, triggering, or providing confirming evidence. This is consistent with the fact that some courses of experience will not improve one's moral judgment, and indeed, are likely to impair or corrupt it. To take an extreme example: a child whose parents are devoted members of the Aryan Nation grows up in an environment in which the opinions it is most natural for her to treat as a standard are unreliable guides. She is at risk, not only of inheriting specific misguided opinions, but of ending up with a warped moral judgment in virtue of having developed that judgment against a defective standard. In the same way, someone might calibrate her estimates of velocity against an unreliable or malfunctioning radar gun, and thus end up worse off than ever. The claim that experience frequently contributes to moral knowledge by conditioning the judgment of would-be knowers is perfectly consistent with the compelling thought that the nature and quality of that experience is crucial, and with the commonplace that prolonged exposure to a morally corrupt environment is likely to have pernicious effects.

C4.P107 Note that the same is true of experience in its evidential role. After all, experience sometimes provides us with misleading evidence, and a person's having such evidence might very well preclude him from knowing the truth. This has no tendency to undermine the idea that experience often plays a role in knowledge acquisition in virtue of providing non-misleading evidence. Similarly, experiences might affect someone's moral judgment in such a way that subsequent exercises of it are *less* likely to yield knowledge. But this does not cast doubt on the idea that experience sometimes contributes to moral knowledge in virtue of conditioning judgment.

C4.P108 Of course, a skeptic about moral knowledge or moral truth would be skeptical about the claim that calibrating your moral judgments against the opinions of others could ever lead to moral knowledge downstream. From a skeptical perspective, such calibration at best serves to get the members of a community on the same page, but does not get anyone closer to the truth, or in a better position to know. However, anyone who is not generally skeptical about moral knowledge should allow that the mechanism described here can play a role in its acquisition.⁴¹

C4.P109 One reason for thinking that the kind of experiential feedback described here plays a significant role in shaping a person's moral judgment is the following: the hypothesis that it does so offers an understanding of fundamental moral disagreement that is superior to standard cognitivist alternatives. For example, suppose that you and I disagree about whether a certain action would be morally wrong. Notoriously, it seems that this disagreement might survive any improvement in our knowledge of the non-moral facts. Indeed, even if the non-moral facts are simply stipulated to be thus and so, we might still disagree. On some views, either (i) our disagreement must at least in principle be rationally resolvable through a priori reflection (e.g. at least one of us must have failed to adequately "unpack" our shared concept), or else (ii) the words "morally wrong" in my mouth must express at least a subtly different concept from the same words in yours, and we are actually talking past one another. However, in some cases, neither of these options will seem attractive as a diagnosis of what is taking place.

C4.P110 Suppose our disagreement takes place against a relatively extensive background of agreement about which actions are correctly described as "morally wrong." In that case, the suggestion that we actually express different concepts by the words and are talking past one another will seem implausible. For surely two people can share a concept even if they are not in complete agreement about its extension. But it doesn't seem right that we should therefore conclude that one of us simply needs to reflect harder on the concept that we share, in order to see that the other is actually correct about whether it applies to the case at hand. Just as disagreement might persist even if there is no

⁴¹ In this connection, it is worth noting that one need not be a moral realist in order to accept the account on offer. For example, according to Bernard Williams (1985), moral truths, unlike the truths that the natural sciences aim to discover, are not part of the "absolute conception of the world". Nevertheless, members of certain moral communities have a kind of ethical knowledge. This knowledge consists in correctly recognizing that the thick ethical concepts that have currency among members of the community apply in particular cases. Someone who holds a picture of this kind can and should acknowledge that experience contributes to moral knowledge in the way described here.

disagreement about the non-moral facts, so too it might persist even if two intelligent and good willed people reflect on their concepts as best they can. Of course, a theorist might simply insist that at least one of us must be guilty of having engaged in defective deliberation. But it is hard to see why we should agree that this *must* be the case.

C4.P111 If experience conditions moral judgment in the way that I have explained, then the fact that we reach different verdicts might be due to subtle differences in the way that our faculties of judgment have been conditioned by our experiences. If that is the explanation, then it is unsurprising that there is no further piece of non-moral information that could be supplied to us now that would resolve our disagreement. Differences in our past experiences explain our current disagreement not because they supplied us with different evidence upon which we base our verdicts (something that could presumably be overcome by our acquiring further evidence, or by simply stipulating the non-moral facts). Rather, experience shapes the way in which we respond to the non-moral information that we share.⁴² Similarly, given the role that experience plays in explaining why we apply the concept in the way that we do, there is no reason to think that further rational reflection on that concept would result in convergence.

C4.S10

4.5 Armchair Moral Knowledge

C4.P112 Consider the claim that genocide is morally wrong. This claim is widely believed, and it is generally taken to be obviously true by those who believe it. Nevertheless, we should not underestimate the role played by contingent historical factors in its being widely believed. Indeed, it is possible that relatively few people so much as had the concept of *genocide* until quite recently,⁴³ and that the occurrence of actual historical events played a crucial role in bringing it about that the relevant proposition is now widely believed. As emphasized in section 4.2, however, even if a given person would not currently believe

⁴² Compare: if you and I carefully observe the same pitch your visual impression might be more or less qualitatively indiscernible from my visual impression. Everything that is currently available to us, for purposes of estimating the velocity of that pitch, might be exactly the same. Nevertheless, we might very well make different estimates, where the fact that we do so reflects differences in past experience. On the significance of the possibility of a lack of convergence among moral inquirers who are fully informed about the non-moral facts, see my “Moral Realism Without Convergence.”

⁴³ The term “genocide” was not coined until the 1940s. Although presumably at least some people had the concept well before then, it is plausible that the coining of the term, and its subsequent popularization, played a crucial role in making the thought that *genocide is morally wrong* available to many of those who currently believe it.

that genocide is morally wrong in the absence of this kind of history, that is perfectly consistent with her having a priori knowledge that the proposition is true. And it does seem as though this proposition, like the proposition that $2+2=4$, is something that most normal adults are in a position to know from the armchair. How should we understand such knowledge?

C4.P113 Perhaps some moral claims can be known a priori by reasoning from other moral claims that are already known a priori, in accordance with the usual rules of valid deduction. This raises the question of how we manage to have a priori knowledge of the ultimate premises employed in the deduction. The crucial question is: how should we understand *foundational* or *basic* a priori moral knowledge? In this section, I will sketch a way of thinking about such knowledge that builds upon some existing accounts, in a way that avoids some of their weaknesses.

C4.P114 On one popular view, the key notion for understanding foundational a priori knowledge is that of an *intellectual seeming* or *intellectual appearance*.⁴⁴ Imagine someone contemplating the proposition that *genocide is morally wrong*, or that *every number has a successor*. The proposition appears to her to be obviously true, and she feels herself drawn or compelled to assent to it. Lacking any reason to dismiss this appearance of truth as misleading, she takes up the relevant belief. Her belief is justified by the intellectual appearance that gives rise to it. Given that the belief is true and circumstances are normal in other respects, it counts as knowledge.

C4.P115 According to the view in question, just as a perceptual appearance as of p (e.g. a particular stick's looking bent) provides prima facie justification for a perceptual belief or judgment with matching content (that the stick is bent), so too a purely intellectual appearance as of q provides prima facie justification for believing or judging that q (e.g. that every number has a successor). As the terminology of "prima facie" suggests, in both cases the justification provided is defeasible. For example, in the perceptual case, if the perceiver can see that the bent stick is partially submerged in water, and she is familiar with the illusion, then she is not justified in believing or judging that the stick is bent. (And if the perceiver nevertheless believes that the stick is bent on the basis of the way the stick appears, her belief is not knowledge even if it is true.) The same holds in the intellectual case: even if Frege's Unrestricted

⁴⁴ On intellectual seemings, see especially Huemer 2005; Bealer 1998, 2002; Pust 2000; Brogaard 2014; and Chudnoff 2011a, 2011b. In the discussion that follows, I present an epistemological view of a familiar general type without attempting to follow any particular philosopher's specific formulation of that view in detail.

Comprehension Axiom seems true upon contemplation, you are not justified in believing it if you know that it leads to contradiction.⁴⁵

C4.P116 Are there limits to the kinds of propositions that can appear as the contents of intellectual seemings? In Chapter 2, I argued that, in assessing whether *considered judgments* (in the semi-technical sense of that term) are good candidates to play the theoretical roles assigned to them, we should not restrict our attention to cases involving considered judgments that we share, or are plausible by our lights.⁴⁶ I argued that we should also consider cases involving considered judgments that are by our lights perverse or absurd, because there is nothing in the notion of a considered judgment (as it is standardly explicated) that would preclude such judgments from counting as “considered” for a given person.

C4.P117 I suggest that we follow a similar procedure here, with respect to *intellectual seemings*. On the face of it, there does not seem to be anything that would preclude an eccentric subject from having intellectual seemings of more or less any proposition that she can understand—including empirical propositions, and/or propositions that we take to be absurd. Although it would be surprising to discover that someone had an intellectual seeming that *there are an even number of stars in the universe*, or that *some numbers have no successor*, there doesn’t seem to be anything impossible about this. Compare: although it would be surprising to find that someone *believed* either of these propositions, there doesn’t seem to be any impossibility here.⁴⁷

C4.P118 Imagine a psychologically eccentric person to whom the proposition *we are morally required to occasionally kill randomly* appears obvious whenever he contemplates it, as obvious as its negation appears to us. Would he be as well justified in believing this proposition as we are in believing its negation, or in believing that genocide is wrong, at least as far as the armchair is concerned?

C4.P119 For a proponent of the intellectual seemings view, there are two ways to go here. She could allow that, at least in principle, a person could have an intellectual seeming of the absurd proposition, and lack any reason to distrust how things seem, and thus be a priori justified in believing that we are

⁴⁵ Another parallel between the intellectual and perceptual cases: just as there need not be any noticeable temporal gap between the having of a perceptual experience and the formation of a perceptual judgment with matching content, so too there need not be any noticeable temporal gap between (for example) its seeming true that genocide is morally wrong and the judgment that it is.

⁴⁶ See the discussion of this point in section 2.4.

⁴⁷ Example: when I was pregnant with my first child, my mother found that the proposition that *Sarah's first child will be a girl* had for her the overwhelming appearance of truth whenever she contemplated it. Although I take it that my mother had good reason to distrust this appearance, she nevertheless came to believe—falsely, as it turned out—that my first child would be a girl on its basis.

morally required to occasionally kill randomly. As far as the armchair is concerned, this belief would be as well justified as our belief that genocide is morally wrong.

C4.P120 But that suggestion seems implausible. The picture of a priori justification that leads to it seems to overestimate the epistemic efficacy of intellectual seemings, and the capacity of such states to bestow justification on any belief, regardless of its content.

C4.P121 As support for this verdict, we might briefly compare this unrestricted version of the intellectual seemings view to another view in epistemology that is widely regarded as making justified belief too easy to come by—and which would be regarded in this way by a proponent of the intellectual seemings view herself. Doxastic conservatism (Harman 1986b) is the view that any *belief* that you find yourself with is *prima facie* justified, i.e. justified in the absence of your having reason to think that it's false.⁴⁸ Doxastic conservatism is often thought to be subject to the following kind of counterexample⁴⁹ Right now I have overwhelming evidence that I do not have the world's greatest singing voice. If I nevertheless started to believe that I do have the world's greatest singing voice in the face of this overwhelming evidence, then my belief would be unjustified, on all accounts. However, according to the doxastic conservative, if I managed to forget that evidence while retaining my belief, then my belief that I have the world's greatest singing voice would be fully justified, despite the fact that I've never had any reason at all to believe it.

C4.P122 That assessment of my belief seems too generous. Of course, a proponent of the intellectual seemings view can agree. On her view, belief, in the absence of counterevidence, is not enough. (After all, if belief, in the absence of counterevidence, were sufficient for justified believing, then there would be no need to appeal to the apparatus of intellectual seemings.)

C4.P123 But now suppose that we add to the story. It's not simply that I *believe* that I have the world's greatest singing voice, I also have an *intellectual seeming* with the same content! It is unclear why simply adding in another propositional attitude—"Wait, it's not just that I *think* so—it also *seems true to me!*"—should

⁴⁸ The terminology of "doxastic conservatism" is not due to Harman, but is often used in the recent literature to distinguish the view in question from Huemer's (2001, 2005, 2007) "phenomenal conservatism." While the phenomenal conservative holds that *appearances* (both perceptual and purely intellectual) are, in effect, innocent until proven guilty—in the absence of any positive reason to distrust the appearances, one is justified in taking them at face value—the doxastic conservative holds that the same is true of *beliefs*.

⁴⁹ See *inter alia* Christensen (1994), Feldman (1989), and Kelly (2016). The example that follows is adapted from an example presented in the last of these works.

make anybody reverse her verdict about the case.⁵⁰ And there is nothing special about this case: in general, if a person is unjustified in believing that p is true given her circumstances, then simply adding to the case that it also *intellectually seems to her* that p will not be enough to render her justified.

C4.P124 Consider again the absurd moral proposition that *we are morally required to occasionally kill randomly*. Again, on one way of spelling out the intellectual seemings view, a person might stand in the relevant relation to this proposition while lacking reason to distrust how things seem to her, and thus be fully justified in believing it, as justified as we are in believing any moral platitude (at least as far as the armchair is concerned). I've argued that this view is too liberal: it makes justification too easy to come by. Suppose instead that the proponent of the intellectual seemings view denies that a person could become fully justified in believing an absurd moral proposition on the basis of an intellectual seeming. On this way of spelling out the view, there is an asymmetry (beyond the difference in truth value) between propositions such as *genocide is morally wrong* and propositions such as *we are morally required to occasionally kill randomly*: while a subject who enjoys an undefeated intellectual seeming that the former is true is fully justified in believing it, there is simply no analogous possibility with respect to the latter proposition. I think this is the more plausible line to take about intellectual seemings with absurd contents.⁵¹

C4.P125 However, if enjoying an intellectual appearance can justify belief in some contents, but not others, then the notion of an intellectual appearance is not really the crucial notion in the explanation of basic a priori moral knowledge. Rather, what seems crucial is some feature of the contents themselves. Now in fact, the historically dominant form of ethical intuitionism—one which has enjoyed a recent resurgence of interest—takes the crucial notion to be that of a *self-evident proposition*.⁵² The basic idea is that a self-evident proposition is one that is capable of being justifiably believed, and known, on the basis of adequate understanding. For example, according to the most thoroughly

⁵⁰ In evaluating this case, it's crucial to bear in mind that *it's seeming to me that I have an exceptionally good singing voice* need not be prompted or inspired by my actual *singing*. Rather, given that what is at issue is an *intellectual* seeming (as opposed to a perceptual seeming), and its capacity to render my belief justified, what we should picture is a case in which the relevant proposition seems true to me when I consider it from the armchair, on the basis of a priori reflection.

⁵¹ Moreover, notice that this line of response seems well-motivated independently of anything having to do with morality in particular. It is similarly implausible that one could become fully justified in believing that $1+1=5$, or that some explicit contradiction is true, on the basis of an undefeated intellectual seeming.

⁵² On self-evident propositions, see especially Audi 2005 (ch. 2) and forthcoming, Hurka 2015 (esp. ch. 5), and the essays collected in Stratton-Lake (2002).

developed and epistemologically sophisticated contemporary account of this kind (Audi forthcoming), self-evident propositions are truths that meet two further conditions: (i) in virtue of adequately understanding them, one has justification for believing them, and (ii) believing them on the basis of adequately understanding them entails knowing them (p. x).⁵³

C4.P126 On this view, given that only true propositions are self-evident, there is clearly no possibility that a person could end up justifiably believing a moral absurdity in virtue of its being self-evident. But although all self-evident propositions are true, not all true moral propositions are self-evident. Indeed, the self-evident moral propositions are taken to comprise a special class even among those moral propositions that would be recognized as obviously true by ordinary people in normal circumstances. For example, even if the proposition that it is wrong to encourage children to smoke is obvious to ordinary people in normal circumstances, it is not a good candidate for a self-evident proposition. For a person might *understand* the claim perfectly and yet not be justified in believing it (as would be the case, for example, if the person lacked any reason to believe that smoking tends to have pernicious as opposed to beneficial effects on the health of children).

C4.P127 Compared to a view on which the notion of an *intellectual seeming* is central, a view that appeals to self-evidence enjoys at least one advantage and one disadvantage. The advantage is that the view does not threaten to entail that an eccentric person's absurd moral convictions are on an epistemic par (as far as a priori justification is concerned) with our beliefs in moral platitudes. The disadvantage is that states of understanding, unlike intellectual seemings, do not present their contents as being true. One of the things that makes it at least somewhat plausible that intellectual seemings can bestow *prima facie* justification on beliefs is that they have this feature; and this is a feature that is characteristic of paradigmatic states or events that justify beliefs. For example, perceptual experiences with propositional content (such as the visual experience as of the stick's being bent) present the world as being a certain way, and it is widely thought that this feature is crucial in explaining why perceptual experiences have the capacity to justify beliefs. Similarly, if my knowledge that p is based on your testimony that p, the fact that your assertion justifies my belief seems to depend on the fact that your assertion presents the world as though p is true. Assertions, perceptual states, and intellectual seemings have a certain *valence*: they are not neutral with respect to their propositional

⁵³ Compare Hurka (2014: 111): "The main intuitionist claim should be that if you understand a self-evident proposition, you are justified in believing it."

146 MORAL KNOWLEDGE

contents, but represent the world (whether correctly or incorrectly) as though those contents are true of it. If we think of the notion of “evidence” broadly, as including anything that tends to justify belief, then it is exactly this feature that makes it plausible to count assertions, perceptual states, and intellectual seemings as evidence, because evidence is similarly valenced: evidence that *p* speaks in favor of *p*.

C4.P128 Contrast the view according to which we know self-evident propositions on the basis of adequately understanding them. On the face of it, a state of understanding—even where what is involved is understanding to a high degree—is *not* a valenced state. It does not present the content that is its object as true as opposed to false, or vice versa. In this respect, it differs from intellectual seemings, perceptual states, experiences with propositional content, assertions, and from beliefs and judgments themselves. States of understanding do not *point*: as a psychological matter, a person’s coming to understand a particular proposition might inspire any number of reactions as to its truth or falsity. If, upon coming to fully understand an allegedly self-evident truth, a person found herself with no inclination to believe it, or even with a strong inclination to *disbelieve* it, it doesn’t seem plausible that she would still be justified in believing it, simply because of the *kind* of proposition that it is.

C4.P129 Let’s take stock. The intellectual seemings view locates the source of a priori knowledge in a state that seems to be at least the right *sort* of state to ground knowledge and justified belief. But in its most straightforward form, it threatens to allow belief in moral absurdities to enjoy a level of a priori justification that is on a par with that enjoyed by belief in moral platitudes. The self-evident propositions view avoids that problem, but it fails to locate the source of a priori knowledge in the right sort of state to ground knowledge and justified belief. Given this dialectical situation, a natural strategy is to try to arrive at an improved account by combining the two views in some way.

C4.P130 Let’s retain the idea that there is an epistemologically special class of propositions among the truths that can be known in a way that other truths cannot. Let’s use the term “self-evident” as a kind of placeholder, to talk about members of that class, while postponing for now the issue of how that notion should be understood. The most straightforward way of combining this idea with the apparatus of intellectual seemings would be to endorse the following principle:

C4.P131 If *p* is a self-evident proposition, and it intellectually seems to *S* that *p*, then *S* is *prima facie* justified in believing *p*.

C4.P132 However, this principle allows for intellectual seemings to provide prima justification for self-evident propositions even when there is absolutely no connection between the proposition's seeming true to the subject and its being self-evident. As an extreme example, consider a psychologically eccentric person who is such that, whenever she considers and fully understands *any* proposition (even an absurd or clearly empirical proposition) she has an intellectual seeming that it is true. Given that this is the case for any proposition, it will obviously be the case for any self-evident proposition. However, it is clear that this would not be enough to ground knowledge or justified belief in a self-evident proposition, for the fact that such a proposition seems true to her has nothing to do with the fact that it is self-evident, or even that it is true. In terms of what she has to go on, she stands in the same relationship to the proposition that we are morally required to occasionally kill randomly, or that there are an even number of stars in the universe.

C4.P133 Here is my suggestion for an improved account. Consider first a non-moral case: upon considering the proposition that $2+2=4$, a child is struck by the sense that it seems clearly true; as a result, she comes to believe it. In what circumstances would we count her belief as knowledge? A natural suggestion is that the child knows *p* just in case *p* intellectually seems true to her because *it is* true. The child knows if there is a (non-deviant) explanatory connection between the truth and the appearance of truth.⁵⁴ Extending this natural

⁵⁴ The requirement that the explanatory connection be *non-deviant* is motivated by cases such as the following. Imagine that a manipulative but epistemically benevolent demon interferes with the child's psychology so that whenever the child considers a true proposition, it intellectually seems true to her. Suppose further that the proposition that $2+2=4$ would not seem true to the child in the absence of the demon's interference, and she would not believe it in that case. In those circumstances, there is a sense in which the fact that the proposition seems true to the child is explained by its truth (the demon would not have interfered for a proposition that lacked that property), but it is at least arguable that the child should not be credited with knowing that $2+2=4$, even though she believes it in response to its seeming obviously true to her.

In short, without the restriction to non-deviant explanatory connections, one might worry that the account on offer is vulnerable to potential counterexamples analogous to those that historically bedeviled classic causal theories of knowledge, for example, Goldman 1967. In what follows, "explanatory connection" can be read as "non-deviant explanatory connection" by those with such concerns. I will not attempt to spell out what makes something a "deviant" as opposed to a "non-deviant" explanatory connection in the relevant sense. Is this objectionable? It would be if the account on offer purported to be a reductive analysis of armchair knowledge, in the way that causal theories of knowledge were historically regarded as attempts to provide a reductive analysis of knowledge. However, I want to explicitly disavow any such ambition: as noted, my aim here is to sketch a way of thinking about armchair moral knowledge that makes it at least plausible that we have some such knowledge (one that improves on existing accounts in various ways), as opposed to providing a reductive analysis of such knowledge in terms of non-circular necessary and sufficient conditions that appeal to more fundamental notions. Among other differences, successful execution of this more modest project is consistent with the possibility that there *isn't* any way to characterize "deviant explanatory connections," apart from "explanatory connections that are too indirect or weak to ground knowledge," or something similar.

account to the moral case: upon first understanding the claim that, e.g. genocide is morally wrong, a person might take up the belief because it seems obviously true to her. In these circumstances, her belief counts as knowledge if it intellectually seems true to her because it *is* true. In contrast, even if the proposition *there is an even number of stars in the universe* seems true to her when she considers it, and even if it *is* true, there is no possibility that she knows the proposition on the basis of the intellectual seeming, because the fact that it seems true to her is presumably not explained by the fact that there are an even number of stars in the universe. We are willing to credit a subject with knowledge on the basis of how things intellectually seem to her when the way that things seem to her is explained by their actually being that way.

C4.P134 Next, here is a modest proposal for how to demarcate the class of “self-evident” propositions. The self-evident propositions are simply those propositions—if any—whose intellectually seeming true to a subject can be correctly explained by their being true. Propositions such as $1+1=2$ or *genocide is morally wrong* are among the strongest candidates for being “self-evident” in an epistemologically interesting sense. On the current proposal, they count as self-evident because their truth can explain why they (intellectually) seem true. Of course, this is consistent with the possibility that, in a given case, a subject will believe one of these propositions on the basis of an intellectual seeming even though the seeming is *not* explained by the truth of the proposition that seems true. On the current proposal, the subject’s belief does not count as knowledge in those circumstances. But the proposition itself still counts as self-evident, since (unlike many other true propositions), it is the kind of proposition whose truth can account for its intellectually seeming true in other cases.

C4.P135 Plausibly, a given proposition’s being true might explain why it seems true to a relatively sophisticated subject even if there is no possibility of a less sophisticated subject grasping its truth in the same way. On one way of precisifying the current proposal, this introduces a kind of subject-relativity into the characterization of a “self-evident proposition”: the same proposition might count as self-evident to one subject but not as self-evident to another. On an alternative precisification, the proposition counts as self-evident *tout court*, in virtue of the fact that some subjects are such that its seeming true to them can be explained by its being true.

C4.P136 Of course, some philosophers would deny that a moral claim’s being true could *ever* figure in the actual explanation of why it seems true. For example, Harman (1977, ch. 1) famously denies that there is ever any need to appeal to moral facts or moral truths in explaining psychological facts about why we

make the moral judgments that we make; he would presumably say the same thing about explaining psychological facts about why certain moral claims seem true to us.

C4.P137 On some accounts, the feature of moral claims that disqualifies them for entering into such explanations is a feature that they share with mathematical claims, for example, the alleged *causal inefficacy* of moral and mathematical properties.⁵⁵ Such accounts imply, implausibly, that the correct explanation of why $1+1=2$ seems true to us is neutral with respect to the question of whether $1+1=2$ is true.

C4.P138 But there are, I think, reasons to doubt that a moral truth could figure in the explanation of its seeming true that do not extend to the mathematical case. For example, at least at first glance, the proposition *slavery is unjust* seems like as natural a candidate for being a self-evident proposition in the moral domain as $1+1=2$ is in the mathematical domain.⁵⁶ However, although the proposition that slavery is unjust now seems obviously true to many people, there is significant historical evidence that this is a relatively recent phenomenon, and that in fact, the same proposition would have *seemed obviously false* to countless human beings who thoroughly understood it.⁵⁷ If so, this might increase the credibility of the hypothesis that whether the proposition seems true or false to a given person can be thoroughly explained in terms of her historically and socially conditioned moral sensibility, in a way that does not entail that the proposition is true (or false). By contrast, it is unlikely that the proposition that $1+1=2$ ever seemed false to any significant number of people who perfectly understood it.⁵⁸

C4.P139 Of course, even if there is a potentially significant difference here, this does not show that the truth of a moral proposition *never* figures in the correct explanation of why it seems true to a given person. Indeed, that claim is extremely strong and speculative. But suppose we did have adequate reason to believe it. (Imagine that an oracle informs us that whenever a moral proposition seems true to us—even obviously true—the full explanation of why it seems that way is completely neutral with respect to whether the proposition is true or false.) In that case, I think, the reasonable conclusion for us to draw is that we don't have any *a priori* moral knowledge after all. (Or at least, that we have no reason to think that we have any such knowledge.) But that

⁵⁵ A clear example is Dworkin (2011). I discuss Dworkin's views in detail in the next chapter.

⁵⁶ As Abraham Lincoln once said: "If slavery is not wrong, then nothing is wrong."

⁵⁷ For relevant evidence, see Sowell (1994), and the further references cited there.

⁵⁸ Note that even those few philosophers who claim that $1+1=2$ is false (e.g. Field 1980) do so for theoretical reasons, reasons that they take to trump the seeming truth of the claim itself.

150 MORAL KNOWLEDGE

possibility is no objection to what I have said. On the contrary, certain possible empirical discoveries would make it rational to conclude that we don't have a priori moral knowledge. No account of what would be involved in having such knowledge should be inconsistent with that fact.

- C4.P140 Indeed, the point holds more generally: in principle, there are things that we could learn about the psychological processes that lead to our moral views that would undermine our claims to moral knowledge. Certain possible empirical discoveries would make it irrational for us to persist in thinking that we have any. This is one of the ideas that I explore in the next chapter, "Losing Moral Knowledge."

C5

5

Losing Moral Knowledge

C5.S1

5.1 Introduction

C5.P1 The discussion thus far might seem unrelentingly optimistic in its assessment of the range of circumstances in which moral knowledge is available to us, and the ways in which we can acquire such knowledge. In Chapter 2, I argued that although critically examining our moral views and attempting to make them more coherent is a potentially fruitful way of expanding our knowledge of morality, neither of these activities is a necessary condition for having full-fledged moral knowledge in the first place. Indeed, if the Moral Inheritance View defended in Chapter 3 is correct, then a person whose earliest moral views are inherited from her social environment might very well have substantial moral knowledge even before she is in a position to begin critically reflecting upon or reasoning about those views. More generally, other people are in principle potentially rich sources of moral knowledge. To the extent that we have reservations about the propriety of forming moral views by relying on others as opposed to through exercising our own autonomous judgment, what is legitimate in those reservations does not derive from its being impossible to acquire moral knowledge in this way. In the last chapter, I attempted to make plausible the idea that experience and observation can contribute to moral knowledge in any of the various ways in which they contribute to ordinary empirical knowledge, and I sketched an account of how some moral knowledge might be available from the armchair.

C5.P2 In contrast, this chapter concerns some of the ways in which it is possible to be overly optimistic about our cognitive relationship to morality, not only in the obvious sense of overestimating how far our moral knowledge extends, but also by underestimating the fragility of such knowledge and the contingency of the circumstances in which it is available to us. If the working hypothesis endorsed in Chapter 1 is correct, then (1) any source of ordinary empirical knowledge is also a potential source of moral knowledge, but also, (2) our efforts to acquire and maintain moral knowledge are susceptible to frustration in all of the ways that our efforts to acquire and maintain ordinary empirical knowledge are. While the arguments offered in the previous three

chapters are intended to support the first conjunct of the working hypothesis, the arguments offered in this chapter are intended to bolster the second, by exploring and defending the possibility of losing moral knowledge in ways that others have denied are genuinely possible.

C5.P3 Let's begin by distinguishing two general ways in which non-moral knowledge can be lost. First, there is the phenomenon of *defeat*: one might cease to know a proposition in virtue of acquiring evidence or information in the light of which one is no longer justified in believing it. In that case, even if one continues to believe the proposition, one does not know it. Secondly, one might lose knowledge simply by forgetting what one once knew. Interestingly, although it is generally agreed that non-moral knowledge can be lost in either of these ways, in each case it has been argued, by philosophers of the first rank, that our knowledge of morality is in an important respect different.

Here I'll explore these issues, beginning with the case of forgetting.

C5.S2

5.2 Moral Forgetting and Ryle's Puzzle

C5.P4 Gilbert Ryle's classic essay "On Forgetting the Difference Between Right and Wrong" is an extended reflection on what he sees as a deep puzzle about our cognitive relationship to morality. The issues raised by Ryle are interesting in their own right, but they also intersect with some of the main themes and arguments of this book. For this reason, I want to look at them in some detail.

C5.P5 Here is Ryle's official statement of the puzzle:

C5.P6 '*Don't you know the difference between right and wrong?*' '*Well, I did learn it once, but I have forgotten it.*' This is a ridiculous thing to say. But why is it ridiculous? We forget lots of things, including lots of important things, that we used to know. So what is the absurdity in the idea of a person's forgetting

C5.P7 the difference between right and wrong? (1958: 394)

C5.P8 Ryle's own view is that talk of "forgetting the difference between right and wrong" is absurd in a very strong sense. It's not simply that people rarely, if ever, forget the difference in practice, in the way that people rarely if ever forget their own names (p. 395). Rather, to speak about forgetting the difference between right and wrong betrays a kind of category mistake: the word "forget" will simply "not go with" the words "the difference between right and wrong."¹

¹ It might be thought that this last claim is clearly an overstatement, because there are at least some contexts in which talk of forgetting the difference between right and wrong seems perfectly in order.

C5.P9 Ryle assumes throughout his discussion that many people do know the difference between right and wrong, including any “properly brought up child” (p. 394). In this respect, knowledge of the difference between right and wrong is “common knowledge,” which Ryle contrasts with the kind of knowledge that is had by an expert (p. 402). Nevertheless, it is not universally shared: for example, someone who is completely indifferent as to whether he breaks or keeps his promises is not someone who can be accurately credited with it (p. 400). For Ryle, the fact that it would be absurd or ridiculous for someone to *claim* to have forgotten the difference between right and wrong is due to the fact that knowledge of the difference is simply not the kind of thing that can be literally forgotten: although not everyone knows the difference between right and wrong, someone who does know the difference will not lose that knowledge by forgetting it.² Although Ryle thinks that many people do know the difference between right and wrong, he regards it as extremely unclear what such knowledge amounts to. His essay is thus largely devoted to exploring what kind of thing such knowledge could be, given that it is not susceptible to being lost by being forgotten, in the way that other kinds of knowledge are.

C5.P10 If Ryle were correct in holding that knowledge of the difference between right and wrong is in this respect fundamentally unlike other types of knowledge, then that would undermine the working hypothesis endorsed in Chapter 1. In defense of that hypothesis, I will argue that what Ryle treats as a datum to be explained is actually false: in principle, there is nothing

Here is one such context. Two friends are discussing whether one of them should pursue a proposed extramarital affair. The person who is contemplating the affair lists a number of considerations that speak either for or against it, none of which is a distinctively moral consideration. The friend, who believes that engaging in the affair would be morally wrong, listens silently for a time and then interjects: “But don’t forget the difference between right and wrong!”

However, I doubt that the kind of “forgetting” at issue in examples such as this one is literal forgetting in the sense with which Ryle is concerned. In urging “Don’t forget the difference between right and wrong!”, the friend who is acting in the advisory capacity is evincing concern that moral considerations will be given insufficient weight, perhaps even ignored entirely, in favor of considerations of other kinds. It’s unlikely that the concern is that the person who is contemplating the affair has literally forgotten the difference between right and wrong, in anything like the way she might have forgotten some piece of factual knowledge that she once knew or forgotten how to speak a language that she used to speak. But the latter seems to be what Ryle has in mind. (Compare: “Don’t you know when the American Civil War occurred?” or “Don’t you know how to speak German?”. With respect to either of these questions, the response “I used to know, but I’ve forgotten” seems like a perfectly good one.) Indeed, one might think the injunction “Don’t forget the difference between right and wrong!” presupposes that the person to whom it is addressed retains his or her grip on the difference between right and wrong; what is being called for is that he or she apply this knowledge to the case at hand.

For more on the distinction between literal and non-literal uses of “forget”, see section 5.5.

² As Ryle notes, a similar claim is endorsed by Aristotle, in brief remarks at 1100b 12 and 1140b 29 of the *Nicomachean Ethics*. There, Aristotle claims that it is characteristic of practical wisdom (*phronesis*) that it is not forgotten by those who possess it. The Aristotelian thesis is endorsed by Aquinas in the *Summa Theologica*; see 2a2ae47.16.

impossible about forgetting the difference between right and wrong, in a quite literal sense.³ Indeed, I'll argue that some cases of corruption—in which a person passes from a better moral sensibility to one that is worse—are best understood as involving such forgetting.

C5.P11 However, simply providing a counterexample to Ryle's impossibility claim would not amount to a satisfactory treatment of the issues raised by his discussion, for two reasons. First, what it is to “know the difference between right and wrong”—as opposed to, say, having propositional knowledge of some moral claim or collection of moral claims—is far from transparent. Because of this, it's far from clear what the contested possibility of forgetting the difference would actually involve. Thus, some unpacking of these notions is in order. Second, although on my view Ryle is mistaken in thinking that there is some deep impossibility here, I share his sense that it would be simply bizarre for someone to sincerely assert: “I used to know the difference between right and wrong, but I don't anymore; at some point, I forgot it.” Thus, what I seek is not simply a vindication of the possibility that Ryle denies is genuine, but a vindication that both leaves intact and accounts for the sense of puzzlement that we would no doubt experience if we were ever actually privy to the kind of exchange that Ryle recounts in his official statement of the puzzle.⁴

C5.S3

5.3 Expertise, Competence, and Knowing the Difference Between Right and Wrong

C5.P12 There are significant connections between Ryle's puzzle and the puzzle about full-blooded moral expertise explored in Chapter 3. Indeed, it's tempting to view Ryle's puzzle as a *special case* of the puzzle about full-blooded moral expertise. For it seems that a person who sincerely claimed that she once knew, but has since forgotten, the difference between right and wrong, would

³ Of course, this claim is consistent with the possibility (which I take to be actual) that such forgetting is a relatively rare occurrence. But even if it's true that knowledge of the difference between right and wrong is seldom if ever forgotten in practice, that wouldn't distinguish it from much ordinary empirical knowledge. What is threatening to the working hypothesis is Ryle's idea that there is a theoretically significant difference in kind here.

⁴ In this respect, I see the dialectical situation as parallel to the one that prevails with respect to the similarly contested possibility of full-blooded moral expertise. A philosopher who holds, as I do, that in principle one could arrive at substantial moral knowledge by treating another person as a full-blooded moral expert should offer some explanation of the common feeling that the practice of the deferential spouse described in Chapter 3 is *not* on all fours with a practice of, say, consistently deferring to one's spouse about driving directions. Similarly, a philosopher, who holds, as I do, that in principle one could literally forget the difference between right and wrong should provide some account of why the exchange recounted by Ryle seems to misfire.

in effect regard her past self as a full-blooded moral expert compared to her current self. (That is, it seems that she should look upon her past self as someone to whom she has a compelling reason to defer in her moral judgments, even when she finds herself inclined to judge otherwise.) When viewed in this way, the case of a person who sincerely judges that she has forgotten the difference between right and wrong stands as the diachronic, intrapersonal counterpart to the synchronic, interpersonal case in which a person sincerely judges that one of her contemporaries is a full-blooded moral expert.

C5.P13 In order to make the parallelism more vivid, imagine someone who at an earlier time takes herself to know the difference between right and wrong, but who is aware that she has an unusually bad memory. As a precaution, she makes a list of all of those actions that she currently takes to be wrong, writing them down on index cards. However, she neglects to write down the reasons why she takes these actions to be morally wrong. (Perhaps there is simply not enough room on the index cards to include this information as well.) Unfortunately, her lack of confidence in her memory proves well founded, and as time passes she begins to forget. Upon consulting the index cards, she finds herself unable to reconstruct the reasons that speak against some of the actions that appear there. Moreover, it now seems to her that at least some of those actions are not morally wrong at all. (If she simply made up her mind now, without worrying about what her past self thought, she would deny that these actions were wrong.) Nevertheless, because she conceives of the situation as one in which a failure of memory has left her with an impoverished moral outlook, she ends up morally opposed to ϕ -ing, even though it now seems to her for all the world as though ϕ -ing is perfectly permissible. In these circumstances, her moral opposition to ϕ -ing seems rather bizarre, in exactly the same way as it would if her opposition was based on blindly deferring to another person.⁵

C5.P14 Below, I'll argue that the analogy between the intrapersonal diachronic case and the interpersonal synchronic case is in fact a fruitful one, and that insights that emerged in the discussion of moral expertise in Chapter 3 can be profitably applied to Ryle's puzzle. There are, however, limitations to the analogy as well. In particular, the knowledge necessary for a person to count as

⁵ Notice that there is nothing at all bizarre about deferring to one's past self in exactly this way about any number of other subject matters. Indeed, those of us with highly imperfect memories often construct lists of exactly this sort, lists containing propositions that we want our future selves to believe, without bothering to record the reasons that we currently possess for thinking that those propositions are true. This too parallels the interpersonal case, in which deferring to other people about various non-moral subject matters seems utterly unproblematic and straightforward.

156 MORAL KNOWLEDGE

“knowing the difference between right and wrong” seems relatively *modest* compared to the kind of knowledge necessary for counting as a genuine moral expert. As Ryle notes, it is natural to credit a “well-brought up child” with knowledge of the difference between right and wrong. (Contrast the relationship that the well-brought up child bears to morality to the position of a significantly younger sibling who has not yet reached a level of development that would warrant treating her as a genuine moral agent.) Similarly, it’s natural to describe ordinary decent people as knowing the difference between right and wrong, whereas it’s not natural to describe sociopaths in the same way. But even if ordinary decent people and well-brought up children can properly be credited with knowing the difference between right and wrong, this obviously does not make them moral experts. In this respect, knowing the difference between right and wrong seems closer to moral *competence* than expertise.

C5.P15 In previous chapters, we’ve focused exclusively on propositional knowledge—for example, the knowledge that a certain action would be wrong. We should thus ask: what is the relationship between *knowing the difference between right and wrong* and having propositional knowledge that certain actions are right and certain actions are wrong? The most straightforward view about their relationship would be this:

C5.P16 (0) One knows the difference between right and wrong just in case one knows which actions are right and which actions are wrong (in a sufficiently wide range of cases).

C5.P17 However, (0) is not correct. In particular, it’s implausible to think that knowing which actions are right and which actions are wrong is *sufficient* for knowing the difference between right and wrong. Consider, for example, a young child who clearly has not yet learned the difference between right and wrong. If the child relied on his own judgment, he would be hopelessly at sea and misclassify many actions with respect to the distinction, including many cases that would strike normal adult members of his community as clear-cut. Suppose, however, that in arriving at his views about which actions are right and which actions are wrong, the child consistently defers to an authority figure whom he has good reason to trust. The authority does know the difference between right and wrong, and as a result, is generally knowledgeable about which actions are right and which are wrong. The child’s views about the moral status of specific actions are thus no less accurate than the views of the person to whom he defers. In Chapter 3, I defended the idea that moral knowledge can be transmitted via deference in the same way that non-moral knowledge can be. If that is correct, then the child can gain knowledge of

which actions are right and wrong in this way. Nevertheless, this possibility is consistent with the original assumption that the child does not yet know the difference between right and wrong. However complete his knowledge is with respect to the purely extensional question of which actions are right and which are wrong, learning the difference between right and wrong might still be something that lies in the child's future, given that he lacks the kind of competence that would allow him to arrive at accurate judgments autonomously.⁶

C5.P18 On the other hand, even though it would be a mistake to simply identify *knowing the difference between right and wrong* with being generally knowledgeable about the moral status of particular actions, it seems plausible that the latter is a necessary condition for the former:

- C5.P19 (1) One knows the difference between right and wrong only if one knows which actions are right and which actions are wrong (in a sufficiently wide range of cases).

C5.P20 The thought behind (1) is that one way of failing to count as someone who knows the difference between right and wrong is to be sufficiently ignorant about the rightness and wrongness of various actions: a person who by our lights regularly misclassified actions with respect to their moral status, or who simply professed ignorance as to their status, is not someone whom we would credit with knowing the difference. Presumably, perfect accuracy is not required: Ryle's "well-brought up child" knows the difference between right and wrong, even if she misclassifies or becomes mystified by some intricate puzzle cases. But a person who knows the difference between right and wrong will generally know which actions are right and which are wrong.⁷

⁶ The point is a general one about the circumstances in which a person can be credited with knowing the difference between x and y. Consider, for example, knowing the difference between beeches and elms. There is a difference between knowing, of those trees that one encounters, which are beeches and which are elms, and knowing the difference between beeches and elms. A person's beliefs about which trees are beeches and which are elms might be arbitrarily reliable if she consistently defers to someone who can reliably discriminate the two, even though she herself does not know the difference. Indeed, her reliability with respect to such judgments might compare favorably with the judgments of others who genuinely *do* know the difference between beeches and elms. Similarly, her access to the relevant facts might in principle be no less robust than theirs is. (Perhaps the master beech-elm discriminator to whom she defers is her conjoined twin.) In the same way, even comprehensive and robustly available propositional knowledge of which actions are right and wrong is insufficient for knowing the difference between right and wrong.

⁷ How reliable does one have to be, given that perfect reliability is not required? Presumably, this is a vague matter, which allows for borderline cases. Perhaps it is also context-sensitive—for example, perhaps in a context in which the characteristics of an Aristotelian *phronimos* are being discussed, Ryle's properly brought up child does not count as someone who knows the difference between right and wrong, but in a context in which she is being compared with her much younger sibling, she does.

158 MORAL KNOWLEDGE

C5.P21 In addition, the reasons offered above for rejecting (0) suggest a second necessary condition:

C5.P22 (2) One knows the difference between right and wrong only if one is a competent judge of which actions are right and which actions are wrong.

C5.P23 In what follows, I will assume that although (0) is false, (1) and (2) are true: the latter are adopted as working hypotheses about *knowing the difference between right and wrong*, assumptions that can lend structure to the discussion that follows. On the picture in question, knowing the difference between right and wrong involves a kind of *autonomous reliability* when it comes to making judgments about which actions are right and which are wrong. Before offering my own treatment of Ryle's puzzle, I want to examine his own solution.

C5.S4 5.4 Forgetting vs. Ceasing to Care

C5.P24 The key element of Ryle's proposal is his commitment to a broadly internalist view about moral motivation. According to Ryle:

C5.P25 There seems to be a sort of incongruity in the idea of a person's knowing the difference between good and bad poetry, while not caring a whit more for the one than for the other... Now whether [our inclination to think this] is justified or not, it exists just as much in our thinking about the knowledge of right and wrong. Here, too, there seems to be an incongruity in the idea of a person's knowing that something wrong had been done, but still not disapproving of it or being ashamed of it; of his knowing that something would be the wrong thing for him to do, but still not scrupling to do it. We hanker to say that, if he has no scruples at all in doing the thing, then he cannot know that it is wrong, but only perhaps, that it is 'wrong', i.e., what other people call 'wrong'. (Ryle 1958, p. 398)⁸

Although I think that there are compelling reasons to reject (0) and accept (1) that are independent of anything having to do with Ryle exegesis, I note in passing that these choices seem to accord with his own understanding of "knowing the difference between right and wrong." For example, he remarks that "knowing the difference between right and wrong is not *just* a competence to label correctly" (Ryle 1958, p. 401, emphasis mine)—which clearly suggests that he regards competence with respect to classification as necessary (although insufficient) for knowing the difference.

⁸ Notice that in an important respect, the view that Ryle endorses here is weaker than standard versions of motivational internalism about moral judgment. Standard versions of motivational judgment internalism entail that (to a first approximation) there is an internal connection between *judging* that ϕ -ing is wrong and having certain motivational dispositions with respect to ϕ -ing. (Anyone who completely lacked the relevant motivational dispositions would not be someone who genuinely judged that ϕ -ing is wrong; at best, such a person judges that ϕ -ing is "wrong".) In contrast, Ryle holds that

C5.P27 After rehearsing such considerations several times (cf. pp. 400, 402), he offers his own solution to the puzzle:

C5.P28 Now we can begin to see why it is ridiculous to say that one has forgotten the difference between right and wrong. To have been taught the difference is to have been brought to appreciate the difference, and this appreciation is not just a competence to label correctly or just a capacity to do things efficiently. It includes an inculcated caring, a habit of taking certain sorts of things seriously.

C5.P29 A person who used to care may, indeed, cease to care. But ceasing to care is not forgetting, any more than ceasing to believe something or to mistrust

C5.P30 someone is forgetting. (Ryle 1958, p. 401)

C5.P31 Thus, for Ryle knowing the difference between right and wrong is quite unlike, say, knowing the difference between beeches and elms. A person who can reliably and autonomously discriminate beeches and elms can rightly be credited with knowing the difference between the two, even if she finds that distinction utterly lacking in significance. In contrast, someone who knows the difference between right and wrong is not only a reliable classifier; in addition, he or she also *cares* about the difference. Such a person exhibits certain characteristic patterns of caring and concern, and is at least disposed to favor what is right over what is wrong in her actions as well as in other ways. Ryle explicitly allows that a person who once exhibited the relevant patterns of caring and concern might no longer do so at a later time. Given his view that the relevant kind of caring and concern is necessary for knowing the difference between right and wrong, Ryle would presumably

there is an internal connection between *knowing* that ϕ -ing is wrong and having certain motivational dispositions with respect to ϕ -ing. Ryle's view is thus silent (in a way that standard versions of motivational judgment internalism are not) about cases in which one makes a genuine moral judgment that does not amount to knowledge.

Of course, the view that Ryle endorses in this passage is perfectly consistent with standard versions of motivational internalism. Indeed, standard versions of motivational internalism not only entail Ryle's claim, but would also explain why it holds if they themselves are true. (Perhaps the explanation of why anyone who knows that ϕ -ing is wrong has certain motivational dispositions is that anyone who judges that ϕ -ing is wrong has those motivational dispositions, and knowing entails judging.) However, one gets the sense from the surrounding text that Ryle has at least strong sympathy for the view that genuine *knowing* has some distinctive motivational import, above and beyond that which attaches to mere judging. If so, that would make his view more akin to the (arguably) Socratic view according to which genuine knowledge of the Good is taken to have distinctive motivational import. One might think that such "knowledge-based" versions of internalism are inferior to more recent versions in this respect, on the grounds that a subject who genuinely judges that p (but fails to know that p) will have all of the same motivational dispositions as he would if he actually knew that p . But for a powerful critique of this assumption, see Williamson (2000: 60–4).

While this potential difference between Ryle's view and more familiar versions of internalism might be significant in many theoretical contexts, I don't think that it makes a difference to the discussion that follows, so I'll ignore it.

allow that it's perfectly possible for a person who once knew the difference between right and wrong to not know the difference at a later time. Thus, on Ryle's view, the solution to the puzzle is *not* that it is impossible for someone who once knew the difference between right and wrong to fail to know the difference at some later time. Rather, Ryle's view is that the process by which the knowledge is lost cannot accurately be characterized as one of *forgetting*, for ceasing to exhibit the relevant patterns of caring and concern is not a matter of forgetting anything.⁹ On Ryle's view, this accounts for why there is nothing absurd about someone's saying, "I used to know the difference between beeches and elms, but I've forgotten it," but there is something absurd about someone's saying, "I used to know the difference between right and wrong, but I've forgotten it."

C5.P32 However, on closer inspection the considerations Ryle adduces fail to deliver a solution to the puzzle. Let's grant that if I ceased to know the difference between right and wrong because I stopped caring about the relevant things, then this would not count as a case of forgetting the difference. Still, the question remains: given that a sufficient level of competence in classifying actions as right or wrong is a *necessary* condition for "knowing the difference between right and wrong"—as Ryle seems to explicitly admit—why couldn't I fall below the relevant threshold simply by losing track of the moral status of certain actions? For example, suppose that I'm initially a *marginal case* of someone who knows the difference between right and wrong, because I'm a marginal case of someone who is reliable enough when it comes to classifying actions with respect to their moral status.¹⁰ If I were any less reliable than I am, I would no longer count as someone who knows the difference. Still, I know of many actions that they are morally right, and of many actions that they are morally wrong. On the face of it, it seems like it should be perfectly possible for me to simply forget the actual moral status of at least a small handful of these actions over time, and thus fall below the relevant threshold.

⁹ Compare St. Thomas Aquinas' explanation, in the *Summa Theologica*, of the putative fact that "practical wisdom cannot be lost through forgetfulness":

Forgetfulness regards knowledge only, wherefore one can forget art and science, so as to lose them altogether, because they belong to the reason. But practical wisdom consists not in knowledge alone, but also in an act of the appetite... Hence practical wisdom is not taken away directly by forgetfulness, but rather is corrupted by the passions. (2a2ae.47.16)

¹⁰ As argued above, a person might know the difference between right and wrong even if he's an imperfect classifier of actions with respect to whether they are right or wrong; on the other hand, someone who can accurately be described as "knowing the difference between right and wrong" is presumably over some minimal threshold of reliability: a person who regularly misclassified actions with respect to this distinction would not be accurately described as knowing the difference. So there is some reliability threshold (short of perfection) satisfaction of which is a necessary condition for knowing the difference between right and wrong.

And if I stopped counting as someone who knew the difference between right and wrong in *this* way, why wouldn't this be a case of forgetting that difference?

C5.P33 Of course, if Ryle is right that knowing that ϕ -ing is wrong and caring about ϕ -ing are inseparable, any case in which I cease to know that ϕ -ing is wrong is also a case in which I cease to care about ϕ -ing in the relevant ways. But it doesn't follow from this (or from the undisputed fact that ceasing to care is not a kind of forgetting) that ceasing to know that ϕ -ing is wrong could not be a matter of forgetting that ϕ -ing is wrong. At least on the face of it, the line of thought advanced by Ryle seems to confuse necessary and sufficient conditions for forgetting the difference between right and wrong. Ryle describes one way in which someone who once knew the difference might cease to know it, and he points out that that way of ceasing to know the difference would not count as a case of forgetting. But this falls short of explaining why one couldn't forget the difference between right and wrong.

C5.P34 Although I think that Ryle's proposal is inadequate as it stands, I don't claim that the line of objection just advanced is decisive against his general approach. Perhaps there is some way of supplementing Ryle's proposal, so that it's not vulnerable to the criticism offered here. For example, one might try to argue that in any case in which there is a constitutive connection between *knowing* and *ceasing to care* (a connection which Ryle seems to think is present in this case) talk of forgetting is out of place. Rather than pursuing such possibilities, I want to offer an alternative approach to Ryle's puzzle. In the course of developing this approach, I will point out additional reasons for doubting Ryle's own solution.

C5.S5 5.5 How to Forget the Difference Between Right and Wrong

C5.P35 Consider the circumstances in which the sentence "Don't you know the difference between right and wrong?" would most naturally be uttered: a context in which one person is chastising another for some piece of alleged misbehavior. The person to whom the speech act is directed has behaved in a way that the speaker finds morally unacceptable, and the speaker is offering a rebuke. In such a context, the utterance functions not so much as a genuine question or request for information but rather as a speech act of another kind. Natural responses to it would include attempting to justify or defend one's behavior, offering an excuse, or simply admitting guilt. Ryle imagines the person to whom the utterance is addressed taking it flatly, as a straightforward request for information, and answering it in the negative. Moreover, the respondent

162 MORAL KNOWLEDGE

adds that he once knew the difference between right and wrong, but then forgot it. (So by his own account, the respondent's current lack of knowledge is not due to his never having been taught the difference, as might be the case if, for example, he had received a particularly bad upbringing.)

C5.P36 It's clear that Ryle thinks that both of the following are true:

- C5.P37 (3) There is something absurd or ridiculous about someone's asserting:
“I used to know the difference between right and wrong, but I've forgotten it.”
- C5.P38 (4) It is not possible for someone to forget the difference between right and wrong.

C5.P39 Notably, in the official statement of the puzzle with which Ryle opens his essay, the absurdity is said to attach to the performance of the *speech act* (i.e. (3)). Similarly, when Ryle introduces his own proposed solution, the focus is once again on the impropriety of the speech act: “Now we can begin to see why it is ridiculous to say that one has forgotten the difference between right and wrong...” (Ryle 1958, p. 401). Here and elsewhere, what calls out for explanation is why a certain *remark* is “paradoxical” (p. 395). But in other places, Ryle writes as though the explanandum is actually (4).¹¹

C5.P40 Given that Ryle is clearly convinced that (4) is true, it's unsurprising that he doesn't distinguish between it and (3). After all, if it's simply impossible (for anyone) to forget the difference between right and wrong, then surely there is something ridiculous about someone's saying of himself that he has done so. More generally, there is something absurd or ridiculous about asserting that something that's impossible has actually occurred or taken place. A philosopher who thinks that (4) is true thus seems committed to thinking that (3) is true.

C5.P41 However, the converse does not hold. That is, a philosopher might coherently maintain that while there is in principle nothing impossible about someone's forgetting the difference between right and wrong, there is something absurd or ridiculous about someone's sincerely asserting that she has forgotten the difference between right and wrong.

C5.P42 In fact, there are good reasons not to pass too quickly from (3) to (4). In paradigmatic cases of forgetting, one's earlier self knew something that one's later self does not know. In claiming to have forgotten something, one simultaneously attributes knowledge to one's past self and ignorance to one's current self. For example, when one declares, “I've forgotten the lecturer's name,”

¹¹ See, for example, earlier on p. 395: “It might be suggested that there is a quite simple reason why we cannot forget the difference between right and wrong.”

one simultaneously attributes knowledge of the lecturer's name to one's past self and ignorance of the lecturer's name to one's current self. However, present tense self-attributions of ignorance can be problematic or absurd, even when they are true. This is a central lesson of Moore's paradox (Moore 1962). Thus, even if I'm ignorant of the fact that it's currently raining, it would nevertheless be absurd for me to assert:

C5.P43 It's raining but I don't know that it is.¹²

C5.P44 Similarly, there would be something absurd about my sincerely asserting:

C5.P45 I used to know that the Battle of Hastings occurred in 1066, but I've forgotten that.

C5.P46 even in a context in which the proposition that I would express by uttering that sentence is true. Of course, in Moore-paradoxical cases, the speaker both puts forward a particular proposition as true and attributes to himself ignorance of that very proposition; in this respect, such cases differ from the case with which we're concerned, in which the speaker attributes to himself ignorance of the difference between right and wrong. However, the general point—that attributions of ignorance to one's current self can be problematic even when they would be true—should at least make us wary of following Ryle in uncritically passing from (3) to (4).

C5.P47 Indeed, the possibility of distinguishing sharply between (3) and (4) suggests an alternative approach to the puzzle. According to that approach, while there *is* something bizarre about someone's asserting, "I used to know the difference between right and wrong, but I've forgotten it" in response to the challenge "Don't you know the difference between right and wrong?", there is nothing in principle impossible about someone's forgetting the difference between right and wrong. After all, there are cases in which talk of *someone else's* forgetting the difference between right and wrong seems perfectly in order. Consider:

C5.P48 *The corruption of my congressman.* When my congressman first ran for elected office, I strongly supported him. That support was in part predicated on my high opinion of his moral sensibility and outlook. Based on his past behavior as a private citizen, I had a high degree of confidence in the moral judgments that he would be disposed to make about (for example) policies

¹² In Moore's original presentation of the paradox, the focus was on *belief*, and sentences like: "It's raining but I don't believe that it is." But as is often noted, the phenomenon also extends to cases involving knowledge. For discussion of Moore's paradox, see especially Sorensen (1988), chapter 1, and the essays collected in Green and Williams (2007).

affecting the poor. In any case, I would have unhesitatingly described him as “someone who knows the difference between right and wrong.” However, in the years since he assumed office, I’ve grown increasingly disillusioned. I now believe that his moral outlook is seriously defective in various ways. It’s not that I believe that I was wrong about what his moral outlook was years ago; rather, I think his moral outlook has changed for the worse. I believe that he’s no longer disposed to make many of the correct moral judgments that he once would have made. In fact, I describe his evolution like this: “He once knew the difference between right and wrong, but somewhere along the way, he forgot it.”

C5.P49 It might be objected that the talk of forgetting here is only metaphorical, and that my congressman didn’t literally forget anything. Whether a literal interpretation can be sustained depends, I think, on how the details of the story are filled in. Consider a variant case, in which my congressman still subscribes to the same moral views, but considerations of political expediency (etc.) frequently prevent him from acting on those views. In that case, while my disillusionment makes sense, the claim that he has “forgotten the difference between right and wrong” seems hyperbolic, and incorrect when taken literally. But suppose instead that what happened is this: in his previous life as a private citizen, my congressman frequently interacted with poor people, and these interactions both led him to form and then retain over time moral views that I take to be correct. Suppose that I currently believe both that (i) he would not now make the same moral judgments and that (ii) this is partially attributable to his relative lack of interaction with the poor in the years since he’s been a member of the government. In this case, talk of literal forgetting seems apt, for the following reason.

C5.P50 Literal forgetting involves ceasing to know what one once knew. Outside the moral domain, forgetting often occurs when the factors that would have sustained the knowledge over time cease to be available, or else diminish in their salience. For example, I once knew the address of my first apartment, but I have long since forgotten it. During the period of time when I knew the address of my first apartment, I was disposed to make a range of correct judgments (in response to questions, etc.), judgments that I would not make now in the absence of that knowledge. In this case, my forgetting is largely attributable to the absence of events that would have sustained the relevant knowledge over time. Similarly, in the case described above, I take my congressman to have once known significant moral truths that he no longer believes, and to

have once been disposed to make a range of correct moral judgments that he is no longer disposed to make, where this is attributable to the absence of experiences that would have sustained the beliefs and dispositions over time. Given this, there seems to be good reason to interpret my talk of his having forgotten the difference between right and wrong in a literal sense. Of course, it wouldn't be plausible to identify *forgetting the difference between right and wrong* with the forgetting of some particular moral truth, or some small number of moral truths. But if the perceived corruption is severe enough, then it will make perfect sense to talk about someone's having forgotten the difference between right and wrong, to the extent that it makes sense to talk about knowing (or not knowing) the difference between right and wrong in the first place.¹³

C5.P51 Imagine that a third party, who is as appalled by my congressman's recent behavior as I am, poses the following question to me: "Doesn't your congressman know the difference between right and wrong?" In response, I might assert:

C5.P52 (5) My congressman used to know the difference between right and wrong, but he doesn't anymore: somewhere along the way, he forgot it.

C5.P53 Given my background views about the evolution of my congressman's character, there's nothing at all bizarre or problematic about my asserting (5), even when it's taken literally. On the other hand, suppose that the third party confronts my congressman directly: "Don't you know the difference between right and wrong?" In response to this challenge, it *would* be bizarre for my congressman to assert:

C5.P54 (6) I used to know the difference between right and wrong, but I don't know it anymore: somewhere along the way, I forgot it.

C5.P55 —and not simply because such an admission would be politically calamitous. Moreover, this assertion would be no less bizarre even if my background

¹³ Compare another variant on this story, in which I express my admiration for the fact that my congressman has managed to retain his admirable moral outlook throughout his career: "Even though he was surrounded by corrupting influences for all of these years, he never forgot the difference between right and wrong."

To be clear, I don't believe that every case of corruption is best understood as one that involves literal forgetting, only that such an interpretation is appropriate in some cases.

views about the evolution of his character are accurate. More generally, first person attributions to the effect that one has forgotten the difference between right and wrong seem problematic in a way that third person attributions are not, even when the same proposition is in play.

C5.P56 Suppose that my congressman *has* forgotten the difference between right and wrong, and that this is something that I've picked up on. Why then should there be anything strange about *his* judging the same thing? Someone who claimed that he once knew the difference between right and wrong but now no longer does so would be attributing to his current self a moral outlook or sensibility that's radically defective compared to one that he used to have. (In effect, he judges that while he used to be a competent judge of moral matters, he's currently an incompetent judge.) He would be in a position to make such an attribution if he knew that (for example) his past self was sensitive to genuine moral considerations to which his current self is insensitive, or which his current self does not recognize. But it's unsurprising that people do not make such judgments about themselves, even when the judgments in question would be true. For a person who currently has the inferior moral outlook or sensibility will generally lack the perspective to recognize it as such. That is, one thing that is typically lost in a process of moral corruption, in which an individual moves from a better moral sensibility to one that is worse, is the perspective required to recognize the superiority of the better moral sensibility. Thus, the description that might very well be available to a third party who does know the difference between right and wrong—that the individual in question has become corrupt, or lost his grip on the difference between right and wrong—will typically not be available to the subject of the transformation himself.

C5.P57 In short, my hypothesis is that the proposition that an individual would express by asserting that

C5.P58 I used to know the difference between right and wrong, but I don't anymore: at some point, I forgot it.

C5.P59 is typically a *blind spot proposition* for that person, in the sense of Sorensen (1988): even if the individual would express a true proposition by uttering the relevant sentence, he will typically not be in a position to recognize that that proposition is true—and this is so even if various third parties are in a position to recognize that it is true.

C5.P60 The mark of a blind spot proposition is *idiosyncratic inaccessibility*: the typical routes by which the proposition is recognized as true by some are

systematically unavailable to others, for perspectival reasons.¹⁴ The lack of perspective at issue is not primarily a matter of self-deception or motivated believing. No doubt, it would be an unpleasant thing to come to believe, or even suspect, that one had “forgotten the difference between right and wrong,” and the psychological incentives to not conclude this might be considerable. Perhaps there are no similar psychological costs involved in concluding that other people have become corrupt, or forgotten the difference between right and wrong. On these grounds alone, we might expect it to be relatively uncommon for a person to attribute a diminished moral outlook to himself, compared to the readiness with which he attributes a diminished moral outlook to others. But such factors do not explain why there would be anything *bizarre* about claiming to have forgotten the difference between right and wrong. After all, people often manage to believe unflattering things about themselves, and that their current selves compare unfavorably to their past selves in various significant respects. For example, perhaps I place a high value on being in good physical condition; it would thus be unpleasant for me to conclude that:

C5.P61 (7) I used to be in good physical condition, but I’m not in good physical condition now.

C5.P62 Perhaps I have no such reluctance to conclude that your physical condition has deteriorated, so I’m somewhat quicker to draw the unflattering conclusion about you. Nevertheless, when I observe that a brisk walk leaves me out of breath, there is nothing odd about my judging that (7) is true.

C5.P63 When a person is in significantly worse physical condition than she used to be, she’s often in a position to recognize that this is so. More generally, physical deterioration does not typically impair one’s ability to accurately assess oneself or others along this dimension. By contrast, moral deterioration *does* impair one’s ability to accurately appraise oneself or others with respect to moral outlook. For in evaluating different moral outlooks, one will inevitably

¹⁴ Perhaps the most straightforward and least controversial cases of blind spot propositions are those expressed by classic Moore-paradoxical sentences such as “It’s raining but I don’t believe that it is.” Other people might very well know that it’s raining and that I do not believe this, in extremely straightforward ways (for example, by seeing rain out the window, and then hearing me express my belief that it’s not raining). However, none of the usual routes by which others could come to know the relevant conjunction are routes by which *I* could come to know it.

Although classic Moore-paradoxical sentences express blind spot propositions, not all blind spot propositions are expressed by such sentences. For example, Egan and Elga (2005) argue persuasively that it would never be reasonable to conclude that one is an “anti-expert” about some subject matter, even if one *is* an anti-expert about that subject matter, and others are in a position to conclude this on the basis of their evidence.

employ or presuppose a great deal of one's current moral outlook. A person whose current moral outlook is seriously defective is thus in a particularly bad position to recognize that this is so.

C5.P64 The crucial point is one that we've encountered before, in a related context. In section 3.5, I endorsed and defended the Platonic thought that, even if genuine moral experts existed, their recognition as such by non-experts would present formidable epistemological challenges, challenges that go beyond those that arise for the recognition of experts in many other domains. This is because arriving at views about the reliability of others with respect to moral questions typically requires one to engage in substantive moral reasoning and judgment of one's own, and this fact tends to limit the capacity of a person whose own moral outlook is relatively impoverished to recognize another person as a full-blooded moral expert. The current point is simply an application of the Platonic insight to the intrapersonal, diachronic case.

C5.P65 Again, it's helpful to consider the contrast between the epistemology of claims like:

C5.P66 (8) I have an unreliable sense of direction.

C5.P67 and

C5.P68 (9) I have an unreliable sense of right and wrong.

C5.P69 As noted in the earlier discussion of moral expertise, having an unreliable sense of direction is not a significant obstacle to knowing that one does: many people who have unreliable senses of direction have excellent evidence that they do. Indeed, it's plausible that many such people have evidence that their sense of direction is weak that is *just as strong* as the evidence that people with a strong sense of direction have that their sense of direction is strong. In this respect, things are symmetrical between the two groups. Although having a weak sense of direction is in many contexts a significant handicap, it's not a significant handicap with respect to knowing the truth about one's sense of direction. Again, at one level of abstraction, the explanation for this fact might be put as follows: a person with a weak sense of direction can and often does have evidence that he has a weak sense of direction, where his having that evidence *does not depend on the reliability of his sense of direction*. (In typical cases, this might consist of a long history of things not going well when he relied on his sense of direction.) In contrast, having an unreliable sense of right and wrong *is* a significant obstacle to knowing that this is the

case. And this is because the reliability of one's judgments about who does and who does not have a reliable sense of right and wrong is in large part hostage to the reliability of one's sense of right and wrong.

C5.P70 Thus, for someone who has lost his grip on the difference between right and wrong over time, the fact that this has occurred will typically have the status of a moral blind spot: it will be among the moral claims that he is not in a position to recognize as true, even if others are, and even if he's not prone to self-deception. My proposal is that the existence of this moral blind spot is what explains the puzzling character of the assertion "I used to know the difference between right and wrong, but I have forgotten it."

C5.P71 Should the proposal on offer be viewed as a candidate *solution* to Ryle's puzzle, or rather as an attempt to *dissolve* that puzzle? That depends on what we take Ryle's puzzle to be. On the one hand, we might take Ryle's puzzle to concern the impossibility of forgetting the difference between right and wrong; alternatively, we might take Ryle's puzzle to concern the impropriety of a certain speech act. If we understand Ryle's puzzle in the first way, then what I've offered amounts to an attempted dissolution of Ryle's puzzle, for I claim (and have argued that) there is in principle nothing impossible about someone's forgetting the difference between right and wrong. On the other hand, if we understand Ryle's puzzle as concerning the impropriety of the speech act, then my proposal purports to be a full-fledged solution to that puzzle, one that accounts for what would otherwise be a mysterious fact.¹⁵

C5.P72 I've claimed that there is a significant asymmetry between first person and third person attributions of the relevant kind of forgetting. The solution to Ryle's puzzle that I've offered both explains and predicts this asymmetry. Of course, one can (and I think, should) accept the asymmetry claim even if one does not accept my proposal, since alternative explanations of the asymmetry are possible. However, the fact that the current proposal predicts the asymmetry does give it an advantage over rival accounts that fail to do so. Consider, for example, the following account:

C5.P73 Knowledge of the difference between right and wrong is genuine knowledge. But it's not susceptible to being forgotten for the same reason that one's knowledge of the most elementary truths of arithmetic isn't susceptible to being forgotten: forgetting requires ceasing to know what one once knew,

¹⁵ As noted above, Ryle himself slides back and forth between the two interpretations, and as a result the text as a whole is perhaps indeterminate with respect to the question of what "Ryle's puzzle" really is. I'll continue to write as though "Ryle's puzzle" concerns the impropriety of the speech act, as is suggested by a literal reading of his official statement of the puzzle, and hence of my "solution" to the puzzle. But this is at least somewhat stipulative on my part.

but the difference between right and wrong is simply so obvious (at least, to someone who has already learned it) that anyone who has once grasped it will also know it at later times. Of course, a significant part of morality is *not* obvious, as is witnessed by the phenomenon of persistent moral disagreement among generally intelligent people. But the *basic core* of morality—what one needs to grasp, in order to count as someone who knows the difference between right and wrong—is obvious, in a way that the answers to contested moral questions are not. (After all, even if there is disagreement about the morality of abortion, there are presumably at least some on both sides of the issue who “know the difference between right and wrong,” in whatever sense Ryle’s properly brought up child does.) Thus, the assertion “I used to know the difference between right and wrong, but I’ve forgotten it” is bizarre for the same reason that the assertion “I used to know elementary arithmetic, but I’ve forgotten it” is.

C5.P74 This account predicts that whatever oddity attaches to “I used to know the difference between right and wrong, but I’ve forgotten it” should also attach to “My congressman used to know the difference between right and wrong, but he’s forgotten it.” The fact that this is not the case suggests that the appeal to obviousness is not the correct explanation of the puzzle.¹⁶

C5.P75 Ryle held that it’s conceptually impossible to forget the difference between right and wrong, and that talk of doing so reflects a kind of category mistake. If it’s impossible to forget the difference between right and wrong, then it is, a fortiori, impossible to recognize that one has done so. Against this, I’ve argued that it is in principle possible to forget the difference between right and wrong, although for any agent who has actually done so, there will be systematic obstacles to recognizing that this is what has happened. Could one ever be in a position to recognize that one has forgotten the difference between right and wrong, given these systematic obstacles? I believe that one could. Consider, for example, the following kind of case:

C5.P76 You undergo a particular kind of brain surgery, a common side effect of which is impairment in those parts of the brain that are responsible for moral judgment. Upon waking up from the surgery, you begin to consider

¹⁶ The same objection also tells against accounts that attempt to explain the puzzle by appealing to the (alleged) fact that knowledge of the difference between right and wrong is part of our innate cognitive endowment.

various moral questions. You find, however, that in each case your mind draws a blank when you attempt to arrive at an answer. You remember once having held many confident moral views (although you do not now remember the content of those views) and are rationally confident in your past self.

C5.P77 In these circumstances, it would be natural for you to judge that while you once knew the difference between right and wrong, you don't currently know the difference. Moreover, given this background story, it would be natural for you to assert, "I used to know the difference between right and wrong, but I've forgotten it," in response to the kind of challenge envisaged by Ryle. Thus, it seems that not only is there no impossibility in forgetting the difference between right and wrong, but that, in at least some special cases, there is no impossibility in someone's recognizing that she has done so. But both of those points are consistent with the idea that, in any ordinary context, the utterance "I used to know the difference between right and wrong, but I have forgotten it" would be bizarre, for the reasons provided here.

C5.P78 More generally, there is something bizarre about judging that one has forgotten the difference between right and wrong, even though there is nothing at all bizarre about judging that one has forgotten any number of other things. It's tempting to think that the phenomenon in question points to some deep fact about the nature of morality: for example, that the subject matter of morality is somehow different in kind from the many things that can be forgotten. On the present proposal, the phenomenon can be accounted for as an artifact of perspectival limitations. Moreover, the fact that such an explanation is available stands in the way of drawing radical conclusions about the metaphysics or epistemology of morality, since such an explanation is consistent with morality being the kind of thing that can be lost in the way that other objects of potential knowledge are. Here as elsewhere, the difference between moral knowledge and non-moral knowledge might be smaller than initial appearances suggest.

C5.P79 As noted in section 5.1, ordinary empirical knowledge can be lost even when nothing has been forgotten. In particular, ordinary empirical knowledge is susceptible to being lost *via* defeat, as when evidence emerges that makes it unreasonable to go on believing what one once knew. So now I want to consider the phenomenon of defeat as it arises in the moral domain, including challenges to the idea that our moral views are susceptible to being rationally undermined in the same ways as our non-moral views.

C5.S6

5.6 Defeat

C5.P80 Consider a paradigm case of non-moral defeat.¹⁷ At some point in time, you know that a particular coin is fair, in the sense that it's disposed to land "heads" more or less 50 percent of the time when flipped in the usual way. You then flip the coin repeatedly, and watch as it lands heads n consecutive times, for some large number n. (The coin is in fact fair; what you observe is simply an extraordinarily improbable sequence of outcomes.) Presumably, at some point (given some sufficiently large value for n) you are no longer justified in believing that the coin is fair, and thus, no longer in a position to know that the coin is fair.

C5.P81 In the coin case, your knowledge that the coin has landed heads n consecutive times serves as a *rebutting* defeater for your belief that the coin is fair. A rebutting defeater for your belief that p undermines your justification for believing p by providing evidence for not-p. Your knowledge that the coin has landed heads n consecutive times is evidence that the coin is biased, which amounts to evidence that it is not the case that the coin is fair; you are thus no longer in a position to know that the coin is fair because you have sufficiently strong evidence that favors the negation of that claim. In contrast, an *undercutting* defeater undermines your justification for believing p without providing justification for believing not-p.¹⁸

C5.P82 Here is one variety of undercutting defeater that will be particularly important for the discussion that follows: evidence about the etiology of a belief that suggests that the belief is held for reasons unrelated to its truth. For example, suppose that you believe that the last time the coin was flipped it landed heads because you seem to remember that it did. You then receive strong evidence that suggests that you have been hypnotized to believe this, and so the same proposition would seem true to you regardless of whether it was true. (The hypnotist himself has no idea whether the coin landed heads or tails.) Evidence of this sort is not itself evidence in favor of the proposition that the coin did *not* land heads; it's not the kind of information which would tend to justify you in believing that the coin landed tails. It's therefore not a rebutting defeater for your belief. But if the evidence of your having been hypnotized is strong enough, then it is rational for you to lose confidence in that belief. Given that you are no longer justified in believing that the coin

¹⁷ The example that follows is based on one due to Williamson (2000).

¹⁸ For the distinction between undercutting and rebutting defeaters, see especially the classic discussion in Pollock (1986), which introduced the terminology of "undercutting" and "rebutting."

landed heads, you are not in a position to know that it did (even if you were in a position to know this before receiving the evidence of hypnosis—as in a case in which the evidence of hypnosis is misleading).

C5.P83 Paradigmatic cases of defeat discussed in the epistemology literature involve a temporal element: there is an earlier time at which one knows, and a later time at which one ceases to know in virtue of acquiring evidence that undermines one's knowledge. Here we will also be concerned with the broader phenomenon of which paradigmatic defeat is a special case: cases in which having evidence of a certain kind *precludes* one from knowing, regardless of when that evidence is acquired, and regardless of whether one would know in the absence of the evidence. Thus, a skeptic might argue that we have evidence that the way in which we arrive at our beliefs about a certain topic is unreliable, and that given this evidence our beliefs about that topic don't amount to knowledge, while bracketing the question of whether we would or might be in a position to know in the absence of this evidence.

C5.P84 It is more or less uncontroversial that ordinary empirical knowledge can be lost in virtue of being defeated.¹⁹ I will argue that moral knowledge is susceptible to defeat in each of the various ways that ordinary empirical knowledge is. But before turning to the case for this, let's consider an argument for the opposite conclusion. Suppose that, as many believe, moral knowledge—or at least, knowledge of the core truths of morality—is *a priori* knowledge. On a traditional view of the *a priori*, it is a characteristic mark of *a priori* knowledge that it is *indefeasible*: if you know that something is true on the basis of *a priori* considerations or reasoning, then the mere acquisition of additional information cannot undermine your knowledge. For example, if you know a mathematical theorem on the basis of a genuine proof, then merely gaining more information will never undermine your knowledge.²⁰ Clearly, if *a priori* knowledge cannot be undermined by being defeated, and moral knowledge is *a priori* knowledge, then moral knowledge cannot be undermined by being defeated.²¹

¹⁹ An exception to the consensus is Neta 2009, who summarizes the state of opinion among epistemologists at the time of his writing as follows: “The account of knowledge that I accept implies that knowledge is indefeasible, and so it will strike virtually all epistemologists as suffering from obvious and devastating problems” (p. 163). For a reply to Neta, see Williamson 2009. Of course, if Neta is correct in his contention that *all* knowledge is indefeasible, then this is not a respect in which moral knowledge and non-moral knowledge differ.

²⁰ Of course, you might later forget the proof and theorem, and thus cease to know, but this would be a case of ceasing to know by forgetting as opposed to a case of defeat.

²¹ As noted in Chapter 4, from the fact that a given subject matter is *a priori*, in the sense that the relevant truths *can* be known *a priori*, it doesn't follow that the relevant truths are known *a priori* by anyone who knows them. Thus, even if the line of argument sketched above is sound, it wouldn't follow that moral knowledge could never be lost via defeat; for example, someone who knows a moral

C5.P85 However, we should reject the view that a priori knowledge is indefeasible: it too is susceptible to being undermined by sufficiently strong defeating evidence. For example, suppose that you add a series of two digit numbers in your head, and correctly arrive at the relevant sum. Assuming that circumstances are otherwise favorable, your belief that the sum is a certain number is something that you know; moreover, this piece of knowledge seems as good a candidate as any for being a priori. But even on the assumption that it's a piece of a priori knowledge, it doesn't follow that there is nothing that you could subsequently learn that would rationally undermine your confidence in the relevant belief. For example, suppose that you redo the calculation and arrive at a slightly different answer, and you have no way of knowing on which occasion you made the mistake. Or suppose that you learn that another person, who you know is generally more reliable when it comes to such calculations, has arrived at a different answer (or even that a number of such people have independently converged on some different answer). At some point, your evidence of having made a mistake will be sufficiently strong that the reasonable thing for you to think that is that you made a mistake. And once the reasonable thing for you to think is that you made a mistake, you will no longer be justified in believing the relevant proposition, despite the fact that, in the absence of the misleading counterevidence, your belief would be a piece of a priori knowledge. Even justification afforded by a priori reasoning can be defeated.²² If moral knowledge is a priori, it might still be susceptible to being lost via defeat.

C5.P86 Consider first cases of rebutting defeat, in which one is precluded from knowing p in virtue of acquiring reason to believe not-p. Here again it's natural to appeal to the phenomenon of interpersonal disagreement, in which justification for believing p is threatened by the fact that others have arrived at the conclusion that not-p. In *The Methods of Ethics*, Sidgwick famously writes:

C5.P87 if I find any of my judgments, intuitive or influential, in direct conflict with a judgment of some other mind, there must be error somewhere: and if I have

truth on the basis of testimony might lose that knowledge in virtue of acquiring sufficiently strong evidence that suggests that the person from whom she acquired the belief is unreliable. Rather, what would follow is that any piece of moral knowledge that is known on a completely a priori basis couldn't be lost in the relevant way. That having been said, I take it that if the line of argument sketched in the text is sound, that would suffice to establish the conclusion that there is a significant difference between moral knowledge and ordinary empirical knowledge.

²² Despite the historical popularity of the opposite view, the point is now widely conceded by friends of the a priori. Indeed, contemporary defenders of the a priori frequently emphasize the defeasibility of a priori justification (including its defeasibility by empirical considerations) as an important part of their project of making plausible the idea that a priori justification exists at all. On this point, see among others BonJour (1998) and Burge (1993).

C5.P88 no more reason to suspect error in the other mind than in my own, reflective comparison between the two judgments necessarily reduces me to a state of neutrality. (Sidgwick 1907[1981], p. 342)

C5.P89 Elsewhere I've appealed to the idea that Sidgwick endorses in this passage in order to argue that, given the kind of persistent and intractable disagreement we find about many moral questions, our beliefs about those questions are not good candidates for knowledge (McGrath 2008, 2011). If this is correct, then whether it's possible for us to know the answers to a significant range of moral questions is in fact quite fragile, and depends on the distribution of opinion about those questions. That conclusion depends on substantive views about the significance of moral disagreement that are themselves controversial.²³ Although the truth of those views would suffice to establish the claim that moral knowledge is vulnerable to being undermined by the right kind of disagreement, the former isn't necessary for the latter. Indeed, we need to assume very little about the epistemic significance of disagreement in order to arrive at the conclusion that there are at least some cases in which the conflicting opinions of others undermine the possibility of moral knowledge in virtue of playing the role of rebutting defeaters.

C5.P90 In particular, we need not accept so-called "conciliatory" views about the epistemic significance of disagreement,²⁴ according to which one is rationally required to treat one's own views and the views of one's peers as on a par, in order to accept the claim that the differing views of others can preclude knowledge of some proposition. In arguing that peer disagreements typically require relatively extensive revisions in the views of the parties who disagree, conciliationists emphasize the apparent symmetries that seem to obtain between them. Of course, in a moral dispute in which one person knows that her view is true (or would know that her view is true in the absence of the disagreement), there is no possibility that the dissenting party is in the same position. So things are asymmetrical in at least that respect. But even anti-conciliationists who hold that such asymmetries can make it justifiable for the party in the right to maintain her original opinion will typically hold that, if a sufficient amount of evidence of the relevant kind were amassed on the other side, then at some point, the person would no longer be in a position to know.²⁵ (Compare: even if you do a non-trivial piece of arithmetic correctly,

²³ Extended critical discussions include Decker and Groll (2013), King (2011), and Sherman (2014).

²⁴ See especially Christensen (2007), Elga (2007), and Feldman (2006).

²⁵ This is, for example, an explicitly drawn implication of Kelly's (2010) "total evidence view" (see pp. 137–44) which is offered as an alternative to conciliationist views and one that's more hospitable to maintaining one's convictions in the face of disagreement.

and thus know that the sum is such-and-such a number, there is surely some amount of evidence consisting of the credible testimony of others to the effect that some other answer is correct that would make it reasonable to think that you are wrong. And if it's reasonable to think that you are wrong, then you are no longer in a position to know.) Of course, a theorist who holds that genuine knowledge is relatively robust in the face of disagreement will think that the cases in which a person loses moral knowledge because of the conflicting opinions of generally reliable (but in this case mistaken) others are relatively rare. By the same token, however, such a theorist will presumably take cases in which *non-moral* knowledge is undermined by disagreement to be relatively rare as well; in that sense, the alleged similarity between moral and non-moral knowledge is preserved.

C5.P91 We should also notice how the claim that moral knowledge can be lost as a result of interpersonal disagreement is supported by some of the conclusions that have already been argued for in this book. In Chapter 3, I defended the view that moral knowledge can be acquired by testimony, or (more generally) by moral deference. In such cases, the fact that another person believes a moral claim is treated as a reason to believe that claim, or to increase one's confidence in it. (In order to maximize the plausibility of the claim that the practice is unobjectionable, we can focus here on cases in which there is positive reason to think that the person to whom one defers is reliable.) Now, it's natural to think that the propriety of adopting or increasing one's confidence in a moral view because it's *held* by a person whom one takes to be reliable, and the propriety of relinquishing or decreasing one's confidence in a moral view because it's *rejected* by a person whom one takes to be reliable, are simply two sides of the same coin. Of course, there's no inconsistency or incoherence in maintaining that knowledge can be gained but not lost by moral testimony. (In particular, it seems easier to be moved from a state of not-knowing to a state of knowing by the testimony of a generally reliable person who knows, than it is to be moved from a state of knowing to a state of not-knowing by the testimony of a generally reliable person who does not know.) Still, if one can be moved from a state of not-knowing to a state of knowing by the testimony of a generally reliable person in the good case, then surely testimony could provide at least *some reason for doubt* in the bad case (even if that testimony is insufficient to undermine one's knowledge). And if testimony provides some reason to doubt, then it seems as though in principle enough additional evidence of the same kind would be enough to undermine knowledge. (Again, consider the effects of

discovering that a large number of people who you antecedently judge to be in a good position to arrive at a true view about the issue turn out to converge on the opposite answer.)²⁶

C5.P92 When the conflicting judgment of a reliable source undermines the knowledge that one would otherwise enjoy, or when one is precluded from knowing by the fact that too many others whom one has reason to trust think differently, the evidence that precludes one's knowing serves as a rebutting defeater: it counts against believing p by supporting not-p. But even if it's possible to lose moral knowledge in this way, that leaves open the extent to which moral knowledge is vulnerable to other types of defeat. As noted above, in the non-moral case, evidence about the etiology of a belief can also preclude knowledge. (Recall the case in which you receive evidence that your belief that the coin landed heads on a given flip is due to hypnotic suggestion, as opposed to a genuine memory.) In the moral domain, the project of offering "debunking" explanations of why we hold the moral views that we do has been famously pursued by thinkers such as Marx ([1867]1972), Freud ([1929]1961), and Nietzsche ([1887]1967). In contemporary moral philosophy, the possibility of explaining our moral beliefs in terms of evolutionary mechanisms that have nothing to do with the truth of those beliefs has been much discussed.²⁷ It has often been assumed, by both proponents and critics of such projects, that if our moral beliefs could be explained in this way, that would undermine the rationality of our continuing to hold them.

C5.P93 In his final book, *Justice for Hedgehogs* (2011), Ronald Dworkin argues vigorously and at length for the claim that this is wrong. According to Dworkin, there is a crucial difference between our moral beliefs and our ordinary empirical beliefs: while compelling evidence that an ordinary empirical belief is held for reasons that are independent of whether it is true undermines the rationality of continuing to hold it, compelling evidence that a moral belief is held for reasons that are independent of whether it is true does not similarly undermine its rationality. On this view, our moral beliefs differ from our ordinary empirical beliefs in that they are relatively immune to being

²⁶ Similarly, section 4.3.5 argued that a potentially important source of misleading evidence in the moral domain is the possibility that people who either are or seem to be morally upstanding people in other respects will engage in some activity or practice that's morally wrong, in a social context in which its wrongness is not generally recognized. If that's correct, then a member of a society in which such misleading evidence is in abundance with respect to some issue might not be in a position to know certain moral truths that he would be in a position to know, if he were located in a more epistemically favorable social environment.

²⁷ Much of the recent literature is inspired by Street (2006). For an overview, see Vavova (2015).

undermined by the provision of information about their etiology. Significantly, crucial premises in Dworkin's case for this conclusion have been endorsed by other leading moral philosophers, including Thomas Nagel and Thomas Scanlon.²⁸ If correct, Dworkin's claim would amount to a direct refutation of the working hypothesis around which this book has been organized. For the correctness of Dworkin's claim would amount to a vindication of the idea that there is a way of losing non-moral knowledge that simply does not arise in the moral sphere. The final section of this chapter examines Dworkin's case for the existence of such an asymmetry.

C5.S7

5.7 Dworkin Against Debunking

C5.S8

5.7.1 Counterfactuals, Conceivability, and Modality

C5.P94 One of the most striking theses of *Justice for Hedgehogs* is that our moral beliefs differ from our ordinary empirical beliefs in the following way: although it's frequently the case that the correct explanation of why an ordinary empirical belief is held invokes the fact that it is true, it's *never* the case that the correct explanation of why a *moral* belief is held invokes the fact that it is true. The fact that Dworkin grants that the truth of an ordinary empirical belief sometimes plays a role in explaining why it is held differentiates his view from one that is significantly more radical: the view that truth never figures in the correct explanation of why *any* belief is held, regardless of its subject matter. Although this more radical view has sometimes been defended (Bloor 1974, 1981), it's clear that Dworkin rejects it:

The radical view: Truth never figures in the correct explanation of why a belief is held.

C5.P95 people's beliefs about the physical world are often caused directly or indirectly by the truth of what they believe, and when they are, that fact confirms the truth of their belief. The best explanation of why I believe that it rained earlier today includes the fact that it did rain. (2011: 71)

C5.P97 Moreover, Dworkin holds that when it comes to ordinary empirical beliefs, the lack of any possible explanatory connection between a belief and the

²⁸ Thus, Nagel (1996: Chapter 6) agrees with Dworkin that it's characteristic of our moral convictions that undermining them always requires first-order moral reasoning and argument, as opposed to purely non-moral information about their aetiology. In his *Locke Lectures*, Scanlon (2014: 26–7) agrees with Dworkin that a well-known “explanatory requirement” associated with Harman (1977) is a test that makes good sense for our empirical beliefs about the natural world but one that it is deeply inappropriate to apply to our moral beliefs. I discuss both of these issues below.

state of affairs that it purports to represent undermines the credibility of that belief:

- C5.P98 Suppose that though you believe it rained in France today, no rain in France could possibly figure in any explanation of why you believe that... You would then have no reason at all to think it had rained there. (2011: 71)

C5.P100 Given that Dworkin is a passionate defender of the view that moral beliefs, like beliefs about the physical world, can be objectively true, it's natural to expect him to tell a parallel story about the moral domain: that in at least some favorable cases, the fact that a moral opinion is true figures in the best explanation of why it's held, and that in cases in which there is no possible explanatory connection between the truth of a moral opinion and its being held, that tends to undermine the credibility of the opinion. However, Dworkin emphatically denies that the domains are parallel in these respects. Rather, his view is that it's a deep confusion to think that the truth of a moral opinion could ever figure in the best explanation of why it is held, but that morality is none the worse off for that. For this reason, our moral beliefs are not hostage to the possibility of being "debunked" by non-truth-related explanations of why we hold them, in the way that our beliefs about the physical world are.

Dworkin is
a realist.
Dworkin:
moral beliefs
cannot be
debunked
in the same
way that
non-moral
beliefs, all

C5.P101 Why should what holds for our beliefs about the physical world not also hold for our moral beliefs? In both *Justice for Hedgehogs* and his seminal earlier paper "Objectivity and Truth: You'd Better Believe It," Dworkin appeals to the following idea: things are asymmetrical between our beliefs about the physical world and our moral beliefs because the former beliefs, but not the latter, are ones for which it makes sense to ask a certain kind of *counterfactual question*. Specifically, it makes sense to ask of beliefs about the physical world whether they would still be held even if the facts that they purport to represent had been otherwise, but it makes no sense to ask the same question of our moral beliefs. Here is a statement of this idea from the earlier paper:

C5.P102 Consider Gilbert Harman's suggestion that we cannot regard any belief as reliable unless we think that the best causal explanation of why we hold it refers to the state of affairs it describes. In some form, this test does seem appropriate to beliefs about the physical world... But nothing in the content of moral (or aesthetic or mathematical or philosophical) opinions invites or justifies such a test. On the contrary, the content of these domains excludes it, because an adequate causal explanation of a belief includes showing that

the belief would not have occurred if the alleged cause had not been present, and we cannot understand or test that counterfactual claim with respect to moral or aesthetic beliefs because we cannot imagine a world that is exactly like this one except that in that world slavery is just or *The Marriage of Figaro* is trash. (1996: 119)²⁹

C5.P103 C5.P104 He offers the same line of argument in *Justice for Hedgehogs*.³⁰

C5.P105 The idea seems to be something like the following. The reason why the best explanation of one's belief that *it rained earlier today* might very well invoke the fact that it rained earlier today is that we can both make sense of and investigate the truth of the following counterfactual:

C5.P106 If it had not rained earlier today, then one would not now believe that it rained earlier today.

C5.P107 In contrast, because we "cannot imagine a world that is exactly like this one except that in that world slavery is just," we cannot understand or investigate the truth of the following counterfactual:

C5.P108 If slavery had not been unjust, then one would not believe that it is unjust.

C5.P109 And according to Dworkin, our being able to understand and evaluate this counterfactual is an (unfulfillable) necessary condition for the fact that *slavery is unjust* to figure in the best explanation of one's believing that slavery is unjust.

C5.P110 This is not a good argument. In general, even if we can't imagine a world that is exactly like ours except that p is false, it doesn't follow that p can't be a part of the best explanation of why someone believes that p is true. To see this, consider another case that Dworkin mentions in the passage above: the case of mathematics. Imagine a mathematician who initially has no opinion about whether some mathematical conjecture is true or false. Suppose that she subsequently succeeds in proving the conjecture and thus comes to believe the relevant proposition on the basis of the proof. In this case, the explanation of why the mathematician currently believes the proposition is that she succeeded in proving the theorem. But of course, that explanation entails that

²⁹ For the claim that Harman's causal explanatory requirement is an appropriate test for our beliefs about the physical world, but that it is a deep mistake to think that it applies to domains such as morality and mathematics, compare the very similar discussion in Scanlon 2014: 26–7.

³⁰ See his discussion of the "crucial counterfactual question" on pp. 73–4.

the relevant proposition is true. So it is essential to the best explanation of why the mathematician believes the proposition that the proposition is true; one could not offer an equally good explanation of why she holds the belief that is neutral with respect to the truth of her belief.³¹

C5.P111 In short, (1) might be every bit as good of an explanation as (2):

- C5.P112 (1) The mathematician believes that p because she proved that p is true.
- C5.P113 (2) The mathematician believes that it rained earlier today because she observed that it rained earlier today.

C5.P114 Of course, given that p is a mathematical truth, we won't be able to imagine a world that is exactly like ours except for the fact that in that world p is false. Because of this, the counterfactual "If p had been false, then the mathematician would not have believed p" will strike us as unintelligible, in the way that counterfactuals whose antecedents consist in the negation of mathematical truths generally strike as unintelligible. ("If 2+2 had not equaled 4, then...") But that has no tendency to cast doubt on the truth-invoking explanation of the mathematician's belief. Indeed, our inability to evaluate or understand the relevant counterfactual does not even mean that we cannot investigate or acquire evidence that bears on the truth-invoking hypothesis as an explanation of why the mathematician believes as she does. (For example, if we learn from the mathematician's diary that she first became convinced of the relevant proposition years earlier, on the basis of a sheer hunch, and that she's believed it unwaveringly ever since, that discovery might very well cast doubt on whether the truth-invoking hypothesis is really the best explanation of the mathematician's belief.)

C5.P115 The same point holds for another of Dworkin's comparisons, one which is in some respects more closely analogous to the moral case: the case of the aesthetic. Given that aesthetic properties supervene on non-aesthetic properties, Dworkin is surely right that we cannot imagine a world which is exactly

³¹ *Objection:* But can't we explain why the mathematician believes the proposition by citing the fact that *she takes herself to have proved it*? One can take oneself to have proved a proposition even if one has not actually proved it, and indeed, even if the relevant proposition is false. So we can explain why the mathematician believes the proposition in a way that's neutral with respect to whether the proposition is true after all, by citing the fact that she takes herself to have proved it. *Reply:* It's a mistake to think that nothing has been lost by substituting in this weaker, less committal explanation. For we can ask the following question: what explains why the mathematician takes herself to have proved the proposition? In a given case, the explanation for why she takes herself to have proved the proposition might be the fact that she proved it. (Given her mathematical competence, she wouldn't take herself to have proved the theorem unless she had actually done so.) So in a sufficiently comprehensive explanation of why she currently believes the proposition, the fact that she successfully proved the theorem will still appear.

like our own except for the fact that the Marriage of Figaro is a piece of trash. Still, given that we think that it's true that the Marriage of Figaro is beautiful (as Dworkin does), it certainly seems as though we can ask whether the best explanation of why a particular person believes that it's beautiful includes her appreciation of its beauty (a truth-invoking explanation), or rather, because certain psychological mechanisms guarantee that she will end up believing that something is beautiful just in case her friends think that it is beautiful, regardless of its actual aesthetic merits. In any case, whether it's ultimately defensible to think that someone's belief could be best explained in this way, no reason for doubting that it could is supplied by the fact that "we cannot imagine a world that is exactly like this one except that...the Marriage of Figaro is trash."

C5.P116 More generally: nothing about whether some truth could be part of the best explanation of someone's believing it follows from its *modal status*. But this observation shows that Dworkin's counterfactual criterion cannot be right, because the unintelligibility of the relevant counterfactuals arises from the fact that their antecedents consist in the negations of necessary truths. In fact, notice that if the orthodox Kripkean view about metaphysical necessity is correct, then Dworkin's own account of what distinguishes the types of beliefs that can be explained by truth-invoking explanations and the types of beliefs that cannot does not yield consistent results. On the one hand, the belief that *water is H₂O* is a belief about the physical world, and thus seems eligible to be explained by citing the fact that water is H₂O, together with a story about how scientists recognized this fact.³² On the other hand, given that water is H₂O in our world, we cannot coherently imagine a world that's exactly like this one except for the fact that water is not H₂O.

C5.P117 I conclude then, that if it's true that the injustice of slavery can play no role in explaining why people think that it is unjust, this has nothing to do with the fact that we cannot imagine a world exactly like ours except for the fact that slavery is not unjust.

C5.S9

5.7.2 Truth, Justification, and Explanation

C5.P118 Dworkin has another, independent argument for the conclusion that the truth of a moral view never plays a role in explaining why it is held. The key idea

³² Dworkin himself explicitly cites our beliefs about chemistry as paradigms of beliefs that can be explained by hypotheses that cite their truth (2011: 69).

↓
it's pointless

here is that, when it comes to our moral views, it's inevitably "pointless" or "otiose" (2011: 74) to appeal to their truth in explaining why we hold them. He develops this line of thought as follows:

- C5.P119 even if we assumed that moral truth does have mysterious causal potency, that assumption could be of no help whatsoever in justifying our moral beliefs. We would have to know, independently, whether those beliefs were true before we could intelligibly cite truth as their parent. That requirement is particularly clear when you offer to explain someone else's moral opinions. You think that affirmative action is unfair but your friend thinks it perfectly fair. You cannot think that his belief is caused by the truth; if you want to explain his belief you must compose a personal-history explanation. You find one that you think complete and persuasive: you cite his education in a knee-jerk liberal family. But now you change your own mind: you are suddenly convinced by his arguments that affirmative action is fair. You now think that what your friend believes is true, but you have discovered nothing that could impeach your earlier explanation of why he believes it. If the personal-history explanation was adequate before, it remains adequate now. You may be tempted now to say that, after all, the truth did play a role in the causal story of how he came to think what he does. But that shows only that [the appeal to truth] is never more than a fifth, spinning wheel in any explanation. (2011: 74)

Why? Because
we would have
to know that
they are true
prior to
citing truths
in our justification.

- C5.P120 Consider first the claim that, in attempting to *justify* a moral belief, it's unhelpful to appeal to its truth. There is an obvious sense in which this is correct. If you and I disagree about whether affirmative action is fair, then it would of course be ridiculous for me to attempt to justify my belief to you by citing its (putative) truth. However, there is no difference here between our moral beliefs and our non-moral, empirical beliefs. If the scientific community is divided about whether some chemical hypothesis is true, then it would obviously be pointless for those who are already convinced of the hypothesis to attempt to justify their belief by citing the (alleged) fact that it is true. But this uncontroversial point about justification has no tendency to show that (for example) the best explanation of why chemists believe that water is H₂O does not invoke the fact that water is H₂O. The same point holds in the moral domain. When William Wilberforce took to the floor of parliament to attempt to justify his belief that slavery is unjust, it would obviously have been pointless for him to cite the fact that slavery is unjust. But this uncontroversial point about justification does nothing to show that the

Appeals to truth are useful
in disputes about the
truth.

184 MORAL KNOWLEDGE

injustice of slavery plays no role in the explanation of why some people believe that it's unjust.

C5.P122 Consider next Dworkin's discussion of your attempts to explain your friend's belief that affirmative action is fair. In Dworkin's example, you originally accept a "personal-history explanation" of your friend's belief, according to which

C5.P123 (H1) My friend believes that affirmative action is fair only³³ because he grew up in a knee-jerk liberal family.

C5.P124 At a later time, you become convinced by your friend's arguments that affirmative action is fair. At this point, it's natural for you to consider another explanation of why your friend believes as he does:

C5.P125 (H2) My friend believes that affirmative action is fair because he has sound arguments for that conclusion.

C5.P126 Notice that H2, unlike H1, is a truth-invoking explanation: if your friend has sound arguments for the conclusion that affirmative action is fair, then it's true that affirmative action is fair. For this reason, the hypothesis would not be considered a live option by anyone who (like your past self in Dworkin's example) is of the opinion that affirmative action is unfair.

C5.P127 Dworkin suggests that, when you change your mind about affirmative action and the quality of your friend's arguments, you should regard this as irrelevant to the acceptability of the explanatory hypothesis that you have accepted up until now, i.e. H1. ("You have discovered nothing that could impeach your earlier explanation...if the personal-history explanation was adequate before, it remains adequate now.") But this claim neglects the way in which the rational credibility of an explanatory hypothesis can depend on which alternative hypotheses are taken to be live options. As philosophers of science often note, an explanatory hypothesis can have its credibility dramatically boosted when a formidable competing hypothesis is eliminated from consideration.³⁴ In such cases, the credibility of the hypothesis increases when the field of alternative hypotheses shrinks. Of course, the opposite can also occur: the credibility of a currently accepted hypothesis might be

³³ Recall that in Dworkin's example, you originally regard this as a "complete" explanation of your friend's belief.

³⁴ For discussion, see Lipton (2004).

dramatically reduced, when a new hypothesis is introduced, or when a previously rejected hypothesis is reinstated as a live option. Dworkin's example is a potential illustration of this last possibility. Back when you thought that affirmative action is unfair and your friend's arguments in its favor were unsound, the truth invoking explanation H2 was not among the live options, because it is inconsistent with what you then took to be the case. After you change your views in the relevant ways, H2 is no longer ruled out by what you believe, so it can reenter the competition. Contrary to what Dworkin suggests, this might very well affect whether it's rational for you to believe that H1 is the actual explanation of why your friend believes as he does.

C5.P128 What would it take to show that an appeal to truth is never more than a "fifth, spinning wheel" in explaining our moral beliefs? Simply this: that whenever we are tempted to accept a truth-invoking hypothesis of why someone accepts a moral belief, there is *always* some superior "personal history explanation" that does *not* invoke truth. (Compare the task of showing that there are no cases in which the best explanation of why someone holds a mathematical belief invokes the truth of that belief.) But contrary to what Dworkin seems to think, there is nothing in the passage quoted above that bears on the possibility of successfully executing this project.

C5.P129 Appreciating the way in which the credibility of an explanatory hypothesis can vary depending on which alternatives are considered live options puts us in a position to answer a question that's made salient by Dworkin's discussion. The question is this: given that (as he correctly notes) one could never *justify* a moral belief to another person by citing its truth, why would anyone *care* whether the truth of a moral belief could ever play a role in explaining why it is held? I take it that one natural answer to this question concerns the apparent possibility broached earlier: the apparent possibility that at least some non-truth-invoking explanations of why we hold the moral views that we do would undermine the rationality of our continuing to endorse those views, as well as our claims to know that they are true. Presumably, *something* explains why we hold the moral views that we do: the fact that one holds a moral belief is never simply a brute fact. If there is some general reason for thinking that the truth of a moral belief could never figure in the best explanation of why it is held, then for any particular moral belief, any truth-invoking hypothesis should not be considered a live option. And it might very well be that, once all truth-invoking hypotheses are removed from the field, the most credible remaining hypothesis about why one believes as one does is an explanation which has the following property: if accepted, it would give one a reason to lose confidence in the moral belief that it explains.

C5.P130 Compare the empirical case: if an oracle informed us that the correct explanation of why many scientists believe that global warming is occurring has nothing to do with the occurrence of global warming, thus eliminating any truth-invoking explanation as a live option, it might be that the next best explanation (the one that it would then be reasonable for us to accept) would be one that undermines the rationality of continuing to believe in global warming. Thus, the reason why a person who acknowledges that one could never justify a moral belief by citing its truth might nevertheless care about whether moral truth is the kind of thing that could play an explanatory role is simply this: if all of the truth-invoking-hypotheses are eliminated from consideration, it might very well be that the explanations of her moral beliefs that it would then be reasonable for her to accept are ones that undermine the rationality of her continuing to endorse those beliefs. More generally, the elimination of all-truth invoking hypotheses might leave debunking explanations as the most reasonable explanations of our moral beliefs left standing.

C5.P131 Although this is a natural line of thought, Dworkin would regard it as resting on a fundamental confusion about the vulnerability of our moral beliefs. For Dworkin holds that our moral beliefs are not susceptible to being rationally undermined by purely psychological hypotheses about why we hold them. Rather, undermining a moral view always requires one to engage in first-order moral reasoning and argument. The same view has been endorsed by Thomas Nagel in *The Last Word*.³⁵ Let's look at the case for it.

C5.S10

5.7.3 “Hume’s Principle”

C5.P132 As we've noted, Dworkin holds that moral beliefs differ from ordinary empirical beliefs in at least two significant respects. First, the fact that a given moral belief is true never appears in the explanation of why it's held. Second, the general lack of explanatory connection between moral truth and moral belief does not cast any doubt on our moral beliefs.³⁶ Given that by Dworkin's own lights the absence of an explanatory connection between a belief's truth and its being held tends to undermine the credibility of an ordinary empirical belief, why wouldn't the same hold for our moral beliefs?

C5.P133 For Dworkin, this difference between ordinary empirical beliefs and moral beliefs is a consequence of an even more fundamental difference between

³⁵ This is a central claim of chapter 6 of that work.

³⁶ Dworkin helpfully locates himself in dialectical space by noting that the first view is one that he shares with a certain kind of “external skeptic” about morality, but that he and the skeptic part ways with respect to the second view (2011: 70).

them. In the case of an ordinary empirical belief, information about its etiology can undermine its credibility. (When I discover that I only believe what I do about the weather because I've been hypnotized to hold those beliefs, this tends to undermine their credibility.) However, Dworkin holds that simply learning information about the etiology of a moral belief is not sufficient to undermine its credibility in the same way. This is because the undermining of a moral belief is something that always requires *moral considerations*; merely learning purely descriptive information about the causal history that accounts for why the belief is currently held is never enough to do the job on its own.³⁷

C5.P134 This is a remarkable view. Imagine that the hypnotist has found altering my beliefs about the weather so enjoyable that he decides to have some additional fun with my moral beliefs. At some later time, I'm presented with a complete list of the things that I've been hypnotized to believe; some of these beliefs concern the weather, others concern moral issues. When I discover that one of my beliefs about the weather is on the list, this undermines the credibility of that belief. But when I find that one of my moral beliefs is on the list, this discovery does *not* similarly undermine its credibility. For anything that could undermine the credibility of my moral conviction would itself have to be or include a moral consideration, and the fact that I hold this belief because I've been hypnotized is not a moral consideration but rather a fact about my psychology.

C5.P135 Consider also traditional debunking explanations of our moral beliefs, of the kind offered by Nietzsche, Marx and Freud. These accounts of why we hold the moral beliefs that we do are not themselves moral or normative or evaluative claims; rather, they purport to be purely descriptive, naturalistic, causal-explanatory accounts of why we end up holding the moral convictions that we do, and why these convictions currently strike us as correct. One might think that these causal-explanatory hypotheses are implausible, or that we lack good evidence that they are true. However, if our moral views are not susceptible to being undermined by the provision of purely non-moral information, then even if we *knew with certainty* that one of these accounts is correct, this would have no tendency to undermine the credibility of our moral convictions. Perhaps there are other possible explanations of our moral convictions that, if known to be true, would seem to cast those convictions in an even worse light than the stories offered by Nietzsche, Marx, and Freud. But if

³⁷ “[A]ny argument that either supports or undermines a moral belief must include or presuppose further moral claims or assumptions” (2011:100). This is a major theme of both Dworkin's (1996) and Part I of his (2011).

Dworkin is correct, then we have a kind of a priori guarantee that nothing that we discover empirically could count as a good reason to lose confidence in our moral convictions, so long as what is discovered is not itself a moral consideration.

C5.P136 Why does Dworkin think that our moral convictions could not be rationally undermined by empirical discoveries? Ironically, Dworkin thinks that the relative immunity of our moral beliefs is guaranteed by the truth of (what he calls) *Hume's principle*, according to which “no amount of empirical discovery about the state of the world... can entail any conclusion about what ought to be without a further premise about what ought to be” (2011: 17). As Dworkin notes, Hume's principle has traditionally been thought to bolster the case for moral skepticism, but he contends that in fact, it undermines the case for such skepticism.³⁸ His thought is this: rationally undermining a moral conviction requires an argument that targets that conviction, but Hume's principle guarantees that the content of moral and non-moral claims is so different that any such argument whose premises consist of non-moral claims will inevitably fail to make contact with its intended target. Therefore, the discovery of purely non-moral information could never provide the premises of a successful skeptical argument.

C5.P137 However, this line of thought neglects the distinction between undercutting and rebutting defeaters. As emphasized in section 5.6, a rebutting defeater for one's belief that p undermines one's justification for believing p by providing a reason to believe not-p, while an undercutting defeater undermines one's justification for believing p *without* providing a reason to believe not-p. Notice that, generally speaking, debunking explanations of our moral beliefs are best understood as attempts to provide *undercutting* as opposed to *rebutting* defeaters. For example, someone who suggests that once we fully understand the evolutionary explanation for why our moral beliefs strike us as correct, we will have good reason to abandon those beliefs, is not suggesting that the same considerations amount to reasons to endorse the opposite moral beliefs (e.g. the belief that affirmative action is fair as opposed to the belief that it is unfair, or vice versa). Rather, she's suggesting that we should hold no moral beliefs at all.

C5.P138 On the other hand, Dworkin's insistence that Hume's principle ensures that our moral beliefs cannot be rationally undermined by arguments from

³⁸ On Hume's principle and its anti-skeptical thrust, see especially (2011: 44–6). For an excellent discussion of Dworkin's use of Hume's principle, see Shafer-Landau (2010). Smith (2010) provides a useful critique of Dworkin's conception of moral skepticism.

non-moral premises seems to depend on the idea that arguments that challenge the rationality of our moral beliefs are best understood as attempts to provide *rebutting* defeaters for those beliefs. When we focus on arguments that purport to provide rebutting defeaters for our moral beliefs, then Hume's principle (assuming for the sake of argument that it's true) does seem at least potentially relevant. For example, if I initially believe that affirmative action is unfair, then the most straightforward way in which you could rationally undermine my belief is by presenting me with a compelling argument for the conclusion that affirmative action *is* fair; if you succeed in doing this, then you've succeeded in providing me with a rebutting defeater for my original belief. Plausibly, an argument for the conclusion that affirmative action is fair will have some moral claim or claims among its premises. And it might be argued that this is generally true: in order to provide a rebutting defeater for a moral belief, one must at some point appeal to moral considerations. Providing purely non-moral information will not be sufficient, for the non-moral information will not warrant taking up the opposite moral conclusion. Thus, it's at least arguable that Hume's principle does entail that *this* kind of undermining cannot be a matter of purely empirical, non-moral discoveries concerning (for example) the origins of our moral convictions, or why those convictions currently strike us as correct.

C5.P139 But not all defeaters are rebutting defeaters. There's a second way in which you might undermine my belief that affirmative action is unfair: by providing me with information that suggests that I'm not in a position to form a reliable opinion about affirmative action one way or the other. To the extent that you succeed in doing this, you've provided me with an undercutting defeater for my belief. However, precisely because you do *not* need to provide me with an argument for a moral conclusion in order to do this, there is no plausible requirement that the considerations that you offer have to include any moral considerations. (For example, it would be enough to show that I arrived at my opinion under the influence of a pill that severely impairs my ability to think about complicated issues.) The point is a general one about undercutting defeaters: a successful undercutting defeater need not stand in any logical relation to the content of the belief that it undermines.³⁹ That is why

³⁹ For example, suppose that I perform some non-trivial mathematical calculation and arrive at a particular number for an answer. Because I know that I'm generally competent when it comes to calculations of the relevant kind, I'm rationally confident in believing a certain mathematical proposition. I then learn that while performing the calculation I was under the influence of a drug that interferes with my ability to think coherently, albeit in imperceptible ways. The information that *I performed the calculation under the influence of a mind-altering drug* undermines my belief in the relevant mathematical proposition, although it is not itself a mathematical proposition.

non-moral considerations can serve as undercutting defeaters for moral beliefs, even if Hume's principle is true.

C5.P140 In sum: debunking explanations of our moral beliefs are best understood as attempts to provide undercutting as opposed to rebutting defeaters for our moral beliefs, but Hume's principle only poses a *prima facie* obstacle for rebutting as opposed to undercutting defeaters of our moral beliefs. Therefore, pace Dworkin, Hume's principle is not relevant to traditional debunking explanations of our moral beliefs. We have no *a priori guarantee* that evidence will not emerge about why we hold our moral views that would undermine the rationality of continuing to endorse those views. Our moral beliefs, like our ordinary empirical beliefs, are susceptible to being debunked if the right kind of evidence emerges.⁴⁰

C5.P141 Is there any significant truth in the vicinity of Dworkin's claims about the normative irrelevance of psychological hypotheses about why we hold the moral views that we do? Perhaps the significant truth in the vicinity is this: even if we learned that the true story about why we originally came to hold our current moral views is not epistemically respectable, this would not amount to a sufficient reason to abandon those views. For it might be that there are nevertheless good reasons to think that those views are true (reasons that played no psychological role in our originally coming to hold them), and if we're currently in a position to recognize those reasons, it would obviously be a mistake to give up the views that they support. Instead, we should continue to hold those moral views on the basis of the compelling reasons that support them. In short, the important truth in the vicinity of Dworkin's claims is simply that the genetic fallacy is indeed a fallacy when it comes to our moral views. But crucially, and contrary to what Dworkin claims, there is no

The possibility of such cases shows what is wrong with Nagel's remark that "someone who abandons or qualifies his basic methods of moral reasoning on historical or anthropological grounds alone is nearly as irrational as someone who abandons a mathematical belief on other than mathematical grounds" (Nagel 1997: 105). One *can* have good non-mathematical grounds for abandoning a mathematical belief, and the way in which this is possible is suggestive of how historical or anthropological findings could in principle give one a reason to abandon or qualify one's methods of moral reasoning.

⁴⁰ A final question: would the truth of Hume's principle at least guarantee that non-moral, empirical discoveries are ineligible to serve as *rebutting* defeaters for our moral beliefs? Not if the substantive conclusions argued for earlier in this book are correct. For example, there are circumstances in which my discovering that you hold a certain moral view would undermine my justification for holding the opposite view; in these circumstances, the fact that you believe as you do counts as a rebutting defeater for my belief. But the fact that you hold a certain moral view is not itself a moral fact. Therefore, there are cases in which non-moral, empirical discoveries about the world provide rebutting defeaters for moral beliefs. Of course, the key fact here is that a consideration can be a good reason to believe a proposition even if it is not a logically conclusive reason to believe that proposition. (Cf. Shafer-Landau 2010: 485–6).

difference here between our moral views and our views about anything else. If we learned that we originally came to hold certain beliefs about the physical world on the basis of bad reasons or for no reason at all, it wouldn't follow that we should abandon those beliefs now; for the relevant possibility is consistent with the possibility that we currently have good reason to think that those beliefs are true.

C5.P142 In this respect, as in many others, our moral beliefs and our empirical beliefs about the physical world are in the same boat.

C6

6

Conclusions

C6.P1 In the previous chapter, I defended the possibility of losing moral knowledge in ways that others have denied are genuine possibilities. Since it's generally agreed that it's possible to lose ordinary empirical knowledge in any of those ways, the arguments of that chapter, like many of the arguments offered elsewhere in the book, support the working hypothesis that I endorsed in the introduction. However, as I noted there, my primary aim is not to vindicate a general, overarching thesis about morality or our cognitive relationship to it. Rather, my primary aim is to further the discussion of the more specific issues that I address. I want to conclude by summarizing the main claims that I've endorsed. The claims are organized thematically, along with references to the specific sections and chapters in which they are discussed.

C6.S1

6.1 General Theses and Methodological Assumptions

C6.P2 (1) We can gain moral knowledge in any of the ways in which we gain ordinary empirical knowledge. (This is the positive half of the “working hypothesis” first introduced in section 1.1; supporting considerations are offered in Chapters 2–4.)

C6.P3 (2) We can lose moral knowledge in any of the ways in which we lose ordinary empirical knowledge. Attempts to acquire it are susceptible to frustration in all of the same ways. (This is the negative half of the “working hypothesis” first introduced in 1.1; supporting considerations are offered in Chapter 5.)

C6.P4 (3) Our access to moral knowledge has an important social dimension. (See especially 1.3, Chapter 3, and also 4.3.5 and 4.4.)

C6.P5 (4) In the absence of some compelling reason for thinking otherwise, we should assume that the *standards* (as opposed to the sources) of knowledge do not vary from domain to domain (1.2).

C6.P6 (5) In testing whether a given kind of state (e.g. “considered judgments about morality” or “intellectual seemings”) can play the epistemological role assigned to it by a given theory, we should not restrict our attention to judgments

or seemings that we ourselves share. (This methodological principle is explicitly defended in 2.4 and plays a key role in both Chapters 2 and 4.)

- C6.P7 (6) *The Knowledge Platitude*: If you know something that is relevant to a question that you are trying to answer, then you should take that information into account in arriving at a view (2.2, 2.4).

C6.S2

6.2 Reflective Equilibrium and Coherence as a Source of Moral Knowledge

- C6.P8 (7) As it is standardly characterized, the method of reflective equilibrium embodies genuine and important insights about moral epistemology and moral inquiry. (See (8)–(12) below.)

- C6.P9 (8) In moral inquiry, we never start from scratch: we typically come to the table already holding any number of substantive moral convictions, and there is no general requirement to set aside or bracket those convictions. It would be a methodological mistake to attempt to pursue moral inquiry from some more austere starting point. (See esp. 2.2.)

- C6.P10 (9) We can often improve our moral views by attempting to make them more coherent and systematic. Given that we have at least some substantial moral knowledge, the process of seeking greater coherence and systematicity among our moral views is likely to be a fruitful way of extending this knowledge (see esp. 2.6).

- C6.P11 (10) Contrary to what many philosophers have supposed, considerations having to do with the levels of generality of our moral judgments are not of deep methodological significance (2.8). Two specific respects in which this is true are captured by the principles (11) *Diversity* and (12) *No Privilege*.

- C6.P12 (11) *Diversity*: There is no specific level of generality that is common to all of the propositions that make up the proper starting point for theorizing about morality (2.8).

- C6.P13 (12) *No Privilege*: When conflicts arise among moral judgments that differ in their level of generality, the mere fact that the conflicting judgments have the levels of generality that they do is not itself a good reason for resolving the conflict in one way rather than another (2.8).

- C6.P14 (13) The closest possible world in which we are unreliable even with respect to those lower level moral judgments that strike us as clearly true is not a world in which we are nevertheless entitled to have justified confidence in those general moral principles that seem true to us. Rather, they are possible worlds

in which we should not have much confidence in the moral judgments that we find ourselves disposed to make at *any* level of generality (2.8).

C6.P15 (14) Although it embodies genuine insights, the method of reflective equilibrium does not have the epistemological significance that has been claimed for it. (See (15)–(17) below.)

C6.P16 (15) An ordinary moral agent might have substantial moral knowledge even if she never engages in reflective equilibrium reasoning. Neither the pursuit nor the achievement of reflective equilibrium is necessary for having such knowledge (2.2).

C6.P17 (16) Neither the pursuit nor the achievement of reflective equilibrium is necessary for having *justified moral beliefs*. Thus, the method of reflective equilibrium is not tenable when it is understood as a general account of the circumstances in which our moral views are justified (2.2).

C6.P18 (17) The method of reflective equilibrium is untenable when taken as a normative account of how we should make up our minds about moral questions. Given standard formulations, the method is consistent with a great deal of good reasoning about morality, but it is equally consistent with a great deal of bad reasoning about morality: it *undercharacterizes* good moral reasoning (2.4 and 2.7).

C6.P19 (18) (17) is true in part because taking “considered judgments” (as standardly characterized) to be the correct starting points for moral inquiry yields an account that is both overly inclusive and overly exclusive (2.4).

C6.P20 (19) On its most defensible interpretations, the method of reflective equilibrium is characterized in terms of moral judgments that have some substantive positive epistemic status: the starting points are pieces of moral knowledge, or at least, things that we justifiably believe *prior* to engaging in the method of reflective equilibrium. Thus, on its most defensible interpretations, the method of reflective equilibrium takes for granted that we have some moral knowledge (or at least, justified moral beliefs) prior to engaging in reflective equilibrium reasoning (2.5).

C6.P21 (20) Because reflective equilibrium reasoning can deliver new moral knowledge only if we have moral knowledge from other sources, the method is not a promising answer to the epistemological challenge of how we can acquire moral knowledge in the first place (Chapter 2).

C6.P22 (21) Although we should accept the method of reflective equilibrium when it is interpreted *modestly*, we should reject it when it is interpreted *ambitiously*. (Chapter 2; for the distinction between “modest” and “ambitious” interpretations of the method of reflective equilibrium, see esp. 2.1.)

- C6.P23 (22) In some respects, the method of reflective equilibrium is a promising account of how we should pursue moral *understanding*, where moral understanding is taken to be the successful subsumption of lower level moral judgments under true principles that account for the truth of those judgments (3.4).
- C6.P24 (23) In the moral domain as well as in other domains, when a person successfully acquires new knowledge by attempting to make her views more coherent, prior knowledge typically plays a role in this process (2.6).
- C6.P25 (24) Although prior knowledge typically plays a role in cases in which a person acquires new knowledge via the pursuit of coherence, there are atypical cases in which the pursuit of coherence leads to new knowledge in a way that does not essentially involve any prior knowledge (2.6).
- C6.P26 (25) When a moral philosopher succeeds in gaining moral insight or knowledge by critically reflecting on her moral views, the description that the reflective equilibrium theorist would offer of this process is often *a* correct description. But typically, this description undercharacterizes the reasoning in theoretically significant ways. This might explain the method's popularity among moral philosophers as an account of their "best practices," in spite of its limitations (2.7).

C6.S3

6.3 Social Aspects of Moral Knowledge

- C6.P27 (26) In the usual case, the fact that a person will have at least some substantial moral knowledge by the time she is in a position to engage in anything like reflective equilibrium reasoning is guaranteed by the fact that she will have absorbed some such knowledge from her social environment (Chapter 3, see esp. 3.1).
- C6.P28 (27) *The Moral Inheritance View*: One way of gaining moral knowledge is by adopting beliefs of those around us. Being able to justify beliefs acquired in this way by citing the reasons is not a necessary condition for having moral knowledge (3.1).
- C6.P29 (28) Even if one lacks the ability to justify a belief, one might still be justified in holding it, in any sense in which *being justified* is a necessary condition for knowledge (3.1).
- C6.P30 (29) The epistemologically interesting issues about moral testimony are really questions about a broader phenomenon, namely moral deference (3.2).
- C6.P31 (30) There are philosophically significant differences between moral and non-moral deference (Chapter 3, esp. 3.4).

196 MORAL KNOWLEDGE

- C6.P32 (31) In some cases, the fact that someone else holds a moral view gives us a reason to adopt that view as our own, even in cases in which this is not due to differences in non-moral information (3.2).
- C6.P33 (32) Skepticism about moral expertise is compatible with the view that there are circumstances in which even a relatively sweeping deference to another person's moral views would be appropriate (3.2).
- C6.P34 (33) Even if ethics is not the sort of subject matter about which there could be a theoretical science, that could explain neither the alleged non-existence of full-blooded moral experts, nor any of the other alleged asymmetries between moral deference and non-moral deference (3.3.1).
- C6.P35 (34) The nonexistence of full-blooded moral experts would be surprising conditional on the truth of some metaethical views and not others, but this distinction among metaethical views cross-cuts the distinctions between realism, constructivism, and noncognitivism (3.3.2).
- C6.P36 (35) Even if the core truths of morality are knowable from the armchair, this cannot explain the putative non-existence of full-blooded moral expertise, or of other putative asymmetries between moral and non-moral deference (3.3.3).
- C6.P37 (36) Attempts to account for the putative non-existence of full-blooded moral experts in terms of “the universal availability of morality” are unpromising (3.3.3).
- C6.P38 (37) Whether evidence is *opaque* or *transparent* does not in general correlate with its probative force (3.4).
- C6.P39 (38) The fact that another person judges that a proposition is true is (at best) opaque as opposed to transparent evidence (3.4).
- C6.P40 (39) When a person holds a moral view because she defers to someone she regards as a full-blooded moral expert, she typically lacks understanding of why that view is true (3.4).
- C6.P41 (40) When a person judges something to be wrong, she is expected to be able to provide some fact about it in virtue of which it is wrong. Where her belief is based on opaque evidence, she will not be in a position to discharge this burden (3.4).
- C6.P42 (41) A person who holds a moral view on the basis of opaque evidence is not in a position to fulfill an important ideal associated with moral agency: that of doing the right thing for the reasons that make it right. Moral deference does not put one in a position to fulfill this ideal even when it delivers moral knowledge (3.4).
- C6.P43 (42) Arriving at views about the reliability of others with respect to moral questions typically requires substantive moral reasoning and judgment (3.5).

C6.P44 (43) In practice, we tend to attribute unusually accurate moral judgment to people who agree with us about specific moral issues. We generally will not judge that someone has superior moral judgment if she frequently makes judgments that conflict with our own. In this respect, morality differs from various other subject matters (3.5).

C6.P45 (44) In the moral domain, we typically lack an independent basis for attributing genuine expertise comparable to the independent bases that we often have in other domains (3.5).

C6.P46 (45) Someone in a good position to recognize another person's genuine moral expertise would have relatively little need to avail herself of that expertise, compared to those who are not in a position to recognize it (3.5).

C6.P47 (46) It is natural for children to treat adult members of their community as full-blooded moral experts (3.5).

C6.P48 (47) Being a member of a moral community, and thus being in a position to compare the moral opinions of others with your own, can contribute to moral knowledge not only by affording evidence for or against your opinions, but also by providing feedback that can be used to *condition* or *calibrate* your capacity for moral judgment, so that future exercises of that judgment are more likely to deliver knowledge (4.6).

C6.S4

6.4 Experience and Observation as Sources of Moral Knowledge

C6.P49 (48) Experience and observation can contribute to moral knowledge in any of the ways in which they contribute to ordinary empirical knowledge (Chapter 4).

C6.P50 (49) Two common ways in which experience contributes to moral knowledge are *enabling* and *triggering* (4.2).

C6.P51 (50) Non-moral observations can sometimes confirm a moral claim *directly*, as opposed to confirming it by way of confirming some non-moral claim that helps to ground its truth (4.3).

C6.P52 (51) The possibility of confirming moral claims by non-moral observations does not entail that moral claims are reducible (in any sense) to non-moral claims. (4.3.2).

C6.P53 (52) The possibility of confirming moral claims by non-moral observations does not entail that moral error theory is false (4.3.2).

C6.P54 (53) Even if the most fundamental moral truths are necessary, knowable *a priori*, and irreducibly normative, they are susceptible to being confirmed and

disconfirmed by observational evidence that delivers contingent information bearing on their credibility (4.3).

C6.P55 (54) Non-moral observation can (dis)confirm a moral belief (even when this is not by way of (dis) confirming some non-moral belief) because it can make one's overall view less coherent, and in a given case the most reasonable way to respond to this incoherence can involve adjusting confidence in a moral view as opposed to (or in addition to) adjusting confidence in a non-moral view (4.3.5).

C6.P56 (55) In the moral domain, an important source of misleading evidence is that people who are morally upstanding in other respects might engage in a practice or behavior that is seriously morally wrong, in a social context in which its wrongness is not generally recognized.

C6.P57 (56) The mechanism for generating misleading evidence in (55) can pose formidable practical difficulties for would-be moral reformers, who seek to convince other members of their society that some common behavior or practice is seriously wrong (4.3.5).

C6.P58 (57) Experience can contribute to moral knowledge by refining or honing the capacity for moral judgment to the point that subsequent exercises of that judgment are sufficiently reliable to count as knowledge (4.4).

C6.P59 (58) The hypothesis that experiential conditioning plays a significant role in influencing our moral judgments offers an understanding of fundamental moral disagreement that is superior to standard cognitivist alternatives (4.4).

C6.P60 (59) Attempts to account for basic a priori moral knowledge either in terms of intellectual seemings or in terms of self-evident propositions are vulnerable to objections (4.5).

C6.P61 (60) On an improved account of basic a priori moral knowledge, self-evident propositions are those propositions that *intellectually seem true* to a subject *because they are true*. Basic a priori moral knowledge arises when the truth of a moral claim explains why it intellectually seems true to a person who believes it on the basis of that seeming.

C6.S5

6.5 Knowing the Difference Between Right and Wrong, Ryle's Puzzle, and Losing Moral Knowledge

C6.P62 (61) Someone *knows the difference between right and wrong* only if she knows which actions are right and which actions are wrong in a suitably wide range of cases. But comprehensive knowledge of which actions are right and which are wrong is not sufficient for knowing the difference between right and wrong (5.3).

- C6.P63 (62) A person knows the difference between right and wrong only if she is *autonomously reliable* with respect to classifying actions as right or wrong (5.3).
- C6.P64 (63) Having poor moral judgment is a significant obstacle to knowing that one has poor moral judgment. In this way, having poor moral judgment differs from having a poor sense of direction or poor eyesight (5.5).
- C6.P65 (64) In a process of moral corruption, in which a person moves from a better moral sensibility to one that is worse, one typically loses the perspective required to recognize the superiority of the better moral sensibility. In this respect, there is a significant difference between physical deterioration and moral deterioration: while physical deterioration is often recognized as such by the subject who undergoes it, it is characteristic of moral deterioration to impair the ability of the subject to recognize that this is happening (5.5).
- C6.P66 (65) *Pace* Ryle, it is in principle possible to “forget the difference between right and wrong” Some cases of corruption are best understood as involving such forgetting (5.5).
- C6.P67 (66) The proposition that an individual would express by asserting

C6.P68 I used to know the difference between right and wrong, but I don’t anymore: at some point, I forgot it.

C6.P69 is typically a *blind sport proposition* for that person: even if the individual would express a true proposition by uttering the relevant sentences, he will typically not be in a position to recognize that that proposition is true, even if some third party is in a position to recognize that it is (5.5).

C6.P70 (67) The fact that there is something bizarre about judging that you have forgotten the difference between right and wrong, even though there is nothing at all bizarre about judging that you have forgotten any number of other things, can be an artifact of perspectival limitations (5.5).

C6.P71 (68) Ryle’s own proposed solution to his puzzle about “forgetting the difference between right and wrong” is inadequate (5.4, 5.5).

C6.P72 (69) Solutions to Ryle’s puzzle that depend on the putative *obviousness* of morality are inadequate because they fail to account for a first-person/third-person asymmetry (5.5).

C6.P73 (70) Moral knowledge is susceptible to being lost via both rebutting and undercutting defeat (5.6, 5.7).

C6.P74 (71) A priori moral knowledge is susceptible to being defeated by empirical considerations (5.6).

C6.P75 (72) *Pace* Dworkin and Nagel, moral beliefs can be rationally undermined by non-moral considerations (5.7).

C6.P76 (73) Questions of the form: “Is the correct explanation of why p is believed a truth-invoking explanation?” are perfectly legitimate questions, regardless of p’s *modal status*. Thus, even on the assumption that our core moral convictions are necessarily true if true at all, we can sensibly ask whether truth figures in the explanation of why we hold them; there can be empirical evidence that bears on this question (5.7.1).

C6.P77 (74) In a context in which one is attempting to *justify* a moral claim, it is typically unhelpful to appeal to its truth (even if the claim in question is in fact true). Despite this, one might nevertheless have strong reasons to care whether the correct psychological explanation of why one endorses a moral claim entails that that claim is true. This is because the most plausible explanation of why one endorses a moral claim, once all of the truth-invoking hypotheses have been ruled out or deemed implausible, might be an explanation that undermines the rationality of one’s continuing to endorse it (5.7.2).

C6.P78 (75) We have no *a priori guarantee* against the emergence evidence that would undermine our moral views. Moral beliefs, like ordinary empirical beliefs, are susceptible to being debunked if the right kind of evidence emerges (5.7).

C6.P79 The Paradox of the Preface (Makinson 1965) concerns the rationality of an author who believes each of the substantive claims presented in her book while simultaneously believing that the book contains at least some errors. In canonical statements of the paradox, the author is portrayed as though she has very strong and unopposed first-order reasons for believing each of her substantive claims; in contrast, her reasons for believing the existentially quantified proposition that there are some errors in the book are not reasons she has for doubting any particular claim, but rather higher-order considerations of a broadly inductive kind (e.g. an acknowledged history of having made mistakes in the past, together with the sheer number of claims put forward in the book). A philosopher who writes a book is likely to have not only strong inductive reasons relating to her own fallibility, but also a sense of how inconclusive many of her arguments are, how many considerations remain unaddressed, and how much might be said on the other side.¹ Although I believe that each of the claims presented here is true, I’m sure—I even *know!*—that some are false. Thus, while it would be too much to hope that the list is free from error, I do hope that one effect of presenting my claims in this way is to make it easier for potential critics to identify and demonstrate exactly where the errors lie.

¹ Interestingly, the actual author whom Makinson quotes in his original presentation of the paradox was not a philosopher or other scholar accustomed to trafficking in less-than-conclusive arguments, but a mathematician.

References

- Adler, J. 2012. "Epistemological Problems of Testimony." *The Stanford Encyclopedia of Philosophy*. ed. E. Zalta. <<http://plato.stanford.edu/archives/fall2012/entries/testimony-episprob/>>.
- Alston, W. 1989. *Epistemic Justification*. Ithaca, NY: Cornell University Press.
- Aquinas, T. 1947. *Summa Theologica*. 3 vols, trans. Fathers of the English Dominican Province. New York: Benziger Brothers.
- Armstrong, D. 1999. "A Naturalist Program: Epistemology and Ontology." *Proceedings and Addresses of the American Philosophical Association* 73.2.
- Audi, R. 1997. *Moral Knowledge and Ethical Character*. Oxford: Oxford University Press.
- Audi, R. 2003. *Epistemology* (2nd edition) New York: Routledge.
- Audi, R. 2005. *The Good In the Right: A Theory of Intuition and Intrinsic Value* Princeton, NJ: Princeton University Press.
- Audi, R. Forthcoming. "Understanding, Self-Evidence, and Justification." *Philosophy and Phenomenological Research*.
- Ayer, A. 1936. *Language, Truth, and Logic*. New York: Dover.
- Ayer, A. 1959. *Logical Positivism*. New York: MacMillan.
- Ayers, M. 1993. *Locke*. New York: Routledge.
- Bealer, G. 1998. "Intuition and the Autonomy of Philosophy." *Rethinking Intuition*. eds. M. DePaul and W. Ramsey. Lanham, MD: Rowman and Littlefield.
- Bealer, G. 2002. "Modal Epistemology and the Rationalist Renaissance." *Conceivability and Possibility*. eds. T. Gendler and J. Hawthorne. Oxford: Oxford University Press: 71–125.
- Berker, S. 2009. "The Normative Insignificance of Neuroscience." *Philosophy and Public Affairs* 37.4: 293–329.
- Blackburn, S. 1993. *Essays in Quasi-Realism*. Oxford: Oxford University Press.
- Blanshard, B. 1939. *The Nature of Thought*. London: Allen and Unwin.
- Bloom, P. 2013. *Just Babies: The Origins of Good and Evil*. New York: Random House.
- Bloor, B. 1974. *Scientific Knowledge and Sociological Theory*. London: Routledge.
- Bloor, B. 1981. "The Strengths of the Strong Programme." *Philosophy of the Social Sciences* 11: 199–213.
- Boghossian, P. and Peacocke, C. 2001. *New Essays on the A priori*. Cambridge: Cambridge University Press.
- Boyd, R. 1988. "How To Be a Moral Realist." *Essays on Moral Realism*. ed. G. Sayre-McCord. Ithaca, NY: Cornell University Press.
- Brandt, R. 1979. *A Theory of the Good and the Right*. Oxford: Oxford University Press.
- Brink, D. 1989. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.
- Brogaard, B. 2014. "Intuitions as Intellectual Seemings." *Analytic Philosophy* 55.4: 382–93.
- Burge, Tyler. 1993. "Content preservation." *Philosophical Review* 102.4: 457–88.
- Byrne, Alex. 2018. *Transparency and Self-Knowledge*. Oxford: Oxford University Press.
- Carnap, R. 1963. "Intellectual Autobiography." ed. P. A. Schilpp. *The Philosophy of Rudolf Carnap*. La Salle, IL: Open Court.
- Chang, H. 2010. "The Hidden History of Phlogiston." *International Journal for Philosophy of Chemistry* 16.2: 47–79.

202 REFERENCES

- Chomsky, N. 1998. *On Language*. New York: The New Press.
- Christensen, D. 1994. "Conservatism in Epistemology." *Nous* 28.1: 69–89.
- Christensen, D. 2007. "Epistemology of Disagreement: The Good News." *Philosophical Review* 116.2: 187–217.
- Chudnoff , E. 2011a. "What Intuitions Are Like." *Philosophy and Phenomenological Research* 82.3: 625–54.Chudnoff, E. 2011b. "The Nature of Intuitive Justification." *Philosophical Studies* 153.2: 313–33.
- Coady, C. 1992. *Testimony: A Philosophical Study*. Oxford: Oxford University Press.
- Cohen, S. 1984. "Justification and Truth." *Philosophical Studies* 46: 279–96.
- Conee, E. and R. Feldman. 1998. "The Generality Problem for Reliabilism." *Philosophical Studies* 89.1: 1–29.
- Conee, E. and R. Feldman. 2004. *Evidentialism*. Oxford: Oxford University Press.
- Cuneo, T. and R. Shafer-Landau. 2014. "The Moral Fixed Points: New Directions for Moral Non-Naturalism." *Philosophical Studies* 171.3: 399–443.
- Dancy, J. 2013. "Moral Particularism." *The Stanford Encyclopedia of Philosophy*. ed. E. Zalta. <<http://plato.stanford.edu/archives/fall2013/entries/moral-particularism/>>.
- Daniels, N. 1996. *Justice and Justification: Reflective Equilibrium in Theory and Practice*. Cambridge: Cambridge University Press.
- Daniels, N. 2003. "Reflective Equilibrium." *The Stanford Encyclopedia of Philosophy*. ed. E. Zalta. <<http://plato.stanford.edu/entries/reflective-equilibrium/>>
- Davidson, D. 1986. "A Coherence Theory of Truth and Knowledge." *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. ed. E. LePore. Malden, MA: Blackwell.
- Decker, J. and D. Groll. 2013. "The (In) Significance of Moral Disagreement for Moral Knowledge." *Oxford Studies in Metaethics* 8: 140–67.
- DePaul, M. 1987. "Two Conceptions of Coherence Methods in Ethics." *Mind* 96.384: 463–81.
- DePaul, M. 1998. *Rethinking Intuition*. "Why Bother With Reflective Equilibrium?" M. DePaul and W. Ramsey. Pp. 293–309. Lanham, MD: Rowman and Littlefield.
- DePaul, M. 2006. "Intuitions in Moral Inquiry." *The Oxford Handbook of Ethical Theory*. ed. D. Copp. Oxford: Oxford University Press.
- Dreier, J. 2007. "Moral Relativism and Moral Nihilism." *The Oxford Handbook of Ethical Theory*. ed. D. Copp. Oxford: Oxford University Press.
- Dretske, F. 2000. "Entitlement: Epistemic Rights without Epistemic Duties?" *Philosophy and Phenomenological Research* 60.3: 591–606.
- Dworkin, R. 1996. "Objectivity and Truth: You'd Better Believe It." *Philosophy and Public Affairs* 25.2: 87–139.
- Dworkin, R. 2011. *Justice for Hedgehogs*. Cambridge, MA: Belknap Press of Harvard University Press.
- Easwaran, K. 2011. "Bayesianism I: Introduction and Arguments in Favor." *Philosophy Compass* 6.5: 312–20.
- Ebertz, R. 1993. "Is Reflective Equilibrium a Coherentist Model?" *Canadian Journal of Philosophy* 23.2: 193–214.
- Elga, A. 2007. "How to Disagree About How to Disagree." *Disagreement*. eds. T. Warfield and R. Feldman. Oxford: Oxford University Press: 175–86.
- Egan, A. and A. Elga. 2005. "I Can't Believe I'm Stupid." *Philosophical Perspectives* 19: 77–94.
- Enoch, D. 2014. "A Defense of Moral Deference." *The Journal of Philosophy* 61.5: 229–58.
- Enoch, D. 2018. "Non-naturalist Realism in Metaethics." *The Routledge Handbook of Metaethics*. eds. T. McPherson and D. Plunkett. New York: Routledge: 29–42.

- Estlund, D. 2008. *Democratic Authority: A Philosophical Framework*. Princeton, NJ: Princeton University Press.
- Feldman, R. 1989. "Review of Gilbert Harman's 'Change in View.'" *Philosophical Review* 98.4: 552–6.
- Feldman, R. 2003. *Epistemology*. Upper Saddle River, NJ: Prentice Hall.
- Feldman, R. 2006. "Epistemological Puzzles About Disagreement." *Epistemology Futures*. ed. S. Hetherington. Oxford: Oxford University Press: 216–36.
- Field, H. 1980. *Science Without Numbers*. Princeton, NJ: Princeton University Press.
- Foot, P. 1967. "The Problem of Abortion and the Doctrine of the Double Effect." *Oxford Review* 5. Included in Foot, 1977/2002, *Virtues and Vices and Other Essays in Moral Philosophy* Killing, Letting Die, and The Trolley Problem.
- Franklin, A. 1997. "Calibration." *Perspectives on Science* 5: 31–80.
- Freud, S. 1929/1961. *Civilization and Its Discontents*. New York: Norton.
- Fricker, E. 2006. "Testimony and Epistemic Autonomy." *The Epistemology of Testimony*. eds. J. Lackey and E. Sosa. Oxford: Oxford University Press.
- Gert, B. 2004. *Common Morality: Deciding What to Do*. Oxford: Oxford University Press.
- Gibbard, A. 1990. *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.
- Gill, M. 2007. "Moral Rationalism vs. Moral Sentimentalism: Is Morality More Like Math or Beauty?" *Philosophy Compass* 2: 16–30.
- Glymour, C. 1980. *Theory and Evidence*. Princeton, NJ: Princeton University Press.
- Goldman, A. 1967. "A Causal Theory of Knowing." *The Journal of Philosophy* 64.12: 357–72.
- Goldman, A. 1979. "What is Justified Belief?" *Justification and Knowledge*. ed. G. Pappas. Dordrecht, Holland: Reidel Publishing Company.
- Goldman, A. 1986. *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.
- Goldman, A. 2015. "Social Epistemology" <https://plato.stanford.edu/entries/epistemology-social/>
- Goldman, A. 2009. "Internalism, Externalism, and the Architecture of Justification." *The Journal of Philosophy* 106.6: 309–38.
- Goldman, A. 2012. *Reliabilism and Contemporary Epistemology: Essays*. New York: Oxford University Press.
- Goldman, A. and T. Blanchard. 2016. "Social Epistemology", *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), <<https://plato.stanford.edu/archives/win2016/entries/epistemology-social/>>.
- Goldman, A. and D. Whitcomb. 2011. *Social Epistemology: Essential Readings*. New York: Oxford University Press.
- Goodman, N. 1953. *Fact, Fiction, and Forecast*. Harvard, MA: Harvard University Press.
- Greene, J. 2003. "From Neural 'Is' to Moral 'Ought': What are the Moral Implications of Neuroscientific Moral Psychology?" *Nature Reviews Neuroscience* 4: 847–50.
- Greene, J. and J. Haidt. 2002. "How (and Where) Does Moral Judgment Work?" *Trends in Cognitive Sciences* 6: 517–23.
- Green, M.S. and J.N. Williams (eds.). 2007. *Moore's Paradox*. Oxford: Oxford University Press.
- Greene, J., S. Morelli, K. Lowenberg, L. Nystrom, and J. Cohen. 2008. "Cognitive Load Selectively Interferes with Utilitarian Moral Judgment." *Cognition* 107: 1144–54.
- Greene, J., R. Sommerville, L. Nystrom, J. Darley, and J. Cohen. 2001. "An fMRI Investigation of Emotional Engagement in Moral Judgment." *Science* 293: 2105–8.
- Groll, D. and J. Decker. 2014. "Moral Testimony: One of These Things Is Just Like the Others." *Analytic Philosophy* 54.4: 54–74.
- Hare, R. 1962. *Freedom and Reason*. Oxford: Oxford University Press.

204 REFERENCES

- Hare, R. 1973. "Rawls' Theory of Justice." *Philosophical Quarterly* 23: 144–155; 241–51.
- Harman, E. 2015. "Is it Reasonable to 'Rely on Intuitions' in Ethics?" *Norton Introduction to Philosophy*, eds. G. Rosen, A. Byrne, J. Cohen, and S. Shiffrin. New York: Norton.
- Harman, G. 1965. "The Inference to the Best Explanation." *Philosophical Review* 74: 88–95.
- Harman, G. 1967. "Enumerative Induction as Inference to the Best Explanation." *The Journal of Philosophy* 64: 529–33.
- Harman, G. 1968. "Knowledge, Inference, and Explanation." *American Philosophical Quarterly* 5: 164–73.
- Harman, G. 1973/2015. *Thought*. Princeton, NJ: Princeton University Press.
- Harman, G. 1977. *The Nature of Morality: An Introduction to Ethics*. Oxford: Oxford University Press.
- Harman, G. 1986a. *Change in View*. Cambridge: MIT Press.
- Harman, G. 1986b. "Moral Explanations of Natural Facts—Can Moral Claims Be Tested Against Moral Reality?" *The Southern Journal of Philosophy* XXIV: 57–68.
- Harman, G. 2004. "Three Trends in Moral and Political Philosophy." *The Journal of Value Inquiry* 37: 415–25.
- Hempel, C. 1934. "On the Logical Positivists' Theory of Truth." *Analysis* 2: 49–59.
- Hills, Alison. 2009. "Moral Testimony and Moral Epistemology." *Ethics* 120.1: 94–127.
- Hopkins, R. 2007. "What is Wrong with Moral Testimony?" *Philosophy and Phenomenological Research* 74.3: 611–34.
- Horwich, P. 1982. *Probability and Evidence*. Cambridge: Cambridge University Press.
- Howell, R. 2014. "Google Morals, Virtue, and the Asymmetry of Deference." *Noûs* 48.3: 389–415.
- Huemer, M. 2001. "Skepticism and the Veil of Perception." Lanham, MD: Rowman and Littlefield.
- Huemer, M. 2005. *Ethical Intuitionism*. New York: Palgrave McMillan.
- Huemer, M. 2007. "Compassionate Phenomenal Conservatism." *Philosophy and Phenomenological Research* 74: 30–55.
- Huemer, M. 2008. "Revisionary Intuitionism." *Social Philosophy and Policy* 25.1: 368–92.
- Hurka, T. 2014. *British Ethical Theorists From Sidgwick to Ewing*. Oxford: Oxford University Press.
- Jacobson, D. 2005. "Seeing by Feeling: Virtues, Skills, and Moral Perception." *Ethical Theory and Moral Practice* 8.4: 387–409.
- James, W. 1995. *The Correspondence of William James*, Vol. 3. Charlottesville, VA: University of Virginia Press.
- Joyce, R. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press.
- Joyce, R. 2016. "Evolution and Moral Naturalism." *The Blackwell Companion to Naturalism*. ed. K.J. Clark. New York: Cambridge University Press: 369–85.
- Kagan, S. 1989. *The Limits of Morality*. Oxford: Oxford University Press.
- Kagan, S. 1998. *Normative Ethics*. Boulder, CO: Westview Press.
- Kagan, S. 2001. "Thinking about Cases." *Social Philosophy and Policy* 18: 2.
- Kant, I. 1785/1981. *Groundwork for the Metaphysics of Morals*. Indianapolis, IN: Hackett.
- Kearns, S. and D. Star. 2009. "Reasons as Evidence." *Oxford Studies in Metaethics* 4: 215–42.
- Kelly, T. 2005a. "The Epistemic Significance of Disagreement." *Oxford Studies in Epistemology* 1: 167–96.
- Kelly, T. 2005b. "Moorean Facts and Belief Revision, or Can the Skeptic Win?" *Philosophical Perspectives* 19.1: 179–209.
- Kelly, T. 2008. "Common Sense as Evidence: Against Revisionary Ontology and Skepticism." *Midwest Studies in Philosophy* 32.1: 53–78.

- Kelly, T. 2010. "Peer Disagreement and Higher Order Evidence". *Disagreement*. eds. R. Feldman and T. Warfield. Oxford: Oxford University Press.
- Kelly, T. 2011. "Following the Argument Where It Leads". *Philosophical Studies* 154: 105–124.
- Kelly, T. 2013. "Disagreement and the Burdens of Judgment". *The Epistemology of Disagreement: New Essays*. eds. D. Christensen and J. Lackey. Oxford: Oxford University Press: 31–53.
- Kelly, T. 2016. "Historical versus Current Time Slice Theories in Epistemology." *Goldman and His Critics*. eds. B. McGlaughlin and H. Kornblith. Malden, MA: Wiley Blackwell: 43–65.
- Kelly, T. and S. McGrath. 2010. "Is Reflective Equilibrium Enough?" *Philosophical Perspectives* 24: 325–59.
- Kim, J. 1981. "The Role of Perception in *A Priori* Knowledge: Some Remarks." *Philosophical Studies* 40: 339–54.
- King, N. 2011. "McGrath on Moral Knowledge." *Journal of Philosophical Research* 36: 219–33.
- Korsgaard, C. 2008. "Realism and Constructivism in Twentieth-Century Moral Philosophy." *The Constitution of Agency*. New York: Oxford University Press.
- Kripke, S. 1972/1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Lackey, J. 1999. "Testimonial Knowledge and Transmission." *Philosophical Quarterly* 49.197: 471–90.
- Lackey, J. 2010. "Testimonial Knowledge." *The Routledge Companion to Epistemology*. eds. S. Bernecker and D. Prichard. New York: Routledge.
- Lance, M., V. Strahovnik, and M. Potrc, eds. 2008. *Challenging Moral Particularism*. New York: Routledge.
- Laudan, L. 1996. *Beyond Positivism and Relativism: Theory, Method, and Evidence*. Boulder, CO: Westview Press.
- Lerner, A. and S. Leslie. 2013. "Generics, Generalism, and Reflective Equilibrium: Implications for Moral Theorizing from the Study of Language." *Philosophical Perspectives* 27.1: 366–403.
- Lewis, D. 1983. "Introduction." *Philosophical Papers, Volume 1*. Oxford: Oxford University Press.
- Lewis, D. 2001. "Elusive Knowledge." *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.
- Lipton, P. 1991. *Inference to the Best Explanation*. New York: Routledge.
- Locke, J. 1979/1685. *An Essay Concerning Human Understanding*. ed. P. Nidditch. Oxford: Oxford University Press.
- Lyons, D. 1975. "Nature and Soundness of the Contract and Coherence Arguments." *Reading Rawls*. ed. N. Daniels. New York: Basic Books.
- Mackie, J. 1977. *Ethics: Inventing Right and Wrong*. New York: Penguin.
- Makinson, D.C. 1965. "The Paradox of the Preface." *Analysis* 25.6: 205–7.
- Markovits, J. 2010. "Acting for the Right Reasons." *Philosophical Review* 119.2: 201–42.
- Markovits, J. 2012. "Saints, Heroes, Sages, and Villains." *Philosophical Studies* 158.2: 289–311.
- Marx, K. 1867/1972. *Capital, vol. 1: A Critique of Political Economy*. New York: Penguin.
- McDowell, J. 1979. "Virtue and Reason." *The Monist* 62.3: 331–50.
- McDowell, J. 1980. "Meaning, Communication, and Knowledge." *Philosophical Subjects: Essays Presented to P.F. Strawson*. ed. Z. van Straaten. Oxford: Clarendon Press.
- McGrath, S. 2004. "Moral Knowledge by Perception." *Philosophical Perspectives* 18.1: 209–28.
- McGrath, S. 2008. "Moral Disagreement and Moral Expertise." *Oxford Studies in Metaethics: Volume 4*. ed. R. Shafer-Landau. Oxford: Oxford University Press: 87–108.

206 REFERENCES

- McGrath, S. 2009. "The Puzzle of Pure Moral Deference." *Philosophical Perspectives* 23.1: 321–44.
- McGrath, S. 2010a. "Skepticism about Moral Expertise as a Puzzle for Moral Realism." *Journal of Philosophy* 108.3: 111–37.
- McGrath, S. 2010b. "Moral Realism without Convergence." *Philosophical Topics* 38.2: 59–90.
- McGrath, S. 2011. "Moral Knowledge and Experience." *Oxford Studies in Metaethics* 6: 107–25.
- McGrath, S. 2014. "On Hastings Rashdall's 'The Limits of Casuistry.'" *Ethics* 125.1: 196–9.
- McGrath, S. 2018. "Moral Perception and its Rivals." *Evaluative Perception*. eds. R. Cowan and A. Bergkvist. Pp. 161–82. Oxford: Oxford University Press.
- McGrath, S. and T. Kelly. 2010. "Is Reflective Equilibrium Enough?" *Philosophical Perspectives*, 24: 325–359.
- McGrath, S. and T. Kelly. 2015. "Soames and Moore on Method in Ethics and Epistemology." *Philosophical Studies* 172: 1661–70.
- McMahan, J. 2000. "Moral Intuition." *The Blackwell Guide to Ethical Theory*. ed. H. LaFollete. Malden, MA: Blackwell.
- McNaughton, D. 1988. *Moral Vision*. Malden, MA: Blackwell.
- McPherson, T. 2009. "Moorean Arguments and Moral Revisionism." *Journal of Ethics and Social Philosophy* 2.
- McPherson, T. 2015. "The Methodological Irrelevance of Reflective Equilibrium." *The Palgrave Handbook of Philosophical Methods*. ed. C. Daly. New York: Palgrave MacMillan.
- Mikhail, J. 2000. "Rawls' Linguistic Analogy: A Study of the 'Generative Grammar' Model of Moral Theory Discussed by John Rawls in *A Theory of Justice*." Cornell: Ph.D. Dissertation.
- Mikhail, J. 2007. "Universal Grammar: Theory, Evidence, and the Future." *Trends in Cognitive Sciences* 11: 143–52.
- Moore, G. 1925. "A Defense of Common Sense." *Contemporary British Philosophy (second series)*. ed. J. Muirhead. London: George Allen and Unwin.
- Moore, G.E. 1962. *Philosophical Papers*. London: Collier.
- Nagel, T. 1986. *The View from Nowhere*. Oxford: Oxford University Press.
- Nagel, T. 1997. *The Last Word*. Oxford: Oxford University Press.
- Nehamas, A. 1998. *The Art of Living*. Berkeley, CA: University of California Press.
- Neta, R. 2009. "Defeating the Dogma of Defeasibility." *Williamson on Knowledge*. eds. P. Greenough and D. Pritchard. Oxford: Oxford University Press.
- Nickel, P. 2001. "Moral Testimony and its Authority." *Ethical Theory and Moral Practice* 4.3: 253–66.
- Nietzsche, F. 1887/1967. *On the Genealogy of Morals*. New York: Vintage.
- Nozick, R. 1981. *Philosophical Explanations*. Cambridge, MA: Belknap.
- Nussbaum, M. 1990. *Love's Knowledge*. Oxford: Oxford University Press.
- Parfit, D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Pew Center Report. 2013. <https://www.peoplepress.org/2013/03/20/growing-support-for-gay-marriage-changed-minds-and-changing-demographics/>
- Plantinga, A. 1993. *Warrant: the Current Debate*. Oxford: Oxford University Press.
- Plato, Laches. (many editions)
- Plato, Protagoras. (many editions)
- Pollock, J. 1986. *Contemporary Theories of Knowledge*. Boston, MA: Rowman and Littlefield.
- Pollock, J. and Cruz, J. 1999. *Contemporary Theories of Knowledge, 2nd edition*. Boston, MA: Rowman and Littlefield.

- Pryor, J. 2000. "The Skeptic and the Dogmatist." *Noûs* 34.4: 517–49.
- Pryor, J. 2001. "Highlights of Recent Epistemology." *British Journal for the Philosophy of Science* 52.1: 95–124.
- Pust, J. 2000. *Intuitions as Evidence*. New York: Garland/Routledge.
- Putnam, H. 1981. *Reason, Truth, and History*. Cambridge: Cambridge University Press.
- Quine, W. 1951. "Two Dogmas of Empiricism." *The Philosophical Review* 60.1: 20–43.
- Quine, W. 1961. *From A Logical Point of View*. Cambridge, MA: Harvard University Press.
- Quine, W. 1979. "On the Nature of Moral Values." *Critical Inquiry* 5.3: 471–80.
- Railton, P. 1986. "Moral Realism." *Philosophical Review* 95.2: 163–207.
- Railton, P. 2003. *Facts, Values, and Norms: Essays Toward a Morality of Consequence*. New York: Cambridge University Press.
- Rashdall, H. 1894. "The Limits of Casuistry." *International Journal of Ethics* 4.4: 459–80.
- Rawls, J. 1951. "Outline of a Decision Procedure for Ethics." *Philosophical Review* 60: 2: 177–97.
- Rawls, J. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Rawls, J. 1974. "The Independence of Moral Theory." *Proceedings and Addresses of the American Philosophical Association* 47: 5–22. (Reprinted in Rawls 1999: 286–302. Page references are to the reprinted version.)
- Rawls, J. 1980. "Kantian Constructivism in Moral Theory." *Journal of Philosophy* 77: 515–72. (Reprinted in Rawls 1999: 303–358. Page references are to the reprinted version.)
- Rawls, J. 1993. *Political Liberalism*. New York: Columbia University Press.
- Rawls, J. 1999. *Collected Papers*. ed. S. Freeman. Cambridge, MA: Harvard University Press.
- Rawls, J. 2001a. *Justice as Fairness: A Restatement*. Cambridge, MA: Harvard University Press.
- Rawls, J. 2001b. *Lectures on the History of Philosophy*. ed. B. Herman. Cambridge, MA: Harvard University Press.
- Raz, J. 1982. "The Claims of Reflective Equilibrium." *Inquiry* 25.3: 307–30.
- Reid, T. 1788/2003. *Essays on the Active Powers of the Human Mind*. Excerpt reprinted in *Moral Philosophy from Montaigne to Kant*. ed. J. Schneewind. New York: Cambridge.
- Rescher, N. 2005. *Commonsense*. Milwaukee, WI: Marquette University Press.
- Roedder, E. and G. Harman. 2010. "Linguistics and Moral Theory." *The Moral Psychology Handbook*. ed. J. Doris. Oxford: Oxford University Press.
- Rosen, G. 2010. "Metaphysical Dependence: Grounding and Reduction." Modality: Metaphysics, Logic, and Epistemology. eds. B. Hale and A. Hoffmann. Oxford: Oxford University Press.
- Rowe, D. and R. Schulmann. 2007. *Einstein on Politics*. Princeton, NJ: Princeton University Press.
- Ryle, G. 1958/2009. "On Forgetting the Difference Between Right and Wrong." Originally in *Essays in Moral Philosophy*. ed. A.I. Melden. Washington, DC: University of Washington Press. Reprinted in Ryle's Collected Papers, ii (London: Routledge). Page numbers are to the reprinted version.
- Sayre-McCord, G. 1996. "Coherentist Epistemology and Moral Theory." *Moral Knowledge?* eds. W. Sinnott-Armstrong and M. Timmons. New Readings in Moral Epistemology. Oxford: Oxford University Press.
- Scanlon, T. 2002. "Rawls on Justification." *The Cambridge Companion to Rawls*. ed. S. Freeman. Cambridge: Cambridge University Press.
- Scanlon, T. 2003. *The Difficulty of Tolerance: Essays in Political Philosophy*. Cambridge: Cambridge University Press.

208 REFERENCES

- Scanlon, T. 2014. *Being Realistic About Reasons*. Oxford: Oxford University Press.
- Schneewind, J. 1988. *The Invention of Autonomy*. New York: Cambridge.
- Schneewind, J. 2003. *Moral Philosophy from Montaigne to Kant*. New York: Cambridge.
- Schroeder, M. 2007. *Slaves of the Passions*. Oxford: Oxford University Press.
- Setiya, K. 2012. *Knowing Right From Wrong*. Oxford: Oxford University Press.
- Shafer-Landau, R. 2003. *Moral Realism: A Defense*. Oxford: Oxford University Press.
- Shafer-Landau, R. 2010. "Dworkin and the Possibility of Metaethics." *Boston University Law Review* 10: 101–17.
- Sherman, B. 2014. "Moral Disagreement and Epistemic Advantages: A Challenge to McGrath." *Journal of Ethics and Social Philosophy* 8.3.
- Sidgwick, H. 1907/1981. *The Methods of Ethics*. Indianapolis, IN: Hackett.
- Singer, P. 1972. "Moral Experts." *Analysis* 32.4: 115–117.
- Singer, P. 1974. "Sidgwick and Reflective Equilibrium." *Monist* 58: 490–517.
- Singer, P. 1988. "Ethical Experts in a Democracy." *Applied Ethics and Ethical Theory*. eds. D. Rosenthal and F. Shehadi. Salt Lake City, UT: University of Utah Press.
- Singer, P. 2005. "Ethics and Intuitions." *Journal of Ethics* 9.3–4: 331–52.
- Sliwa, P. 2012. "In Defense of Moral Testimony." *Philosophical Studies* 158.2: 175–95.
- Smith, M. 1994. *The Moral Problem*. Malden, MA: Blackwell.
- Smith, M. 2000. "Moral Realism." *The Blackwell Guide to Ethical Theory*. ed. Hugh LaFollete. Malden, MA: Blackwell.
- Smith, M. 2007. "In Defence of Ethics and the A Priori: Selected Essays on Moral Psychology and Metaethics: Reply to Enoch, Heironymi, and Tannenbaum." *Philosophical Books* 48: 136–49.
- Smith, M. 2010. "Dworkin on External Skepticism." *School of Law* 90.2: 509–20. Boston University Law Review, Boston University.
- Soames, S. 2014. *The Analytic Tradition in Philosophy, volume 1: Founding Giants*. Princeton, NJ: Princeton University Press.
- Soames, S. 2015. "Reply to Critics of the Analytic Tradition in Philosophy. Vol.1: The Founding Giants". *Philosophical Studies* 172: 1681–96.
- Sorensen, R. 1988. *Blindsights*. Oxford: Clarendon Press.
- Sosa, E. 2007. *A Virtue Epistemology*. Oxford: Oxford University Press.
- Sowell, T. 1994. *Race and Culture*. New York: Basic Books.
- Stitch, S. 1990. *The Fragmentation of Reason*. Cambridge, MA: MIT Press.
- Stratton-Lake, P. 2002. *Ethical Intuitionism: Re-evaluations*. Oxford: Oxford University Press.
- Street, S. 2006. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127.1: 109–66.
- Street, S. 2010. "What is Constructivism in Ethics and Metaethics?" *Philosophy Compass* 5: 363–84.
- Sturgeon, N. 1985. "Moral Explanations." *Morality, Reason, and Truth*. eds. D. Copp and D. Zimmerman. Pp. 49–78. Totowa, NJ: Rowman and Allanheld.
- Sturgeon, N. 1986. "Harman on Moral Explanations of Natural Facts." *Southern Journal of Philosophy* 24.1: 69–78.
- Sturgeon, N. 2002. "Ethical Intuitionism and Ethical Naturalism." *Ethical Intuitionism: Re-evaluations*. ed. P. Stratton-Lake. Oxford: Oxford University Press.
- Tersman F. 2018. "Recent Work on Reflective Equilibrium and Method in Ethics." *Philosophy Compass*. e12493. <https://doi.org/10.1111/phc3.12493>
- Thomson, J. 2013. "The Way It Was." *Portraits of American Philosophy*. ed. S. Cahn. Lanham, MD: Rowman and Littlefield.

REFERENCES 209

- Thomson, J.J. 1976. "Killing, Letting Die, and The Trolley Problem." *The Monist* 59.2: 204–17.
- Timmons, M. 1999. *Morality Without Foundations: A Defense of Ethical Contextualism*. Oxford: Oxford University Press.
- Vavova, K. 2015. "Evolutionary Debunking of Moral Realism." *Philosophy Compass* 10.2: 104–16.
- Walden, K. 2013. "In Defense of Reflective Equilibrium." *Philosophical Studies* 166: 23–256.
- Wallace, R. 1998. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press.
- Warnock, M. 1985. *A Question of Life: The Warnock Report on Human Fertilisation and Embryology*. Malden, MA: Wiley Blackwell.
- Wedgwood, R. 2007. *The Nature of Normativity*. Oxford: Oxford University Press.
- White, R. 2015. "The Problem of the Problem of Induction." *Episteme* 12.2: 275–90.
- Williams, B. 1985. *Ethics and the Limits of Philosophy*. Cambridge, MA: Harvard University Press.
- Williams, B. 1995. *Making Sense of Humanity*. Cambridge: Cambridge University Press.
- Williamson, T. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press.
- Williamson, T. 2007. *The Philosophy of Philosophy*. Malden, MA: Blackwell.
- Williamson, T. 2009. "Reply to Ram Neta." *Williamson on Knowledge*. eds. P. Greenough and D. Pritchard. Pp. 350–56. Oxford: Oxford University Press.

