

Epistemic Authority

Epistemic Authority

A Theory of Trust, Authority,
and Autonomy in Belief

Linda Trinkaus Zagzebski

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

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

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

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

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

© Oxford University Press 2012

First issued as an Oxford University Press paperback, 2015.

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 license, or under terms agreed with the appropriate reproduction rights organization. Inquiries 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.

Library of Congress Cataloging-in-Publication Data
Zagzebski, Linda Trinkaus, 1946–

Epistemic authority: a theory of trust, authority, and
autonomy in belief / Linda Trinkaus Zagzebski.
p. cm.

ISBN 978-0-19-993647-2 (hardcover : alk. paper); 978-0-19-027826-7 (paperback : alk. paper)

1. Authority. 2. Knowledge, Theory of. 3. Self.

4. Trust. 5. Belief and doubt. I. Title.

BD209.Z34 2012

121'6—dc23

2012008783

Dedicated to my sister, Rita Trinkaus Cooney

Acknowledgments xi

Introduction 1

Chapter 1 The Rejection of Epistemic Authority 4

1. Authority, Equality, and Self-Reliance in the Epistemic Realm 4
2. The Epistemological Case for Epistemic Self-Reliance 8
 - 2.1. Mistrust of Taking Beliefs from Others 8
 - 2.2. Self-Reliance and the Nature of Knowledge: Plato and Locke 12
 - 2.3. Self-Reliance and Cartesian Doubt 16
3. The Case from Ethics: Self-Reliance and Autonomy 19
4. Authority and Autonomy in the Intellectual Domain 23
5. The Value of Reflective Self-Consciousness 26

Chapter 2 Epistemic Self-Trust 29

1. The Natural Authority of the Self 29
2. The Natural Desire for Truth and the Prereflective Self 33
3. The Desire for Truth and the Reflective Self 38
4. Self-Trust and the Alternatives 43
5. The Conscientious Believer and the Nature of Reasons 45

Chapter 3 Epistemic Trust in Others 52

1. Epistemic Egoism 52
2. The Need for Trust in Others 55

- 2.1. Why Epistemic Egoism Is Unreasonable 55
 - 2.2. Epistemic Egocentrism 61
 - 3. Trust in Others and the Two Kinds of Reasons 63
 - 3.1. The Distinction between Deliberative and Theoretical Reasons 63
 - 3.2. The Two Kinds of Reasons and Parity between Self and Others 66
 - 4. Epistemic Universalism and Common Consent Arguments 69
- Chapter 4 Trust in Emotions 75
- 1. The Rational Inescapability of Emotional Self-Trust 75
 - 2. Trustworthy and Untrustworthy Emotions 84
 - 3. Admiration and Trust in Exemplars 87
 - 4. Trust in the Emotions of Others 93
 - 5. Expanding the Range of Trust 96
- Chapter 5 Trust and Epistemic Authority 99
- 1. Authority in the Realm of Belief 99
 - 1.1. Can We Believe on Authority? 99
 - 1.2. The Epistemic Authorities 103
 - 2. The Contours of Epistemic Authority: The Principles of Joseph Raz 105
 - 3. Preemption and Evidence 113
 - 4. The Value of Truth versus the Value of Self-Reliance 117
 - 5. Standard Epistemic Egoism and Epistemic Authority 119
- Chapter 6 The Authority of Testimony 120
- 1. Conscientious Testimony 120
 - 2. Testimony and Deliberative versus Theoretical Reasons 128
 - 3. Principles of the Authority of Testimony 131
 - 4. Testimony as Evidence and the Authority of Testimony 135
 - 5. The Parallel between Epistemic and Practical Authority 136
- Chapter 7 Epistemic Authority in Communities 140
- 1. Epistemic Authority and the Limits of the Political Model 140
 - 2. Authority in Small Communities 144
 - 2.1. Justifying Authority in Small Communities 144
 - 2.2. Justifying Epistemic Authority in Small Communities 149
 - 3. Communal Epistemic Authority 151
 - 4. The Epistemology of Imperfection 157

Chapter 8 Moral Authority 159

1. The *Prima Facie* Case for Moral Epistemic Authority 159
2. Skepticism about Moral Authority 162
 - 2.1. Skepticism about Moral Truth 162
 - 2.2. Moral Egalitarianism 163
 - 2.3. Autonomy 166
3. Moral Authority and the Limits of Testimony 170
 - 3.1. Emotion and Moral Belief 170
 - 3.2. Moral Authority and Understanding 174
4. Communal Moral Authority and Conscience 178

Chapter 9 Religious Authority 181

1. Religious Epistemic Egoism 181
2. Religious Epistemic Universalism 184
3. Believing Divine Testimony 189
 - 3.1. Faith and Believing Persons 189
 - 3.2. Models of Revelation 191
4. Conscientious Belief and Religious Authority 199

Chapter 10 Trust and Disagreement 204

1. The Antinomy of Reasonable Disagreement 204
2. Disagreement and Deliberative versus Theoretical Reasons 210
3. Self-Trust and Resolving Disagreement 211
4. Communal Epistemic Egoism and Disagreement between Communities 221

Chapter 11 Autonomy 229

1. The Autonomous Self 229
 - 1.1. The Norm of Conscientious Self-Reflection 229
 - 1.2. Autonomy from the Inside and the Outside 233
2. Attacks on the Possibility of Autonomy: Debunking Self-Trust 237
3. Epistemic Authority from the Outside 247
4. Self-Fulfillment 251

Bibliography 255

Name Index 269

Subject Index 273

This book began as eight Wilde Lectures in Natural and Revealed Religion, delivered in Oxford during Trinity term 2010. I am grateful for lively discussions in the ideal atmosphere of Oxford, particularly with Brian Leftow, Lizzie Fricker, Joseph Raz, Gail Fine, Richard Swinburne, and John Cottingham, and for the generous hospitality of Oriel College, as well as delightful meals at several other colleges and academic houses, including St. Benet's Hall, Campion Hall, and Blackfriar's Hall, and the home of Kelly James Clark.

In the fall 2010 I taught a graduate seminar on the manuscript at the University of Oklahoma while revising the lectures into a book. My students were extraordinarily good. In addition to helping me improve the work, they made it a real joy to teach the class. Their contributions can be found throughout this book. Several faculty members in my department read portions of the manuscript and were very helpful consultants, including Hugh Benson, Wayne Riggs, Ray Elugardo, Martin Montminy, Monte Cook, Zev Trachtenberg, and Neal Judisch.

In May 2011, I presented a version of the manuscript as nine Kaminski Lectures at the John Paul II Catholic University of Lublin, Poland. I was touched at the generous hospitality of both the faculty and students, who were attentive to my every need. My host, Agnieszka Lekka-Kowalik, was wonderful. She and Rafal Lizut were delightful tour guides as well as philosophical interlocutors. I am especially grateful to Paulina Broszkiewicz, who showed us the sights of Warsaw and took care of us in Lublin, and to Fr. Andrzej Bronk, who hosted a "fire party" at his home. Students Tomasz Łak and Łukasz Cięgetura were very helpful in making arrangements, and I thank Łukasz and his wife for a wonderful meal at their home. The

Kaminski Lectures were followed by three Olaus Petri Lectures at the University of Uppsala, Sweden, hosted by Carl Reinhold Bråkenhielm. The participants in all these events were generous with their time and their insights.

I owe a special debt of gratitude to the John Simon Guggenheim Foundation for awarding me a fellowship for academic year 2011–12 to finish the manuscript. The Guggenheim Fellowship has permitted me to complete the book under conditions that for me are ideal—at home. The book would have taken much longer to complete without the luxury of release from teaching and service duties.

At the beginning of November 2011, Michael Beaty organized a two-day symposium on the manuscript in the Department of Philosophy at Baylor University, in which six papers were presented by faculty and graduate students. I thank Ryan Byerly, Chris Tweedt, Ben McMyler, Trent Dougherty, Heidi Chamberlin, Nate Jackson, and Karl Aho for their helpful papers, as well as many members of the audience. In addition, Ben McMyler read much of the manuscript and gave me detailed comments for which I am very grateful. It was a pleasure to meet him and to engage in conversation outside of the formal part of the symposium.

I presented material from chapter 1 on the ideal of epistemic self-reliance as a public lecture at Fort Lewis College in Durango, Colorado, in October 2011, and to the Department of Philosophy at Loyola Marymount University, Los Angeles, in November 2011. Participants in both those discussions gave me very helpful suggestions that have been incorporated into the first chapter of the book.

A version of chapter 5 was presented at an epistemology conference at St. Louis University in 2009, hosted by John Greco, and I thank the participants in that conference for their feedback. In addition, I thank E. J. Coffman, who made very valuable suggestions on chapter 10, and Rodrigo Borges, who read and commented on most of the manuscript.

I particularly appreciate my hardworking research assistants: Bryan Reece, who worked on the project during the research stage, Gary Osmundsen (2009–10) and Max Parish (2010–11), who worked on the project during the writing stages, and Shyam Patwardhan (2011–12), who worked

on the project in its final stages. Samuel Robert Byer gave me assistance during the summer of 2010.

Finally, I want to thank my husband, Ken, who traveled with me when I gave all of these lectures, and who is an exemplar of patience and caring.

Norman, Oklahoma

December 29, 2011

Introduction

Everyone recognizes that some forms of authority are inescapable, but authority in the epistemic domain gets virtually no attention. Moral and political philosophers understandably focus on authority over actions, presumably assuming that authority over beliefs is a topic for epistemology. But epistemologists do not discuss it either. If they occasionally use the term “epistemic authority,” that is only by courtesy. What they mean by an epistemic authority is an expert. Rarely do we get any explanation of what makes an expert authoritative, or any attempt to connect epistemic authority with the literature on authority in moral and political philosophy.

I made the same observation about the concept of intellectual virtue some years ago (Zagzebski 1996). In both cases there is an important concept in moral philosophy that applies to the domain of intellectual inquiry and belief, but which is disregarded by both moral philosophers and epistemologists. No doubt some philosophers neglect epistemic authority unconsciously, but I think that many find the idea suspicious. Historical experience provides one kind of reason for that suspicion. The Protestant Reformation, the political turmoil of the early modern period, and the rise of modern science all contributed to shattering the idea of authority, and epistemologists are not immune to the effects of those events. Distrust of authority is now as pervasive and as invisible as trust in authority once was. I think that when we do philosophy it is important to be aware of what we trust and what we distrust because that affects the lines of inquiry we resist and the kinds of conclusions we applaud. Suspicion of epistemic authority is a natural consequence of the general suspicion of authority we have inherited.

But philosophers cannot refer to historical events to justify the rejection of epistemic authority. For that we need to turn to philosophy, and philosophy

in the Enlightenment is standardly interpreted as beginning a new focus on the self. In my opinion, one of the most significant philosophical turns of the modern period for epistemology, as well as ethics and political philosophy, was the development of the idea of autonomy. The general distrust of authority might be the cause of the rejection of epistemic authority, but its philosophical defense is its apparent incompatibility with autonomy.

The idea of autonomy is another idea that has a large literature in moral philosophy but is ignored in epistemology. If intellectual autonomy exists, we will want to see how it connects with autonomy in the moral realm and authority in any realm. If we should reject authority in the epistemic domain, there is much work to be done to explain why that is the case. If we should not reject it, then again, we need to find out what it is, why we should care about it, and what we lose if we ignore it.

One thing we lose is a way to understand the justification of a great number of religious beliefs. There have been claims to religious authority for thousands of years, and if there is religious authority, it is both moral and epistemic. More precisely, religious authority includes epistemic authority in the realm of the teaching authority of a given faith tradition, which usually includes both metaphysical and moral teachings. Some religions have an authoritative teaching structure, as does the church to which I belong, the Roman Catholic Church. But given the modern suspicion of authority, it often appears from an outside perspective that a person is unjustified in accepting epistemic authority in her community. From a perspective inside the community, authority is usually justified by reference to other beliefs that arise from the community (e.g., "Christ founded the Church," "Sharia law was revealed by God," etc.), so the justification is circular. That is not necessarily problematic for the members of the community, but it means that their beliefs are insulated from criticism from the outside. It also makes it easy for those on the outside to disregard a community's justification for its authoritative beliefs. I assume that this situation is undesirable.

In this book I will defend the existence of epistemic authority on grounds that almost all modern philosophers would accept. My argument will proceed wholly from the point of view of the subject—a self-reflective person who asks herself how she should get beliefs she accepts upon reflection. Is it ever reasonable for her to take someone as an authoritative source of her beliefs? Can she reasonably accept an authority in her community only if she already accepts what the authority proclaims? I will argue that if we

make some minimal assumptions about the self, we are all committed to accepting epistemic authority, and authority in religious communities can be defended in the same way. My argument permits both external recognition of the justification of authority in a community and external critique of that authority.

I will proceed as follows. In the first chapter I will explore the historical roots of the rejection of epistemic authority from the fields of epistemology and ethics. Autonomy in the intellectual realm has often been identified with epistemic self-reliance, and self-reliance has been treated as an ideal by epistemologists for other reasons. I find defenses of self-reliance inadequate, but our discussion of autonomy will reveal a value that is real and important: the value of self-reflective consciousness. The method of this project is to explore the implications of self-reflective consciousness, starting from a position that is neutral on the issues of authority and autonomy. Beginning in chapter 2 and continuing in the rest of the book, I will give an extended argument that the self-reflective person is committed to belief on authority, and that epistemic self-reliance is incoherent. I will argue that epistemic self-trust is both rational and inescapable, that consistent self-trust commits us to trust in others, that among those we are committed to trusting are some whom we ought to treat as epistemic authorities, and that some of these authorities can be in the moral and religious domains. I will argue that intellectual autonomy is compatible with epistemic authority, but the ideal of self-reliance must be rejected.

The Rejection of Epistemic Authority

The best man of all is he who knows everything himself,
Good also the man who accepts another's sound advice;
But the man who neither knows himself nor takes to heart
What another says, he is no good at all.

HESIOD, *quoted by Aristotle NE 1095B 11–13*
(Burnyeat 1980: 71)

1. AUTHORITY, EQUALITY, AND SELF-RELIANCE IN THE EPISTEMIC REALM

Close to a half century ago, Hannah Arendt remarked that one characteristic of the modern world is the disappearance of authority, not just in philosophy, but in the practices of modern life.¹ I doubt that we can overestimate the collapse of the idea of authority, given that it never recovered. It is impossible to live without government, so some of the most important work of early modern philosophy was devoted to reconstituting the justification of political authority in social contract theories, starting from the premise that each person has ultimate authority over himself.² Moral and epistemic authority disappeared, and religious authority was reserved for sectarian communities whose place in the public square is still disputed. Much of

1. In "What Is Authority?" Arendt says, "In order to avoid misunderstanding, it might have been wiser to ask in the title: What was—and not what is—authority? For it is my contention that we are tempted and entitled to raise this question because authority has vanished from the modern world" (Arendt 1968: 91).

2. Classic sources include Locke's *Two Treatises of Government*, Bk. II, chap. II, sec. 4; Rousseau's *Social Contract*, Bk. I, chap. 2; and Hobbes' *Leviathan*, Pt. I, chap. 14.

contemporary epistemology takes all of this for granted. It is not surprising, then, that authority in the realm of belief gets almost no attention.

Epistemologists occasionally use the term “epistemic authority” to refer to experts—people who are reliable sources of information in some domain—but that usage indirectly confirms my point. An expert is an authority in only a very weak sense since the expert and her “subject” may have no relationship with each other.³ What’s more, the existence of expertise is typically denied in any domain pertaining to value or religion. The fields in which experts are recognized are carefully circumscribed, and experts are not treated as authorities in the sense in which political authorities are authorities. Most philosophers assume that they may not command belief, and nobody has an epistemic duty to believe them.⁴

Why is it normally taken for granted that nobody may command belief? In *A Letter Concerning Toleration*, Locke gives one influential answer: it is not possible to believe on command. Locke says, “[I]t is absurd that things should be enjoined by laws which are not in men’s power to perform. And to believe this or that to be true does not depend upon our will” (Locke 209: 24). This claim has a measure of truth and no doubt has many adherents, but I do not see that it is any harder to believe on command than to believe testimony. It depends a lot on the context. If I can believe some proposition *p* when somebody says “*p*,” then surely I can believe *p* when someone says “Believe me, *p*.” But then it is but a small step to believe *p* when somebody says “*p* is true. Believe it.” In each case I know that the speaker intends for me to believe *p*, and I also know that the speaker himself believes *p*. I may not like the commanding mode of the last case, but I see no reason to think that I am unable to follow the command. If I can accept the speaker’s testimony that *p*, why would I find myself unable to do so once the speaker turns his testimony into an explicit imperative?

I think this is an important issue and I will return to it in chapters 5 and 6, but my project does not require acceptance of the claim that we can believe

3. Alvin Goldman (2001: 95) refers to persons who claim to be epistemic experts in some domain, implying that they expect their expertise to be recognized by others, but Goldman does not investigate the relationship between an expert and her epistemic subject.

4. Occasionally writers on authority will mention epistemic authority briefly. For instance, R. B. Friedman says that an epistemic authority is one who is entitled to be believed, and remarks that it should not be assumed that there is no connection between epistemic and political authority (1990: 57). Joseph Raz also alludes to intellectual authority in a couple of places. His work will be discussed in later chapters.

on command. I am primarily interested in epistemic authority from the viewpoint of the subject. Whether or not an authority has the right to command, am I ever justified in taking a belief on the authority of another person? Am I ever required to do so if I am rational? This chapter is a prelude to answering this question in the affirmative. What I want to do first is to investigate why there is a strong proclivity to answering in the negative.

The most important theoretical reason for the rejection of epistemic authority, I think, is the perceived conflict between epistemic authority and two deep modern values: egalitarianism and autonomy. Since the beginning of the modern period, most theorists have assumed that normal adult humans are roughly equal in basic human capacities, including epistemic capacities. At least, they are equal enough that the conditions for applying such epistemic concepts as knowledge, justified belief, reasonable doubt, and so on, are satisfied by any adult human taken at random. We all admit that some persons have more inherited intelligence than others, or have honed their intellectual skills to a greater degree than others, or have greater access to information in some domains, but we take for granted that these differences are not important for the way epistemologists approach epistemological issues, and they are not enough to ground authority in the domain of belief in an interesting and robust sense. Contemporary philosophers rarely mention the assumption of equality directly, but there are many indications of it both in the way we formulate questions, and in the range of solutions that we consider candidates for acceptance.⁵ We treat individual knowers as interchangeable, differing only in situation, and any one can be used to represent the rest.⁶ Egalitarianism is also implied in the adoption of the first-person perspective in discussing epistemological problems. The philosopher or teacher of philosophy addresses everyone, proposing that

5. Richard Foley (2001) is one of the few who mentions intellectual egalitarianism. Foley accepts intellectual egalitarianism, although his defense seems to be mostly a matter of giving its advantages. He argues that (1) it is needed for the freedom to dissent, which combats the laziness that stems from uniformity; (2) nonexperts can form credible opinions over intellectual communities and establishments; (3) it fosters open-mindedness; (4) laypeople can at times be in a better position to evaluate data because they lack biases rooted in the training experts receive in their school of thought or tradition; (5) it allows for metacritiques over those who dominate and influence culture (122–30).

6. Louise Antony calls this "the interchangeability thesis" (1995b: 63). Antony defends the thesis, arguing against a dominant strand of feminist epistemology. She also defends it in Antony (1995a).

they follow her thoughts as if they were conducting the investigation themselves. In doing so, she assumes that the method can be used by anybody.⁷ But that only works if everyone is sufficiently alike in epistemic capacities to make the technique applicable to all of them.

The other value connected with the rejection of epistemic authority is autonomy. Robert Paul Wolff (1998, first edition 1970) argued in a well-known work several decades ago that autonomy conflicts with authority, and that the only defensible political arrangement is anarchism. I have never heard anyone call the systematic rejection of epistemic authority epistemic anarchism, but I suspect that if Wolff's argument were applied to the domain of believing, it would find many adherents. That is because autonomy in the epistemic domain is generally equated with epistemic self-reliance. The intellectually autonomous person has no need for authority. She may take a belief on the word of another person, but she is free not to do so. There is no rational claim on her arising from another person's authority.

The conceptual relations among the ideas of epistemic authority, egalitarianism, and self-reliance are interesting. The rejection of epistemic authority presumably follows from the self-reliant person's reluctance to get beliefs from others. The rejection of authority also follows from egalitarianism since if everyone has the same epistemic powers, then nobody has the superiority needed for epistemic authority. But it is interesting that there is tension between self-reliance and egalitarianism. If the powers of other persons are equal to mine, on what grounds can I be more skeptical of beliefs obtained from them than from myself? We will consider potential grounds for such skepticism later, but notice that if others are equal to me, it appears that I have no reason to mistrust beliefs taken from them due to the inferiority of their powers.

So there are different values connected with the rejection of epistemic authority, and it is not always clear which one drives the position. It is quite a different matter to reject authority because of egalitarianism than because one adopts a view of the self that includes a devaluing of taking beliefs from others. For that reason it is also unclear what the alternative is to adopting beliefs on authority. If one is truly egalitarian, one might reject authority but

7. Consider, for instance, the way we teach Descartes's *Meditations* to introductory students. The common practice of asking students to insert themselves into the *Meditations* as they read it implies that the teacher is egalitarian in the relevant sense.

not place any special value on relying upon the self either. In contrast, one might reject authority because of a mistrust or devaluing of beliefs taken from others, with epistemic self-reliance as the dominant value. This value is not straightforwardly inconsistent with egalitarianism, but it means that even if others are equal to me epistemically, that does not affect the way I ought to form my beliefs.

I think that the value of epistemic self-reliance touches a nerve in the modern Western mind that is even stronger than the egalitarian nerve. In any case, acceptance of the ideal of epistemic self-reliance makes epistemic authority a nonstarter, and I will devote the next section of this chapter to canvassing arguments for it from epistemology. I will look at the idea of autonomy and its relationship to authority in the following section, and the application to epistemic authority and autonomy in the concluding section. I will return to egalitarianism in chapter 3.

2. THE EPISTEMOLOGICAL CASE FOR EPISTEMIC SELF-RELIANCE

2.1. Mistrust of Taking Beliefs from Others

What is epistemic self-reliance and why would anybody value it? To begin answering this question, notice first that there is a trivial sense in which we are all epistemically self-reliant. We must rely upon our epistemic faculties at all times, even in the process of relying upon others. The self has operational primacy, whether or not it has justificatory primacy. In this sense of self-reliance there is no point in defending the value of self-reliance because nobody can do otherwise than to be self-reliant. But if we want to claim that being self-reliant is something it is better to be than not to be, something we can legitimately criticize someone for failing to be, then it has to be something it is possible not to be. "Self-reliance" is not merely a word for something we do in any case.⁸

8. As Gail Fine pointed out to me in conversation, we might think that God cannot fail to be self-reliant, yet it is better for him to be self-reliant than not to be. That seems right. I am assuming that the value of epistemic self-reliance for humans is connected with the admonition to be self-reliant, which does imply that some people fail to be self-reliant in the relevant sense. In general, I think there are difficulties in drawing conclusions about what is good for humans from what is a good feature of God.

As we will see, some philosophers blur the trivial sense of epistemic self-reliance with the sense in which it is espoused as an ideal, but Elizabeth Fricker describes the value in a way that does not make this mistake: "This ideal type relies on no one else for any of her knowledge. Thus she takes no one else's word for anything, but accepts only what she has found out for herself, relying only on her own cognitive faculties and investigative and inferential powers" (2006b: 225). Fricker clearly thinks this value is common, and she says that the ideal of extreme self-reliance was supported by Descartes, Locke, and many others. Later in her paper she argues that it is an ideal that we must give up because of its impracticality, but a superior being could live up to it, and would be superior for being able to do so. Fricker says:

[A] superior being, with all the epistemic powers to find out everything she wanted to know for herself, could live up to this idea of complete epistemic autonomy without thereby circumscribing the extent of her knowledge. Given the risks involved in epistemic dependence on others . . . , this superior being is, I suppose, epistemically better placed than humans are. That is, if she knew at first hand just as much as I myself know in large part through trust in others' testimony, she would be epistemically more secure, hence both practically more independent, and—in some abstract sense—more autonomous than I am. In the same way that I might regret that I cannot fly, or live to be 300 years old, I might regret that I am not such a being. (2006b: 243)

This ideal is interesting in a number of ways. Fricker contrasts epistemic self-reliance (notice she calls it "epistemic autonomy") with epistemic trust in others and says that the latter is due to our inferiority. It is clear from Fricker's characterization of the ideal of self-reliance that it goes far beyond the rejection of epistemic authority, involving the rejection of testimony, and even the rejection of epistemic community. Ideally, we would live in an epistemic state of nature. In fact, to the extent that any kind of community depends upon epistemic trust, epistemic self-reliance as an ideal has the consequence that ideally we would not live in communities of any kind. Of course, Fricker is not proposing that we live in a state of nature because in the world of her imagined ideal, our nature is different. It is one in which an epistemic state of nature does not have any of the consequences that it has in human life as it actually exists. But if Fricker is right that philosophers find epistemic self-reliance ideal, that tells us something about what they

think is worthy of aspiration even if the ideal has to be severely altered in practice. If extreme epistemic self-reliance is an unattainable ideal, it is still an ideal. If so, that suggests that an attainable degree of self-reliance is more valuable than a lesser degree.

In the above passage Fricker alludes to two grounds for valuing self-reliance. One is the epistemic untrustworthiness of others. If we knew just as much by the direct use of our powers as we know now through trust in others, we would be “epistemically more secure.” If we would be more secure by trusting ourselves than by trusting others, that implies that we are more trustworthy than others. The second ground is the value of autonomy in general. Whatever autonomy is, it presumably requires, or at least is aided by, epistemic self-reliance. The rejection of epistemic authority clearly follows.

Let us begin with the first ground for valuing self-reliance mentioned by Fricker. The idea that one should treat other persons as less trustworthy than oneself occurs repeatedly in modern philosophy, although often without argument.⁹ In the practical domain it is reflected in the common aphorism that if you want something done right, do it yourself. The epistemic version of that view is that if you want to make sure your question is answered correctly, find out the answer yourself. This way of thinking has an intuitive attraction, but it is hard to make sense out of it. How is it possible that each person’s own faculties are more trustworthy than the faculties of any other person? Should those other persons rely upon themselves or upon her? There is, of course, a class of beliefs about which each person has first-person privilege: beliefs about one’s own mental states. Presumably, each person is a more trustworthy source of beliefs in that category than any other person, but that is a special case. It is not the one relevant to the claim that getting beliefs on one’s own is more trustworthy than getting beliefs from others.¹⁰ Very few of my beliefs are or reduce to beliefs for which I have a special privilege.

9. An example from Rousseau: “Would we seek the truth . . . in sincerity, we must lay no stress on the place or circumstance of our birth, nor on the authority of fathers and teachers; but appeal to the dictates of reason and conscience concerning every thing that is taught us in our youth.” *The Profession of Faith of the Savoyard Vicar*, par. 133 (2009). The commonality of the value of self-reliance in modern epistemology has been observed by a number of philosophers in addition to Fricker. See, for instance, Code (1991: 111).

10. Even this category of belief is not immune to external critique. See Fricker (1998: 155–206).

One reason we might think that each person should treat others as less trustworthy than herself is that each individual has far greater inductive evidence of the reliability of her own faculties than of the faculties of others since the sheer quantity of evidence about her own faculties vastly exceeds the quantity of her evidence about the faculties of any other person.¹¹ That is true, but even so, it seems to me that the evidence she has does not indicate that her faculties are more trustworthy than the faculties of others. At most, her greater quantity of evidence about herself will give her a greater degree of conviction about her own trustworthiness, but even that is doubtful, given that the greater quantity of evidence in her own case is what one would expect whether or not she was more trustworthy than others.

However, the idea that one should treat others as less trustworthy than oneself might not be due to the idea that they have less trustworthy faculties. Even if we assume that others are equally trustworthy in their beliefs, taking a belief from another person might be untrustworthy because of the difficulty of communicating a belief in a trustworthy manner, a point Fricker observes (2006b: 229–31).¹² As Locke remarked, “Passion, interest, inadvertency, mistake of his meaning, and a thousand odd reasons, or caprichios, men’s minds are acted by (impossible to be discovered), may make one man quote another man’s words or meaning wrong” (*Essay*, IV, xvi, 11). And even if communication is generally trustworthy, it is not *as* trustworthy as getting a belief through the direct use of my own faculties since getting a belief from another person includes extra steps between the grounds for her belief and my own belief. As Locke observed, there are more places at which error can arise.

This view is plausible, but it does not get us very far in the direction of supporting the ideal of self-reliance as described by Fricker. It shows at most that given the option of believing through someone else’s faculties, and believing the same proposition directly *in a way that is just as trustworthy*, the latter is more trustworthy. But that is a long way from giving us the conclusion that getting beliefs from others is less trustworthy in general than

11. I thank Ray Elugardo for raising this objection in conversation.

12. This reason for suspicion of taking beliefs from others could explain why people generally are not suspicious of relying upon instruments such as microscopes and measuring devices, assuming that they rely upon their own powers in interpreting the reading of the instrument.

believing on one's own since we are not often in a situation in which we have such an option. The main reason I depend upon others for many beliefs is that I cannot get the same beliefs on my own, or I can only get them directly in a less trustworthy manner. In such cases, even when I take into account the risk of error in communication, relying upon the faculties of another is more trustworthy than using my faculties directly. Furthermore, even when I get a belief on my own, there are often reasons to rely upon others as checks on my belief, given that I sometimes have a tendency to self-deception or can make an honest error that other persons can correct. The problem of passion, interest, and capricchios mentioned by Locke applies to myself as well as to others. So although this argument may show that believing by the direct use of one's faculties is more trustworthy under certain conditions, we do not get a generalizable defense of epistemic self-reliance from it.

Another way in which getting beliefs from others can be less trustworthy than beliefs I get on my own is that the former may be less sensitive to defeating evidence. This also is mentioned by Fricker (2006b: 235–36). If a belief is outside my area of expertise, typically the identification of defeaters is also. But if I trust the source, presumably I trust her both to form the belief correctly and to be aware of relevant defeaters. It seems to me, then, that mistrust of taking beliefs from others for the reasons I have discussed in this subsection is not due to a plausible argument *for* the untrustworthiness of doing so, but is due to a realization that trust occurs in conditions of vulnerability, which makes it preferable to avoid trust when we can. That is to say, the default position is lack of trust rather than trust. I will return to this point in chapter 3.

2.2. Self-Reliance and the Nature of Knowledge: Plato and Locke

There is another historically important reason for valuing epistemic self-reliance. Some philosophers have defined knowledge as a kind of state that can only be achieved by the direct use of one's faculties. This position can be used as a defense of self-reliance only if knowledge is more valuable than other epistemic states, but most philosophers do make that assumption. The first and most important source of the view that knowledge is superior to true belief or opinion is Plato, and some Plato commentators claim to find

the position that knowledge must be achieved by the direct use of one's own faculties in the Platonic dialogues. In the *Theaetetus* (201B/C) Socrates says that the jury members in a court case do not have knowledge (*epistēmē*) of what the eyewitness knows because the jury believes only by "hearsay" (*ex akoē*). In this passage hearsay could be ruled out as a source of knowledge either because hearsay is untrustworthy, or because knowledge is a state that requires the direct use of one's own powers, and Plato is interpreted differently by different commentators.¹³ I have already said why I think the first interpretation is not very plausible as a general position, and it is even less plausible as an interpretation of Plato when we consider that earlier in the dialogue, Socrates distinguishes teaching from persuasion, and defines teaching as the transmission of knowledge (*epistēmē*) from one person to another. It is unlikely that he means to insist that someone can have knowledge of an event only if he is an eyewitness.

But leaving aside the interpretation of the *Theaetetus*, there is little reason to think that Plato valued epistemic self-reliance as a general policy, given his many admonitions to learn from someone who already has *epistēmē*.¹⁴ The

13. Compare M. F. Burnyeat and J. Barnes: "The jury cannot be expected to attain knowledge, first because what they experience is persuasion rather than teaching, and second because they are not eyewitnesses but dependent on testimony" (Burnyeat and Barnes 1980: 177). "It thus appears that the first key contrast, that between teaching and persuasion, is in conflict with the second, between an eyewitness and the juror who relies on testimony. The contrast between teaching and persuasion is most naturally taken to imply that knowledge can be transmitted from one person to another; indeed, a short while back in the dialogue, teaching was apparently defined as the transmission of knowledge (198B; cp. *Gorg.* 454E-455A with 453D, *Tim.* 51E), which in turn suggests that when a fully adequate explanation has been given to someone, his true belief can be sufficiently well-founded to constitute knowledge. Just this, however, is denied by the contrast between eye witnessing and testimony. Here Socrates asserts that knowledge cannot be transmitted, only belief, implying that in such a case no matter how well-founded a person's true belief may be, if he was not himself an eyewitness to the facts that have been explained to him, it falls short of knowledge. The one contrast invites us to think that knowledge is well-founded true belief, the other that it is not" (179). Other commentators interpret Plato differently. R. S. Bluck thinks the jury example shows that acquaintance is necessary for knowledge, according to Plato (Bluck 1963: 259–63). That interpretation is rejected by Gail Fine (2004: 41–81).

14. In the second chapter of Hugh Benson's manuscript, *Clitophon's Challenge: Platonic Dialectic and the Philosophical Method in Plato's Classical Dialogues*, he discusses the way in which Plato's view of learning from others differs from the traditional paternalistic model of pedagogy. Notice that both ways involve relying upon others. I thank Benson for conversation on the relevant passages in Plato and for help in interpreting them. I also thank Gail Fine and Russell Jones for conversations on these and similar passages.

problem, as Plato's Socrates sees it in his speech in the *Apology* (21b–23b4), is to find a person with knowledge, and Socrates bemoans the fact that there is a paucity of such persons. But that is not a reason to think Plato valued getting beliefs directly rather than indirectly. Consider also the end of the *Laches*. Socrates says:

[I]t would be a terrible thing, Lysimachus, to be unwilling to join in assisting any man to become as good as possible. If in the conversations we have just had I had seemed to be knowing and the other two had not, then it would be right to issue a special invitation to me to perform this task; but as the matter stands, we were all in the same difficulty. Why then should anybody choose one of us in preference to another? What I think is that he ought to choose none of us. . . . What I say we ought to do, my friends—since this is just between ourselves—is to join in searching for the best possible teacher, first for ourselves—we really need one—and then for the young men, sparing neither money nor anything else. What I don't advise is that we remain as we are. And if anyone laughs at us because we think it worthwhile to spend our time in school at our age, then I think we should confront him with the saying of Homer, "Modesty is not a good mate for a needy man." (200e–201b5) (Sprague trans)

It seems to me, then, that the evidence that Plato valued epistemic self-reliance is at best mixed. The problem for Plato's Socrates is not relying upon others, but finding the right others upon whom we ought to rely. What we do learn from the model of Socrates is the value of being a rigorously self-reflective person. Self-reflection leads us to critically evaluate both ourselves and others, but there is no reason to think that it leads to the judgment that one ought to pursue a strategy of relying upon one's own faculties rather than the faculties of others.

When we move to the modern era, the idea that knowledge requires the direct use of our faculties is common. My example is Locke, again because of the subsequent influence of his arguments. Consider what Locke says in a frequently quoted passage:

For, I think, we may as rationally hope to see with other Men's Eyes, as to know by other Men's Understandings. So much as we ourselves consider and comprehend of Truth and Reason, so much we possess of real and true Knowledge. The floating of other Men's Opinions in our brains makes us not a jot more knowing, though they happen to be true. (*Essay*, I, 4, 23)

Locke's point here seems to be that the problem with relying upon others is not that other people are untrustworthy, but that there is something about knowledge that rules out relying upon others, no matter how trustworthy the source.

But how does the argument go? Locke begins with the truism that we cannot know through another person's understanding, the trivial sense of epistemic self-reliance I mentioned earlier. He concludes that there is a difference between a state that counts as knowledge, one that requires the direct use of one's reason, and one that does not, one that involves "the floating of other men's opinions in our brains," and the former is better than the latter. But we do not get *that* difference from the trivial sense of self-reliance, nor is the difference explained by the classical empiricist picture of knowledge as ultimately based on our own sensations. Empiricism can be understood either as a descriptive or as an evaluative doctrine. If it is the former, it is a view about the way human beings form beliefs, some of which count as knowledge. Interpreted this way, empiricism does not distinguish between a person who believes p by the direct use of her faculties and a person who believes p by relying upon someone else, since in both cases the person's belief is formed as the empiricist describes. That is, in both cases, the belief is ultimately based on sense perception. Interpreted descriptively, then, empiricism is not a defense of any value, and so it cannot give us a defense of the value of self-reliance. On the other hand, empiricism can be interpreted evaluatively, as a doctrine about the nature of knowledge, where it is assumed that knowledge is an evaluatively superior epistemic state. Knowledge is defined in a way that distinguishes it from lesser states in part by the claim that in a state of knowledge the knower relies upon her own faculties in coming to form a belief. But that cannot be because the alternative is "the floating of other men's opinion in our brains." However, since Locke elsewhere defines knowledge as the perception of the comparison of the agreement or disagreement of ideas in one's mind (*Essay*, VI, 1, 2), he does seem to think that knowledge is a state in which one exercises one's own powers directly, and he might also have thought of that as a superior epistemic state.

What should we conclude about the value of self-reliance from Locke's definition of knowledge? I think it unlikely that Locke is concerned with defending the ideal described by Fricker. Locke is not a skeptic about testimony. He argues that testimonial belief is justified in the same general way that any kind of broadly inductive belief is justified. (See Locke's *Essay*, IV,

chap. 11.) So even if he thinks that knowledge in his sense is a state superior to belief on testimony, it seems to me that he is not advocating trust in one-self over trust in others. There is no suggestion that we should ideally adopt a policy of forming beliefs by the direct use of our faculties.¹⁵

The argument for valuing epistemic self-reliance I have briefly treated in this subsection seems to me to involve a confusion. There are philosophers who believe (a) knowledge is a better epistemic state than mere true belief, possibly the highest epistemic state to which we can aspire, and (b) the state of knowledge involves the direct apprehension of truth by our own faculties. Belief obtained from others does not qualify as knowledge. But it does not follow from (a) and (b) that we should adopt a policy of forming beliefs by the direct use of our faculties rather than through the faculties of others. It depends upon what the alternatives are. If my options are true belief obtained from others versus no true belief, then clearly I should opt for the former. If my options are true belief obtained from others versus the superior state of knowledge, where the latter involves the direct use of my faculties, then other things equal, I should opt for the latter. But even then, the process of eventually getting to knowledge may involve considerable dependence on others, as Plato argued. Similarly, Locke's view of knowledge involves the comparison of ideas in one's own mind, but often one cannot get the ideas to compare without relying upon other people. As I see it, then, the mistake is in going from "The highest epistemic state has feature F" to "We ought to pursue epistemic states with feature F." If we pursue feature F we may end up with very few of the highest epistemic states.

2.3. Self-Reliance and Cartesian Doubt

When Fricker mentions sources of the ideal of epistemic self-reliance, she refers to another philosopher who is even more important historically than Locke: René Descartes. Fricker does not cite any particular passage, but Descartes is often treated as the champion of epistemic self-reliance because his method of doubt is conducted in solitude, and it involves a cleansing of the mind from any influence from other persons. At the beginning of the *Meditations* and in the *Discourse on Method*, Descartes complains that people

15. Benjamin McMyler argues that what was new in Locke's view of testimony is the way he divorced it from authority. Locke is skeptical about the latter, not the former. See McMyler (2011a: chap. 1).

he had trusted epistemically let him down, and he places his ultimate epistemic trust in something in himself. It is tempting to infer that Descartes is an advocate of self-reliance, but I think it is a mistake to interpret either the Cartesian method of doubt or his foundationalism as a justification of self-reliance. Even if one's epistemic foundation is a power or state in oneself, there is no reason to think that that will lead to greater suspicion of the senses of others than of one's own senses, or greater suspicion of their memory than of one's own memory, and so on for one's other epistemic faculties. What Descartes trusts at the end of the *Meditations* is the use of human powers purified by his method. Descartes maintains that if I have followed his directives, I will reasonably trust my powers much more than I trusted them before I subjected them to the method, and I will trust them more than the powers of others who have not subjected their faculties to the method. But the reason for trust is the trustworthiness of the method. I do not see any evidence that Descartes thought I should trust my own faculties more just because they are mine. On the contrary, his *Discourse on Method* was intended to be the basis for a new science, one in which people must rely upon the method used by others.¹⁶ Furthermore, there is even evidence that Descartes accepted belief on the authority of the Church.¹⁷ This is not a person who advocates self-reliance as a method of getting beliefs.

16. Descartes says near the end of the *Discourse on Method*: "For, desiring to spend my entire life searching for so needed a science, and having found a road that seems to me such that, by following it, one ought infallibly to find that science, were it not for the fact that one is stopped either by the brevity of life or by a lack of experience, I judged there to be no better remedy against these two impediments than thus to convey faithfully to the public what little I had found and to urge good minds to try to advance beyond this, by contributing, each according to his inclination and ability, to the experiments one must have and also by conveying to the public everything they learned, so that later inquirers, beginning where their predecessors had left off, and thus, in joining together the lives and works of many, we might all together advance much further than a single individual could on his own" (VI: 63).

17. Desmond Clarke (2006) describes Descartes's attempts to avoid theological clashes with the Church over the new astronomy, but he argues that "Descartes seems to have accepted the authority of the church to decide what the Bible means, even in passages about natural phenomena. However, he distinguished between the authoritative teaching of the church, as expressed by the Pope or a general council, and the decisions about what books should be published or censored, which was an administrative decision by a small group of cardinals or theologians. That provided an alternative resolution to the one proposed by Galileo, and it gave him enough wiggle room to avoid Rome's decision without directly addressing the fundamental issue about how to interpret Scripture" (113–14). I thank Monte Cook for directing me to this passage and for advice on the interpretation of Descartes.

So far I have not identified an influential epistemological argument that clearly supports greater skepticism about getting beliefs through the epistemic powers of others than through ones own, but that is not to say that there are no arguments. Epistemic self-reliance is a value, and if it is like most values, it is probably adopted because of its perceived connection to other values. In the passage quoted from Fricker above, she gives a second reason for valuing epistemic self-reliance: its connection to the value of autonomy. Fricker identifies epistemic self-reliance with epistemic autonomy and indicates that there is a connection between autonomy in the intellectual and practical domains. Epistemic autonomy, she says, makes one more "practically independent." I think this remark is the key to the reason epistemic self-reliance is so highly prized in modern philosophy and epistemic authority is rejected. Epistemic self-reliance became an ideal because it was thought to be demanded by autonomy. This is particularly interesting because moral and political philosophers typically disagree with Wolff's view that autonomy conflicts with authority, nor are they inclined to say that autonomy requires self-reliance. Why, then, are autonomy and self-reliance equated in the intellectual domain, as we see in the passage from Fricker, and why are intellectual autonomy and epistemic authority thought to be inconsistent?

The fact that most moral philosophers do not find authority and autonomy inconsistent might be a reason to think that intellectual authority and intellectual autonomy are not inconsistent, but one could react in the other direction. Wolff's position that authority is incompatible with autonomy has *prima facie* plausibility, and we might think that moral philosophers resist it only because everyone agrees it is necessary to find a justification for the political state. Nobody is comfortable with anarchy, not even Wolff (1998: 72). But epistemic anarchism is not so frightening. Philosophers arguably have no motive to resist it if that is where the argument leads. But philosophers cannot have it both ways, at least not without argument. If Wolff's argument fails in the political domain, it may very well fail in the epistemic domain. On the other hand, if we take it for granted that it succeeds in the epistemic domain, we need special argument to explain why it does not succeed in the political domain.

My last historical exhibit will be Kant, and I will look next at whether the Kantian idea of the authority of the self leads to the rejection of epistemic authority or to a defense of the value of epistemic self-reliance.

3. THE CASE FROM ETHICS: SELF-RELIANCE AND AUTONOMY

According to the standard interpretation of philosophical history, Immanuel Kant revolutionized ethics by making the ultimate moral authority one's own rational will. I take this to be the heart of the idea of autonomy. It is undoubtedly one of the most influential ideas in modern moral and political philosophy, and it has been adopted as a personal ideal by many philosophers who in other ways lack sympathy with Kant's ethics.¹⁸ For some philosophers, autonomy is a capacity of a person that gives her a right that can be either respected or violated. Alternatively, autonomy can be identified with the successful exercise of that capacity, an ideal state to which we aspire.¹⁹ The dominant idea is that of rational self-governance, although autonomy has sometimes been identified with a variety of other values, including the Stoic idea of self-sufficiency, the existentialist idea of authenticity, the idea of integrity, and especially the idea of independence.²⁰ But I think that the core idea of autonomy is this: It is the right or ideal of submitting to nothing but one's own rational will. As Joel Feinberg expressed it, "I am autonomous if I rule me, and no one else rules I" (1980: 23). Autonomy as I interpret it, then, is fundamentally about authority. I have the authority to rule myself. It is also a value; it is better if I exercise that authority than if I do not.

Autonomy is most interesting if it is a right or ideal that is pre-moral. What I mean by that is that the idea of autonomy does not rest upon any moral intuition, and it is strong enough to be a constraint on our understanding of what morality is. If I am right about that, that would explain why the idea of autonomy can support the modern shift from thinking of morality as obedience to law to understanding morality as self-governance, masterfully described by J. B. Schneewind in *The Invention of Autonomy*.²¹ What was new in

18. J. S. Mill endorsed the personal ideal in this way: "If a person possesses any tolerable amount of common sense and experience, his own mode of laying out his existence is the best, not because it is the best in itself, but because it is his own mode" (1947: 67).

19. Feinberg (1989: 27–53). A similar distinction between autonomy as a capacity and the successful exercise of that capacity is made by Darwall (2006: 263–84).

20. For a survey of views of autonomy, see Christman (2002) and Buss (2002).

21. If the idea of autonomy is pre-moral or extramoral, that would also explain why some philosophers think that a person can act autonomously when she is evil. See Feinberg (1989: 37–38) and Coeckelbergh (2004: 17).

the early modern era was not the idea of self-governance, which arguably was an important value since the ancient Greeks.²² What Kant gave us was the idea that morality is self-governance, the rules by which a rational being governs itself. The additional premise is that morality comes in the form of a command; it is the product of a will. So if morality comes in the form of a command, and a person should not submit to any will but her own, it follows that morality is a command I give myself. It gets its authority from my self. The Kantian position on the nature of morality, then, follows from the nature of autonomy as self-governance combined with the view that morality comes in the form of a command.

Why would anybody think, either on moral or premoral grounds, that I should not submit to anything but my own rational will? To ask the question another way, why did philosophers decide at some time in the early modern era that the ultimate authority over myself is my self or some part of my self, and how did that differ from what they thought earlier?

In the ancient and medieval periods there were two different but not necessarily competing grounds for authority—one Greek, one Judeo-Christian, both of which were either rejected or severely modified in the modern West. The first way to ground authority was in God, the creator and governor of the universe (e.g., “There is no power but from God” [Romans 13:1]). If we belong to God as his creatures, God is the ultimate authority, and any other authority derives from the divine plan for governance of the world. The second ancient ground of authority was reason. For the Greeks, reason is naturally and intrinsically authoritative; the authority of reason is self-evident. A person is self-governing only insofar as she has a share in the force of reason that governs the universe. In *Republic* VIII and IX Plato argues that it is the rational part of the person that governs, and that part governs the parts of the soul other than reason. The authority of the person to govern herself therefore is not primary; it is derivative from the authority of reason. Since it is reason in the person that is authoritative, and since some individuals have more reason than others, that also explains why some persons have authority to govern others. So if reason is what is ultimately authoritative,

22. Anthony Flood argues in his dissertation, “Self-Governance in Aquinas and Pre-Modern Moral Philosophy” (2003), that there are numerous ancient sources of the idea of self-governance stretching back to Socrates, and that Aquinas had a robust notion of self-governance in his moral philosophy.

the extent to which reason entails self-governance does not rule out external authority.

This Platonic theme can be found in most of the major Western philosophers until the modern period. The stress is on the authority of reason, with self-governance derivative from that authority.²³ In the later Middle Ages there was a split between those philosophers who continued to claim that the source of authority is in the divine reason, and those who claimed that the source of authority is in the divine will rather than the divine reason. Duns Scotus was one of the earliest proponents of the latter view.²⁴ This shift from reason to will as the ground of authority had significant consequences since the relation between the divine and human wills is quite different from the relation between the divine and human reason. Our reason is arguably a share in the divine reason, which is why a measure of self-governance for the ancients and medievals is compatible with being governed by God. In contrast, our wills obviously do not share in the divine will. A will by its nature is individual.

In the century leading up to Kant, attitudes towards both of the premodern grounds of authority changed. Reference to God as the source of authority was given up, and simultaneously, there was a shift from the idea of the natural authority of reason to the idea of the natural authority of the self over the self. Kant brilliantly combined the modern idea of the authority of the self with the ancient idea of the authority of reason in his view that to be

23. For instance, Aquinas argues that the virtue of prudence (*prudentia*) is necessary to live a good life (*ST I-II*, q. 57, a. 5, obj. and reply 2), and he argues that self-rule is an intrinsic component of prudence. Aquinas says:

Prudence is in the reason. Now ruling and governing belong properly to the reason; and therefore it is proper to a man to reason and be prudent in so far as he has a share in ruling and governing. . . . Since however, every man, for as much as he is rational, has a share in ruling according to the judgment of reason, he is proportionally competent to have prudence. (*ST II-II*, q. 47, a. 12, corpus)

Aquinas's view of self-governance as deriving from reason is compatible with his position that authority comes from God because God promulgates his laws through reason (*ST I-II*, q. 94, a. 2). This position had Stoic roots, an expression of the ancient Greek view that reason is the divine in the human. See Cooper (2003) for an argument that the Stoic Dio Chrysostom claimed that the laws of nature are the expression of the will of Zeus, a position that Cooper interprets as similar to Kant.

24. See Schneewind (1998: chap. 2, sec. 3) for a historical overview of voluntarism, or the view that moral authority is grounded in a will. For a history of Divine Command theory, see Idziak (1980).

governed by oneself and to be governed by reason are the same thing because the true self is one's rational will. But the important question is which is more basic. Is the point of the Kantian view of autonomy that I should not submit to anything but my *rational* will, or is it that I should not submit to anything but *my* rational will? If it is the former, reason is still the primary authority, and there is no explanation for why it should be my will that governs me rather than any other rational will. If instead, autonomy means I should not submit to anything but my own rational will, my rationality is not sufficient to explain why other wills do not have authority equal to or greater than mine. Many modern philosophers will say that I do not need a justification for the authority of my own will, but since philosophers before the modern period did not see it that way, we need a defense for the shift from the idea that the authority that needs no justification is reason to the idea that the authority that needs no justification is the authority of the self over itself.

Christine Korsgaard offers such a defense. She argues that Kant's answer to our question is that the self's authority over itself does not derive from the authority of a rational will; rather, reason is authoritative because it is the rules that the self must set to govern itself (2009: xi). The self just is a being with an executive function. It must take control of itself because of the operation of self-consciousness. The rules of reason are the rules of a self-conscious being. This idea constitutes a monumental change in the idea of authority. Korsgaard interprets the authority of reason as derivative from the authority of the self-conscious self, the reverse of the traditional view.

As I interpret the historical development of the idea of authority, then, there was a progression from the ancient idea that authority over me resides in *reason*, to the idea that authority resides in the *rational will*, to the idea that authority resides in *my* rational will. Notice that this progression involves two shifts: from reason to the will, and from reason to the self. The first shift occurred first in strong Divine Command theory, in the idea that the divine will is the source of authority, not just the form through which the authority of the divine reason operates. Kant put the two shifts together in the idea that *my* rational will is the source of moral authority. Furthermore, in making the authority of the self primary and the authority of reason derivative, there was room for a loosening of the connection between the authority of the self and the authority of reason, culminating in the Nietzschean view that the self is the authority and my own will is the only thing to which I should

submit, whether or not it is rational.²⁵ Of course, this progression is not inevitable and there are philosophers who will get off the train somewhere along the line. But it is natural to ask, "How did Western philosophers get from the idea that the fundamental bearer of authority is reason, a faculty in human beings that is the closest we get to the divine, to the idea that each person's will is the fundamental bearer of authority over herself, whether or not it is rational?" My suggestion is that the progression in views on the source of authority (and the corresponding rise of the idea of autonomy) followed a path roughly like this:

reason → the rational will → my rational will → my will (rational or otherwise)

The first two were rooted outside the individual person, generally in the divine reason or will, with human authority based on the human being's submission to or imitation of the divine ground of authority. The third constituted a radical shift, although Kant did not give up the idea that authority is grounded in universal reason. What was radical was the idea that universal reason is attached to my own will. It was that feature that permitted later degeneration into the view that my will, unconstrained by anything, including reason, is the only authority over me.

4. AUTHORITY AND AUTONOMY IN THE INTELLECTUAL DOMAIN

Let us now look at what all this has to do with intellectual authority. Even if the ultimate bearer of practical authority is one's will, nobody but Hobbes would say that the ultimate bearer of intellectual authority is a will.²⁶ Nevertheless,

25. For an interesting discussion of the culmination of the idea of authority in the individual will, see Taylor (1976: 288–94).

26. In *Behemoth*, Dialogue 1, Hobbes argues that the Sovereign is the intellectual authority in the state, holding authority over the Church in its teachings. He also argues that the universities need to be disciplined so that they teach what the sovereign wants since the universities are the core of rebellion, as happened in the English Civil War. This part of *Behemoth* includes a long diatribe against the Catholic Church and the Presbyterian Church for their "pretensions" to authority. In contrast to Hobbes, Locke denies that authority over religious belief is possible, as I noted earlier. He says: "[T]he care of souls

there is a very close connection between intellectual autonomy and autonomy of the will, and a corresponding connection between intellectual authority and practical authority.

Notice first that autonomy of the will presupposes autonomy of the intellect. It is unlikely that we can autonomously make a choice unless the beliefs upon which the choice is based are autonomous.²⁷ This point does not depend upon a particular view of autonomy, and it is enough to indicate that at least some of our beliefs ought to be formed autonomously, assuming that it is good that acts are autonomous.

Notice also that the ways in which a will can be heteronomous according to Kant have a parallel in the formation of beliefs. A will is heteronomous in one way when it is controlled by a will outside of it. Similarly, an intellect can be too greatly influenced or even controlled by someone else's will. It is plausible to say such an intellect is heteronomous. Intellectual coercion is generally thought to be impossible, as I have noted, but commercial and political advertising are common ways of pressuring people to form particular beliefs even though the beliefs are not literally coerced. It is understandable that people sometimes complain that autonomy is violated, at least to some degree, by the use of such methods of influencing belief. An amusing and more extreme view on coercion over belief is Robert Nozick's claim that *rational* persuasion is coercive, and philosophers are guilty of coercing people's minds. Nozick says:

The terminology of philosophical art is coercive: arguments are *powerful* and best when they are *knockdown*, arguments *force* you to a conclusion, if

cannot belong to the civil magistrate, because his power consists only in outward force: but true and saving religion consists in the inward persuasion of the mind, without which nothing can be acceptable to God. And such is the nature of the understanding, that it cannot be compelled to the belief of any thing by outward force. Confiscation of estate, imprisonment, torments, nothing of that nature can have any such efficacy as to make men change the inward judgment that they have framed of things" (Locke 2009: par. 16). Rousseau (1968) seems to have an intermediate position. He agrees with Hobbes that the sovereign should fix the articles of faith, but he says it cannot compel a citizen to believe them, although it can banish whoever does not believe them as an antisocial being (*Social Contract*, Bk. IV, chap. 8). I thank Zev Trachtenberg for conversation on these passages.

27. I will not raise the issue of the metaphysical freedom of the will since I assume that metaphysical freedom is neither necessary nor sufficient for an autonomous will. Both compatibilists and incompatibilists about free will and determinism maintain that the kind of control a person must have over her own will in order to be autonomous is possible. What is relevant in autonomy of will is not the metaphysical conditions needed for choice, but what it contrasts with: a heteronomous will.

you believe the premisses you *have to* or *must* believe the conclusion, some arguments do not carry much *punch*, and so forth. A philosophical argument is an attempt to get someone to believe something, whether he wants to or not.... Why are philosophers intent on forcing others to believe things? Is that a nice way to behave toward someone? (1981: 4–5)

We may not all agree on whether anyone is entitled to influence the beliefs of others by rational argument or any other method, but fortunately, we do not have to agree on that issue because the first kind of intellectual heteronomy is relevant either way. Surely another person can unduly influence my beliefs, and that can happen even when the other person is not claiming authority, is not commanding or attempting to coerce belief, and may even be unaware of my existence. Of course, we will want to know what undue influence amounts to, but it is plausible that it exists and that it is problematic for roughly the same reason Kant gives for thinking that the first kind of heteronomy of the will is problematic. We are not fully rational if either our wills or our intellects are pushed around, whatever “pushed around” amounts to.

A will is heteronomous in the second way, according to Kant, when it is determined by forces within the self other than reason—by inclination or fancy. An intellect also can be determined by inclination or fancy, something other than reason, and there is a *prima facie* case for calling such an intellect heteronomous.²⁸ It seems to me that the value of intellectual autonomy that contrasts with a heteronomous intellect in this sense is not very controversial. Nobody denies that reason is good for the intellect, and that forming beliefs by inclination or fancy is a bad idea. I think that we can expect, then, that if autonomy is valuable, intellectual autonomy is also.

We are now ready to look at whether the value of intellectual autonomy gets us to the value of epistemic self-reliance. Does Kantian autonomy justify relying upon one’s own epistemic powers without relying upon others? The answer is no. More accurately, if it does, it cannot be by way of an argument that Kant himself endorsed. Kant argued that the value of noninterference in the freedom of thought and expression of others depends upon the *rejection* of epistemic self-reliance, the position he calls logical egoism. Towards the beginning of his *Anthropology*, Kant argues that the judgment of

28. I am leaving open the possibility that emotion can be a determinant of belief that does not undermine intellectual autonomy. But the distinction Kant draws remains relevant. Some forms of emotion undermine the autonomy of the intellect even if others do not.

others is a test of the veracity of our own judgment. This applies even to judgments in mathematics and some categories of sense perception:

The logical egoist considers it unnecessary to test his judgment by the reason of others, as if he had no need of a touchstone (*criterium veritatis externum*). However, it is so certain that we cannot dispense with this means of ensuring ourselves of the truth of our judgment, that it is perhaps the most important reason why learned people insist so emphatically on the freedom of writing. If this freedom is denied, we are deprived of an effective means of testing the correctness of our judgment, and we expose ourselves to error. It should not even be said that a mathematician is privileged to make judgments on his own authority, for if the perceived and verified agreement between the judgments of one geometer and those of all others who devote themselves with talent and industry to the same subject did not prevail, then even mathematics would not be free from having somewhere fallen into error. There are also certain cases where we do not trust solely in the judgment of our own senses, where we find it necessary to inquire of other people if they seem to have had the same *impression* [italics in source] as ours, for example, whether the ringing of a bell was real or only in our ears. And although, when philosophizing, we are not even permitted to appeal to the judgment of others for establishing our own..., every writer who finds that no one agrees with his clearly expressed and important views is suspected by the public of being in error. (2006: 11)

Kant's rejection of epistemic self-reliance is very clear. At a minimum, it should lead us to suspect that a line of argument connecting Kantian autonomy and epistemic self-reliance is based on a misreading of Kant. Of course, the defender of epistemic self-reliance on grounds that it is required by autonomy need not be Kantian. But it is a mistake to think that we can identify epistemic self-reliance with intellectual autonomy without argument. I have agreed that practical autonomy probably requires intellectual autonomy, but that does not mean that practical autonomy requires epistemic self-reliance.

5. THE VALUE OF REFLECTIVE SELF-CONSCIOUSNESS

In this chapter I have given a preliminary search for the roots of the rejection of epistemic authority. The answer I proposed is that there are two deep values, either of which can be used to reject epistemic authority, but which

are at odds with each other: egalitarianism and an interpretation of autonomy that identifies it with self-reliance. Egalitarianism will become important in later chapters. This chapter focused on self-reliance, and I argued that a defense of epistemic self-reliance from Plato, Descartes, Locke, or Kant is at best problematic, and probably based on some confusions. Whether self-reliance is interesting and important as an ideal depends upon both the way we describe self-reliance and the function of ideals. Fricker knows that the ideal she describes is impossible for humans to attain. I said earlier that I think her implied assumption is that if something is an unattainable ideal, it is nonetheless an ideal, and an attainable degree of what is ideal is a goal worth striving for. If so, we should try to be as self-reliant as we can be, consistent with our other goals, such as getting as much knowledge as we can in the short span of our lives. It might turn out that epistemic self-reliance is in fact something we find good in the sense of the desirable or, what is more likely, in the sense of the admirable, and the fact that it conflicts with some other values need not be a reason to declare that it is not a value. There might also be senses of epistemic self-reliance I have not considered and that have been defended by philosophers I have not mentioned. There are reasons to think that if autonomy is valuable, intellectual autonomy is also. But the connection between autonomy and self-reliance is puzzling, and part of what I hope to do in this project is to disentangle them.

There might be a much simpler defense of epistemic self-reliance than anything I have considered in this chapter. Epistemic self-reliance is a component of self-reliance, and self-reliance might just be a basic value, a value we detect through intuition, or a value that is grounded in natural human desires. We do not like to be dominated. We hate submitting to the will of others. Many of us would rather dictate our own lives and do it poorly than to submit to somebody else, even if the somebody else is wiser than we are. In any case, most somebody else is not wiser than we are and may have motives that are deeply suspect.

So we might value self-reliance because we spurn the idea of domination by another person. More importantly, the positive attraction of the popular image of the self-reliant person (almost always a man) can be a potent force in shaping our idea of the kind of person we want to be. Think of Ralph Waldo Emerson's classic essay "On Self-Reliance," which gets its plausibility not from arguments, which are virtually nonexistent, but from its passionate appeal to the beauty and nobility of a certain image. Many people resonate

with Emerson's remarks that one has to "take with shame" the opinion of another, or that "imitation is suicide," and that "Nothing is sacred but the integrity of your own mind" (1903: 30). Trust in one's admiration for the kind of person Emerson so vividly describes should be taken seriously. One of the themes that I will develop in this book is the idea that ultimately we have to decide what we trust, and some of what we trust is more basic than the beliefs we would use in a justification for any view on authority. Trust in oneself is reasonable. Whether it is reasonable to trust oneself more than others remains to be seen, but I have no objections to taking important positions based on what we trust upon reflection, including the kind of persons we admire.

In the rest of this book I will navigate the issues of epistemic authority, autonomy, and self-reliance by starting from the self as a self-reflective being. I want my method to be acceptable to any contemporary philosopher, regardless of her view on Enlightenment philosophy, and regardless of her view on authority and autonomy. I will attempt to follow the Socratic dictum to be rigorously self-reflective, but I will do it in the modern manner, by unraveling the implications of reflective self-consciousness, the method Korsgaard attributes to Kant. I hope to show by the last chapter that a reasonable view of intellectual autonomy falls out of this investigation.

Epistemic Self-Trust

1. THE NATURAL AUTHORITY OF THE SELF

In this chapter I begin an extended argument for epistemic authority by examining the implications of reflective self-consciousness. I will assume that a self is conscious of itself; it is the inner world of a person. It is probably possible to be conscious without being self-conscious, so self-consciousness is a distinctive form of consciousness. It includes awareness of the distinction between the subject and object of consciousness, and in self-consciousness the subject is able to direct consciousness to itself.

To be conscious of oneself includes consciousness of a variety of mental states that traditionally have been divided into rather vague and overlapping categories, including beliefs, desires, emotions, sensations, attitudes, judgments, and decisions, as well as imaginary versions of each. So we can imagine believing, desiring, having a certain emotion or sensation, making a certain decision, and so on, and each of these states is something of which we are aware as self-conscious beings. At a very early age we become aware of a difference between ourselves and a world outside our minds towards which various of our mental states are directed, and we use this difference to distinguish some of our mental states from their imaginary versions. One need not presume that there is such a world to notice that we think that there is, and that the way we divide ourselves and the world is partly constitutive of what we take ourselves to be.

Beliefs, emotions, desires, and decisions can conflict with one another. We experience conflict between our mental states as dissonance. It seems to me that the experience of dissonance is basic. It cannot be explained or analyzed in terms of some other experience. I do not mean that conflict is

defined by the experience of dissonance since there can be conflict that is unconscious, but we detect conflict through the experience of dissonance. Many times when there is dissonance, the self automatically adjusts by giving up one of the states that conflict. This often happens when there is conflict between a belief and a perception. I believe that I turned off the watering system, but then hear the sprinklers go on. I give up my belief without any thought to the conflict. The conflict is short-lived and psychic harmony is restored effortlessly and without conscious attention.

I think that the awareness of dissonance resolved without effort gives us our initial model of what rationality is. I say that because I think that rationality is a property we have when we do what we do naturally, only we do a better job of it. To be rational is to do a better job of what we do in any case—what our faculties do naturally. Granted, “rationality” is often used to apply to the cognitive faculties in particular, but there is an extended sense of rationality that applies to other faculties. I am speaking now of rationality in the extended sense.

Sometimes there is conflict that is not automatically resolved. If we are aware of dissonance, we typically feel a need to resolve it by giving up one of the components of the self that creates the dissonance, and we may not want to do so. That might be a reason not to become aware of the conflict, and it is one of the sources of self-deception. I think that what Sartre called “bad faith” is a conflict within the self that is suppressed at the conscious level. It is conflict that a self attempts to resolve by suppressing the experience of dissonance. But the conflict does not go away by suppressing the feeling that normally accompanies the awareness of it.

It is possible to be aware of conflicting beliefs without experiencing dissonance. Well-known paradoxes such as the Lottery Paradox and the Paradox of the Preface are probably like that. People typically feel no dissonance even after they see the conflict.¹ These paradoxes are not handled the way other belief conflicts are handled; they are treated as intellectual puzzles. I conjecture that that is due to the fact that they do not produce dissonance. We can judge that two beliefs conflict without experiencing dissonance even though we get the idea of conflict from the experience of dissonance. Paradoxes are cases in which a set of beliefs share features with cases that do generate dissonance. If I am right

1. I thank Jonathan Rutledge for pointing this out to me.

that we do not experience dissonance in the case of the paradoxes, that just makes them more interesting to us.

Some forms of dissonance do not need to be resolved; we can get along well enough with the dissonance. This often happens with conflicting desires, or with a desire that conflicts with a belief. I believe that I will go on a trip tomorrow, but I do not want to go. I am aware of the dissonance between the belief and the desire, but I do not feel an urgency to give up either the belief or the desire the way I do when I am aware of conflicting beliefs. Nonetheless, it is better if dissonance is resolved. Conflict between desires or between a belief and a desire can also be resolved unconsciously, and the experience of the resulting harmony gives us a model of a kind of rationality that is desirable for the same reason we desire harmony in our beliefs: We naturally desire and attempt to achieve a harmonious self.

When we are aware of a conflict within the self, we might find it hard to give up either side to the conflict, but may make the judgment that a certain one *ought* to be given up. This situation often occurs when a decision conflicts with a set of beliefs. We may judge that we ought to change the decision, but find it difficult. Perhaps we are able to do so after a struggle; we call that continence. Perhaps we are not able to do so and we call that incontinence or *akrasia*. *Akrasia* is often called moral weakness, but the weakness need not have anything to do with morality. Indeed, one can be akratic when what one judges one ought to do is opposed to morality. Donald Davidson (1970) made this point in a famous paper many years ago. I think the point can be generalized further. A form of *akrasia* occurs any time there is one component of the self that conflicts with another, and we judge that a certain one should be given up, and we do not do so. The conflict need not be between moral judgment and desire. It can be between two beliefs or between two desires or between a desire and a decision, or between an emotion and a belief, and so on for many other states of the self.² A mild form of *akrasia* exists as long as the conflict exists and we are aware of the dissonance but do not resolve it. A stronger form of *akrasia* occurs when we resolve the conflict in favor of the wrong side to the conflict—what we judge we ought not to do.

2. I briefly discussed *akrasia* in belief in Zagzebski (1996: 154–55). See also Hookway (2001).

I have proposed that we begin reflection with a model of what we ought to do to resolve conflict, which is, very roughly, what we would do if we were doing it automatically and without effort. What we do reflectively builds upon a base in what we do prereflectively. Of course, part of the point of reflection is to change some of what we do prereflectively, but reflection operates on processes that already exist in our prereflective state. We judge ourselves reflectively with the prereflective experience of successful resolutions of conflict.

In cases in which we do not do what we judge we ought to do to resolve dissonance within the self, we are not rational because it is not rational to fail to do what we judge we ought to do. Failing to follow our own judgment of what we ought to do to resolve dissonance within the self creates more dissonance in the self. But there is a way in which our judgment that we ought to change when we do not do so brings harmony to the self at the reflective level. There is harmony between the reflective judgment and what would restore harmony in the self. This gives the reflecting part of the self a kind of harmony even when we do not follow the judgment.³

My position, then, is that there is a connection between rationality and our reflective judgment of what produces harmony in the self.⁴ I have added the idea that there is a connection between rationality and what people do automatically. I think we would have a lot of trouble distinguishing rational from irrational behavior were it not for the experience of making an automatic adjustment when there is dissonance in the states of the self. Of course, I am not suggesting that rationality should be defined by such behavior. But what we do automatically gives us our initial standard of rationality, a standard for what it is to make the adjustment in the self correctly. The criterion works only if there is a close connection between the way the self naturally

3. I thank Wayne Riggs for making this observation.

4. We have no guarantee that our judgment of what will restore harmony in the soul is correct, and ideal rationality may require a correct judgment. If so, and if I make an incorrect judgment about what will resolve dissonance between A and B, I will not be ideally rational no matter which way I go. Suppose that harmony is better restored by resolving a dissonance between A and B in favor of A, but I incorrectly judge that it ought to be resolved in favor of B. It is not rational in such a case to give up A and resolve the dissonance in favor of B since that does not correctly resolve the dissonance between A and B. But it cannot be rational to give up B and resolve the dissonance in favor of A since that creates dissonance with my judgment that I should opt for B. Ideal rationality is very demanding and involves a lot of luck. I will mention this feature of rationality occasionally in later chapters.

operates and what the self ought to do. This means that there is a connection between the natural and the normative, in particular, a connection between the self as it naturally operates and the way it should operate.

When parts of the self adjust automatically, no executive is needed. The self exercises its executive function when we have to make up our mind. Choice in action involves an executive function, but other changes in the self do also. Sometimes resolution of dissonance within the self requires the exercise of the executive function of the self. It does so when the resolution of dissonance does not occur automatically. The executive self can also be called an agent. The self is an agent in its role of taking charge of itself, correcting itself, thereby becoming a more harmonious self, and hence, in some deeper way, more of a self. A self-conscious being has an executive function in virtue of being a self. This is the sense in which the self has natural authority over itself.⁵

I said that a change in the self is a response to disharmony and the result is a more harmonious set of states. But the story is more complicated than that because there is always more than one way to resolve dissonance, and some ways are better than others. Some ways result in dissonance later. If I give up the belief that someone is irritated with me and replace it with the belief that he is anxious, I might end up with more dissonance in the future. That may happen if he really *is* irritated with me and it was my perception of anxiety that was misleading. I suspect that a resolution of dissonance in favor of a false belief or an inappropriate desire or a wrong decision tends to produce more dissonance somewhere else in our vast mental network at some future time, but it is not necessary to agree with me on this Platonic point to follow my line of inquiry.

2. THE NATURAL DESIRE FOR TRUTH AND THE PREREFLECTIVE SELF

I want now to focus on a desire that is part of every prereflective self: the natural desire for truth. To say there is a natural desire for truth is out of fashion in at least two ways. For one thing, contemporary philosophers generally do not

5. Does God have an executive self? It would seem that God has no dissonance to be resolved, and hence no need to have an executive self at all, yet God is an agent. I thank Rodrigo Borges for pointing out this problem in understanding divine selfhood.

speak of natural desires, although it is part of a long philosophical tradition to do so, and reference to natural desires is common in many fields.⁶ It is particularly passé to claim that there is a natural desire for truth. I think there are probably many natural desires, but the desire for truth enjoys primacy in the map of our desires, and I think that everyone has it.⁷ It is possible that the Pyrrhonian skeptics succeeded in giving it up, and I will return to that, but even the Pyrrhonians recognized that skepticism is not the natural way to be since they thought they had to go through therapy to get there.⁸ So I assume that with possibly a few exceptions, everyone desires truth, and it is natural for everyone, even if there are a few who do not desire it.

Second, many philosophers are suspicious of the word “truth.” Deflationists about truth say that if I want to know whether the proposition expressed by the sentence “St. Augustine was bishop of Hippo” is true, that is just to say that I want to know whether St. Augustine was bishop of Hippo.⁹ The word “true” adds nothing to what can be said without it, and any proposition can be expressed equally well without the word “true.” I accept this position in one way but not another. I agree that we can say that there is a natural desire for truth without using the word “true.” We could say that the natural desire for truth is the natural desire to have our questions answered. We want to know who or what something is, or whether something is the case, or how something works. When I say there is a natural desire for truth, I mean at least this much: It is natural to ask questions and to want answers to them. But we do not want just any answers that make us feel that the desire is satisfied. We want the *right* answers. To say that we want the right

6. There are many historical examples of the idea that there is a natural desire for truth. Aristotle begins his *Metaphysics* with the famous sentence, “All men by nature desire to know.” Thomas Reid said, “The love of truth is natural to man, and strong in every well-disposed mind” (*Essays*, chap. 8, 1993: 529). Here is a contemporary example from the physicist Euan Squires: “I will close by giving what seems to me to be the pragmatic case for realism, i.e. for the belief in an external reality. It is the belief that a reality exists that provides the motivation for seeking it. Built into our nature is our desire to know the truth; it is, along with beauty and love, one of the things for which we crave. Though we will surely never find it, in any absolute sense, we will go on seeking it. Certainly we have far better chances of finding it if we search for it, than if we accept that it doesn’t exist and stop looking” (1994: 127).

7. I defend the centrality of caring about truth in Zagzebski (2005).

8. See Sextus Empiricus’s *Outlines*, Bk. I, chaps. xiii–xvii.

9. See, for example, Ramsey (1927: 153–70); Ayer (1935: 28–32); Quine (1986); Field (1994: 249–84); Horwich (1994).

answers to our questions is another way of saying we want the truth. I assume that “truth” and “right answer” amount to the same thing.¹⁰ Moreover, I think that the desire for truth goes beyond the desire to have our questions correctly answered. Sometimes we do not even know enough to know what question to ask, as Socrates taught us. The desire to be shown the question to ask as well as its correct answer is part of the desire for truth.

Suspicion of truth is rampant in many fields of the humanities for reasons different from those that motivate deflationist views on the semantical function of the words “true” and “truth.” Generally, Anglo-American philosophers do not take this suspicion seriously,¹¹ but I think that those who have these worries can accept my assumption about the desire for truth. I am leaving it open that conflict within the self revealed by critical self-reflection can lead a person to modify the areas in which truth is possible, perhaps due to unresolved disagreement or mistrust of the conventional sources of truth in some domains. At the extreme, one can decide upon reflection that there is no truth to be desired. As I interpret suspicion about truth, then, it arises from the desire for truth that, upon reflection, seems to go awry. But it takes reflection to see that it goes awry. Nobody would pay attention to skepticism about truth if there were not a natural desire for it.¹² The postmodern suspicion of truth arises from the same process that generates any form of skepticism. Indeed, it arises from the same process that generates conscious antiskepticism. It arises from reflection upon the natural desire for truth.

In addition to the natural desire for truth, I assume that there is a natural belief that the natural desire for truth is satisfiable. I believe I can get the answers to my questions, and I depend upon those answers in the way I live my life, so I think that the desire for truth can be and often is satisfied. Again, skepticism about the possibility of getting truth in some domain might result from reflection, but I begin with the belief that truth is attainable. I do not

10. I agree with Dennis Whitcomb (2010) that the desire for truth is like hunger. It is a desire that can be satisfied correctly or incorrectly. We want nourishment, not perceived nourishment. Similarly, we want the truth, not the perceived truth.

11. A notable exception is Bernard Williams (2002).

12. I am assuming that the natural desire for truth is not a desire for truth as the pragmatist understands it, although I am not offering an argument against pragmatism. See Williams (2002: 59–60 and chap. 6) for arguments that the person using a pragmatic notion of truth needs the nonpragmatic notion of truth.

believe that it is attainable only after reflection. As soon as I think about whether my questions can be answered, I realize that many of them were answered before I asked myself whether my questions can be answered. That is, I asked questions and got answers, and I believed the answers. And I use many of those beliefs now as I undertake reflection upon my beliefs. I also notice that many of my beliefs come unbidden or change without reflection. That is the way beliefs are. Very few of them are the result of the exercise of my reflective capacities. So if I think I can get truth, I must think I can sometimes get it without the aid of reflection, and probably do so often. But if so, I must have a basic trust in the connection between my faculties and the truth. Furthermore, since I relied upon my faculties to answer my questions before I started reflecting, trusting my faculties was already part of the pre-reflective self.

So far I have argued that there are a number of components of the pre-reflective self related to truth: (1) a natural desire for truth, (2) a natural belief that the natural desire for truth is satisfiable, and (3) a natural trust in the suitability of my faculties for getting the truth. I want to comment on the third component, the component of trust. I think of trust as a three-place relation.¹³ One trusts something for some purpose or in some respect. I trust my car to get me to campus. I trust my friend not to betray a confidence. I do not trust a politician caught in an unlawful act to accurately answer questions about the incident, and I do not trust the Oklahoma weather to be what I expect it to be on any given day. Trust in persons is more complicated than trust in nonpersons because the former includes trust in their goodwill towards me, whereas the latter obviously does not.¹⁴ When the trust in question is epistemic trust in myself, I trust my epistemic faculties to get me to the truth. Trusting myself in this sense is like trusting my car. It does not involve relying upon my goodwill towards myself.¹⁵

13. This view of trust as a three-place relation appears in Baier (1995a) and in Jones (1996), among others.

14. See Baier (1995a: 98). Joseph Godfrey distinguishes reliance trust from trust in persons, calling the latter "I-Thou" trust (Godfrey 2012: chap. 2).

15. I am not denying that there may be cases in which self-trust does require relying upon my goodwill towards myself. For instance, forgiving myself may involve trusting myself not to punish myself. Rodrigo Borges has pointed out to me that trusting my goodwill towards myself might apply in the epistemic domain as well since I might need to trust myself not to engage in self-deception in the way I get beliefs.

I think that the state of trust is a hybrid of epistemic, affective, and behavioral components. As a first approximation, I propose that when I trust x for purpose y , (1) I *believe* x will get me y , (2) I *feel* trusting towards x for that purpose, and (3) I *treat* x as if it will get me y . I do not claim that all three components of trust are necessary in every instance, but I think that they are present in standard cases. I have already mentioned the belief element of trust, but a look at an example shows us that it is not sufficient, even when accompanied by the appropriate behavior.

Suppose I want to get to my niece's wedding, but I am phobic about flying. I may believe that the plane will get me safely to my destination and act as if it will do so, but I might still feel fear, have doubts, face indecision about getting on the plane, and regret my decision as the plane is taking off. All of this is compatible with believing that the plane will get me safely to my destination and acting as if it will. But it seems to me I do not trust the plane to get me safely to my destination if I am in the grip of fear and doubt. Trust includes an emotional element, a feeling that cannot be identified any more precisely than simply "the feeling of trust." I lack that feeling in the case I am imagining.

Annette Baier defines trust as "acceptance of vulnerability to harm that others could inflict, but which we judge that they will not in fact inflict" (1995a: 152). This suggests to me that trust includes an element of dispelling doubt with the awareness that doing so makes one vulnerable. Trust is an attitude opposed to doubt.¹⁶ But I have proposed that it is an attitude that can be appropriately directed towards nonpersons, so even though it may include awareness of vulnerability to harm, the harm need not be something another person can intentionally inflict upon us.

I think that the three components of trust are included in my prereflective trust in my faculties. I *treat* them as if they will get me to the truth, I *believe* that they will get me to the truth, and in addition, I *feel* a trusting emotion

16. Joseph Raz has suggested to me that the connection between trusting and the ability to doubt indicates that a creature incapable of doubt would be incapable of trust. In fact, it might mean that self-trust is required even to maintain a belief for any appreciable length of time since having a belief means that one treats an issue as settled sufficiently to dispel doubts. If, so, very young children do not trust because they lack the conceptual resources to doubt. When I say that there is prereflective epistemic self-trust, the state to which I am referring is not the state of an infant. It is prereflective in that the agent has not reflected on his epistemic faculties to any appreciable extent, but such a person has many beliefs, has reflected on many things, and is capable of doubting and dispelling doubts.

towards my faculties in that respect. Again, I do not claim every prereflective person has all of these components all of the time, but I think they are present in the prereflective state of most normal persons prior to reflection upon the connection between their faculties and success in reaching the truth.

It follows from what I have said that the natural belief that the natural desire for truth is satisfiable is included in the natural trust in one's epistemic faculties. That is, the prereflective self includes both the natural desire for truth and a natural trust in my epistemic faculties for the purpose of satisfying that desire. Trust includes relying upon my faculties for the purpose of getting my questions answered, believing that they can get my questions answered correctly, and feeling the attitude of trust towards my faculties for that purpose.

The faculties I rely upon in forming beliefs operate on an environment, so trusting my faculties includes trusting that the environment is appropriate to the faculties. My faculties may operate on the environment directly, or they may operate indirectly through the faculties of others. It is natural to believe what other people tell me. Trust in my faculties and environment includes trust in the faculties of many other persons.¹⁷ Trust in others, like trust in the self, is the starting point. It is a component of the prereflective self.

3. THE DESIRE FOR TRUTH AND THE REFLECTIVE SELF

I cannot say how common it is to reflect upon one's desire for truth or one's belief that the desire is satisfiable, but philosophers reflect upon everything. As Thomas Nagel says, philosophy is nothing more than the most disciplined form of self-consciousness (2010: 9). Reflecting upon the belief that truth is attainable is something to which a disciplined self-reflective being is eventually driven. When she does so, does she gain or lose confidence in it? Does she decide that the belief is reasonable? What does that do to her feeling of trust in herself and others? In this section I will argue that trust in oneself is rational upon reflection.

The simplest way to see the rational need for epistemic self-trust is to notice the phenomenon of epistemic circularity, or what Keith Lehrer

17. The classic expression of this point is Thomas Reid, "An Inquiry Into the Human Mind," Pt. VI, sec. 24, in Reid (1997: 196–97).

(1997) has called “the loop of reason.” A number of philosophers have observed that there is no noncircular way to determine that the natural desire for truth is satisfiable, or to put the claim in the preferred idiom, there is no noncircular way to tell that our belief-forming faculties are reliable as a whole.

Richard Foley (2001) links the phenomenon of epistemic circularity with the lack of answers to the radical skeptic and the failure of the project of strong foundationalism. We can do everything epistemically that we are supposed to do, including following the evidence scrupulously, but we have no assurances that the results will give us the truth or even make it more probable that we will get the truth. Foley concludes that we need self-trust in our epistemic faculties taken as a whole, together with our prereflective opinions. Self-trust is necessary, and further, he argues that it is rational in that it is a state to which we are led by the process of rational self-criticism. One is rationally entitled to self-trust and one is entitled to the degree of confidence one has in one’s opinions and faculties after critical reflection (25, 47). Foley therefore connects the theory of rational belief with a theory of self-trust.

Foley’s thesis that we are entitled to our confidence when it withstands self-criticism seems to me to be right. But for Foley, the need for epistemic self-trust depends upon the failure of adequate answers to skepticism. Self-trust is not a component of the prereflective self, but is a state to which we must move when we reflect upon the skeptical hypotheses and the failure of responses to them, particularly the failure of foundationalism. While Foley does not say that self-trust would be unnecessary if there was an adequate answer to the skeptic, his argument explicitly arises out of his view of the skeptical threat. He implies that trust is a state to which we retreat when we do not have adequate justification, or a “guarantee” of the reliability of our faculties and opinions taken as a whole.

William Alston offers a more detailed argument for a related conclusion about circularity in his final book, *Beyond Justification* (2005), which modifies an argument in Alston (1986). Alston argues that we cannot justify any belief arising from a basic practice of belief-formation (perception, memory, introspection, rational intuition, induction, and others) without justifying the well-groundedness of the practice, but we cannot do that without using that same practice. For instance, I cannot justify my belief that a dive-bombing hawk just swooped by my window without a justification of the

reliability of my perceptual faculties, but I cannot justify my belief in the reliability of my perceptual faculties without using perception.¹⁸ This is a stronger claim than the one made by Foley. Alston argues that circularity arises in the attempt to establish the reliability of individual basic sources of belief such as perception, memory, and deductive reasoning, whereas Foley claims only that circularity arises in the attempt to establish the reliability of our epistemic faculties and beliefs taken as a whole.¹⁹

Another interesting difference between Alston and Foley is that Alston does not think that the problem of epistemic circularity is necessarily tied to the threat of skepticism. He says that the specter of skepticism is a dramatic way to put the issue, “but it is not necessary for a calm, fully mature consideration of the problem” (2005: 216).²⁰ The problem as Alston sees it is that the fact that the justification of our beliefs is ultimately circular prevents us from being “fully reflectively justified” in our beliefs. We need not be especially worried about evil geniuses and brains in vats to notice circularity, and we need not think that the alternative to full reflective justification is skepticism.

18. Alston (2005) says that circularity can be avoided by keeping the targets very narrow (e.g., the reliability of perceptual beliefs about pies (205)), but as long as the issue is the reliability of broad sources of belief, the attempt to establish the reliability of beliefs deriving from that source will inevitably take us back to the source from which we started (209–10).

19. Alston’s position on the circularity of justifying basic sources of belief has an interesting twist. He argues that circularity does not prevent us from using an inductive argument to establish the conclusion that a doxastic practice such as sense perception is adequately grounded (1986: 202–3). An inductive argument of this sort is not logically circular, given that the conclusion does not appear in the premises, but it is epistemically circular in that one’s confidence in each premise depends in practice upon the assumption of the reliability of sense perception (SR). Nonetheless, epistemic circularity does not prevent us from being justified in believing each of the premises in the argument, nor does it prevent us from being justified in believing that SR follows from the premises. And so epistemic circularity does not prevent us from being justified in believing the conclusion, SR. Furthermore, epistemic circularity does not prevent us from knowing the premises of the argument and the conclusion. In fact, it does not even prevent us from being justified in the higher-order belief that we are justified in believing SR. All of this follows as long as (a) sense experience is in fact reliable, and (b) one can be justified in believing *p* without the ability to justify each premise of an argument showing that one is justified in believing *p*, and justifying each premise of *that* argument, and so on.

20. He says, however, that he will pursue the discussion in the following pages in terms of the “more dramatically attractive” skeptical challenge. His response to epistemic circularity two pages later is therefore framed as a reply to the Pyrrhonian skeptic.

I think Alston is right about that. A problem arises as soon as a person reflects upon her desire for truth and carries reflection upon that desire as far as she can. She is doing what every self-reflective being does, only more thoroughly and scrupulously. She feels dissonance when she lacks full reflective justification for her beliefs, and that is a problem even if she does not fear skepticism or even pay any attention to skepticism.

Although we have no noncircular way to tell that our faculties are reliable, we can sometimes tell in a noncircular way that our faculties are *unreliable*. Consider a simple analogy. Suppose the thermostat on your air conditioning system is set at 72 degrees Fahrenheit, and imagine that the system can read itself. If it is set at 72, it “knows” it is set at 72, and it knows what it reads when it goes on and off. The system could tell that it is malfunctioning if it failed to go on when the thermostat read above 72, but it would not know that it is accurate if it regularly went on at 73 and off at 72. Accuracy requires a certain relation between the system as a whole and the ambient air temperature. If it gives unstable or inconsistent readings, it “knows” it is unreliable, but it has no way to tell that it is reliable if it gives consistent readings. In the same way, if we regularly had inconsistent memories or unstable perceptions, we would know that something was amiss with these faculties; they could not be trusted. But consistent and stable memories and perceptions are not sufficient to tell that they are reliably accurate.

Alston’s response to epistemic circularity is this:

Proceed to form beliefs and rely on them (take them to be credible, take them to be at least probably true), using various modes of belief formation that we find ourselves in possession of and the reliability of which we find ourselves strongly inclined to trust. All this without already having shown them to be reliable....

But, the critical philosopher might say, isn’t that arbitrary, dogmatic, or deserving of some other term of epistemic censure? How can it be rationally acceptable to employ a way of forming beliefs without having established its claim to a sufficient degree of reliability? Well, it must be conceded that this goes against a strong aspiration of epistemology—to refuse to use any way of forming beliefs unless it has successfully run the gauntlet of philosophical criticism. But what we have just seen is that achieving this is strictly impossible, as impossible as squaring the circle or being in two widely separated places at the same time. Hence the better part of wisdom is to recognize that fact and get over the yearning for the impossible. (2005: 218)

Like Foley, Alston thinks of self-trust as the outcome of a sophisticated line of argument. We are forced into it by careful reflection on the human epistemic condition. If we could be fully reflectively justified in our beliefs, presumably we would not need to “take” our faculties to be reliable and our beliefs to be credible. We would not need self-trust because we would have something in principle better, but impossible to achieve. Similarly, Foley implies that if strong foundationalism had succeeded or if we had some other adequate answer to skepticism, trust in the self would not be needed. So for both of them trust is a fallback position, something we accept when we cannot get what we really want—proof or a strong form of justification. And for both of them trust is a state to which we retreat after philosophical reflection. It is an end state, not the state from which we start.

I differ from them on both points. I have suggested that there is prereflective self-trust. Before we reflect upon the justification of our beliefs or the reliability of our faculties, we already trust ourselves and our environment, including other people. Foley and Alston think of trust as a fallback position because they think of it as something we have when we do not have something else that in principle would be better: proof. They think that trust in our faculties rests upon the realization that we lack proof of the connection between our faculties and getting the truth. The assumption seems to be that we do not trust the connection between x and y unless we lack proof of the desired connection between x and y and realize that we lack the proof.

This way of looking at trust makes sense in certain contexts. People often say they trust their mates when they lack proof of fidelity and are aware that they lack proof. But trust in your spouse need not be a state to which you retreat when you lack the proof you would prefer to have. If you had proof, you might not change your disposition with regard to your spouse at all, and it would be odd to say, “I no longer trust you” as a response to proof of fidelity. As I see it, the relevant feature of trust is that it does not *depend* upon proof, not that it requires the lack of proof. When there is lack of proof, we say you must “fall back” on trust, but the belief in the spouse’s fidelity and feeling of trust must have preceded the realization of lack of proof. Similarly, I do not think that trust in one’s epistemic faculties depends upon the outcome of philosophical arguments. If we could escape epistemic circularity by getting a guarantee that our faculties get us the truth, that would be like the case in which you have proof of your spouse’s fidelity. I do not see that trust would disappear. The difference is that awareness of epistemic circularity

forces us to confront the prereflective trust we have in ourselves at the reflective level.²¹

4. SELF-TRUST AND THE ALTERNATIVES

Is it rational to have self-trust after reflection? That depends, of course, on what we mean by rationality, and whether it applies to all three components of trust, including the behavioral and feeling components. I said above that I think of rationality in the broad sense as doing a better job of what we do naturally in the use of any of our faculties, and I see no reason to exclude our emotion dispositions and overt behavior from the domain of the rational. Reflective self-trust resolves the dissonance we have when we discover epistemic circularity, and that seems to me to be rational. It is rational to believe that my faculties are trustworthy for the purpose of getting the truth; it is rational to treat my faculties as if they will get me to the truth, and it is rational to feel trusting of them in that respect. Of course, if someone thinks it is not rational to have a belief without noncircular justification, then self-trust is not rational, but then none of our beliefs is rational. However, I see no reason to think that that is what rationality is.

Would it be rational for a person to respond to the problem of epistemic circularity by not trusting her faculties? Since trust has more than one component, there is more than one way she might lack trust. After reflection I could give up any of the aspects of trust. It is hard not to treat my faculties as trustworthy even after understanding circularity, but I know philosophers who claim that they do not *believe* that their faculties are trustworthy even though they act as if they do. I am not sure whether the people I have in mind feel trusting or not, but it seems to me that to treat something as deserving of trust without believing it is deserving of trust creates dissonance in the self that becomes noticeable once we reflect upon it. When I become aware that I treat myself as epistemically trustworthy, I feel pressure within myself to either believe that I am trustworthy or to stop treating myself as trustworthy. It is possible to accept the dissonance or not to notice it, so I do

21. It could be argued that whereas the disposition to trust does not depend upon the lack of proof, the disposition is only activated when proof is lacking. I thank Martin Montminy for this point. This way of looking at trust leads to a different way to distinguish prereflective from reflective trust. The former is merely dispositional. The latter is activated.

not insist that it is impossible to live a normal life without believing that our epistemic faculties are trustworthy.²² But the self-reflective person at some point will become aware of the dissonance if she does not believe her faculties are trustworthy, and will then have to decide whether to accept the natural belief that her natural desire for truth is satisfiable or else live with dissonance.

The same point applies to the dissonance produced by lacking the feeling component of trust. Someone might judge that her faculties are trustworthy in getting her to the truth and treat them that way, but she might continue to be plagued by doubts. She might feel this way because she obsesses over the skeptical hypotheses, but the more interesting case is one in which she is a person who just reflects excessively and never *feels* that the issue is settled even when she judges that it is. Karen Jones (forthcoming) gives an amusing example of a woman who believes she has her passport in her purse, in fact knows that it is there, but checks obsessively in the taxi to the airport to make sure that she has it. The problem is not that she thinks an evil genius might have stolen it; there is no particular hypothesis that generates her doubts. She simply feels untrusting, even when she not only believes that the passport is in her purse, but believes she has done everything a reasonable person can do to believe truly. It seems to me that a person can obsess over the trustworthiness of her faculties in the same way. Even when she believes they are trustworthy and acts as if they are by living a normal life, she may not feel trusting of them. The feeling of doubt may continue to plague her, and she cannot dispel it because she lacks the feeling aspect of trust. Is she rational? As I have said, many people think that rationality is not at issue when we are talking about feelings, and I will not argue about that usage, but in the broad sense of rationality I have been using, she is not rational because of the dissonance between her feelings and her beliefs. It is rational to resolve dissonance between components of the self, and she has not done so.

22. Alvin Plantinga says proper functioning demands that we trust the reliability of our faculties so as to prevent "cognitive disaster," and he quotes Hume's remark that if we find reason to doubt the reliability of our faculties and sink into philosophical melancholy, nature will, fortunately, "cure me of this delirium, either by relaxing this bent of mind, or by some avocation... which obliterates all these chimeras. I dine, I play a game of backgammon, I converse, and am merry with my friends" (Plantinga 2002b: 210).

There is, finally, the most radical response to epistemic circularity—the skeptic who, upon reflection, neither believes her faculties are trustworthy, feels trusting of them, nor treats them as trustworthy. Perhaps the Pyrrhonians were like that. Maybe they lacked all of the components of trust I have identified. If there are such people, they would not face dissonance between and among their beliefs, feelings, and behavior. To avoid dissonance, they would also have to give up the natural desire for truth, and maybe they are able to do that as well. I have my doubts that there are such persons, but let us suppose that there are. Are they rational? The radical skeptic I am imagining does not have the irrationality of dissonance, but she attains that by foregoing much of what we do naturally. I have suggested that rationality is, roughly, doing a better job of what we naturally do. The extreme skeptic I have described is not doing a better job of what we do naturally because she is not doing what we do naturally. But I have no interest in critiquing skepticism of any sort. Epistemic self-trust is rational in the sense I have described, and it is more rational than alternatives in that it requires making the fewest adjustments in the prereflective self. Self-reflection is what a self-conscious being does, and a rational being does it carefully. However, it is possible to go on reflecting forever about whether x is trustworthy in respect to y . Trust ends the process of reflection, and it is rational because excessive reflection is not rational. Trust prevents excessive reflection, and in my view, it is an essential component of a rationally self-reflective being.

5. THE CONSCIENTIOUS BELIEVER AND THE NATURE OF REASONS

Our investigation of the place of self-trust in our epistemic lives leads to a way of thinking about reasons for belief that makes them derivative from what we do when we seek the truth. When we want our questions answered, what we typically do is to look for reasons that, upon reflection, can be put together in such a way that they seem to support a given conclusion, a conclusion that we then take to be true. What we call justification (in one of its senses) is the state we are in when we succeed in finding reasons of that kind. The desire for truth in a self-reflective person leads to the search for reasons in this way, and the arguments for epistemic circularity by Foley and Alston

make this assumption. We trust that there is a connection between the possession of reasons for belief and getting the truth.²³

A self-reflective person who desires truth may not search for reasons for every belief. The issue of whether there are beliefs that a self-reflective person accepts without reasons is an important one, and I am not assuming that there are no such beliefs. What is not disputable is that a self-reflective person looks for reasons for many of her beliefs, and she considers it a good thing to have reasons for any of her beliefs. I do not think there is any explanation for why she does this or what would justify her in doing it except that that is what self-reflective persons who desire truth do. To have reasons for her beliefs produces psychic harmony, and to fail to have reasons produces psychic dissonance.

This leads to the question whether we have reasons to think that our reasons for belief lead to the truth. The same question can be posed in terms of the related notion of evidence.²⁴ Do we have evidence that evidence leads to truth? In any sense of evidence that would eliminate the need for trust in the relation between evidence and truth, the answer is no. For one thing, we do not have evidence that evidence leads to truth. What we have is evidence that evidence for p leads to more evidence for p , enough that at some point we declare p true. But in any case, why should we pay attention to the evidence that evidence leads to truth unless we trust the connection between evidence and truth? No matter how much evidence we have, its connection to truth will always be something that cannot be established without circularity.

23. It is possible that a self-reflective person would desire to be justified in her beliefs independent of the desire for truth. Being justified might be judged by a self-reflective person to be intrinsically valuable. I am not denying the intrinsic value of justification. My point is that the natural desire for truth is sufficient to lead to the desire to be justified in the sense we are discussing, not that it is necessary.

24. The notions of reasons and evidence are closely connected, but there are some differences. One difference is that we usually speak of reasons *for* a given belief, whereas evidence can be gathered when any belief for which it is evidence is not yet in play. So we would not normally speak of having reasons without indicating what the reasons support, whereas we might say we have evidence when we have no idea what the evidence indicates. Some philosophers make evidence a narrower category than reasons, limiting evidence to reasons of a certain kind—e.g., propositional beliefs. For instance, Plantinga (1991) does this in his well-known attack on evidentialism. In this terminology, an experience can give a person a reason to believe a proposition, but it is not evidence. Another difference is that evidence is sometimes thought to be objects that point to truth, such as fingerprints. A fingerprint could be evidence, but it is not a reason. I mention evidence in this sense below.

I do not think there is any deeper answer to the question why it is rational to trust evidence except that trusting the connection between evidence and truth is something rational people who desire the truth do. Even the skeptic trusts this connection. In fact, it is because the skeptic trusts this connection that the skeptic becomes a skeptic. It is the failure to complete the search for evidence that leads her to skepticism.

Circularity is relevant to the desire for truth because we make certain assumptions about the nature of mind and the universe. We want truth—our questions answered, and we notice that the process of attempting to answer those questions can never be completed. But this is a problem because we assume (a) there is a connection between successfully getting the truth and what we do when we attempt to answer our questions (what we call finding reasons), and (b) what we attempt to do can never be completed. The discovery of epistemic circularity is the discovery of (b), but what about (a)? We do not discover (a); we trust it. The need for trust in (a) is independent of (b), and we can see that by looking at what our situation would be like if, *per impossibile*, we were able to complete the search for reasons in a noncircular way. We would still need trust that there is *any* connection between reasons and truth. Whether or not we have the reasons we seek, we need to trust that reasons are the sorts of things that give us the answers to our questions, that connect us to truth. So even if strong foundationalism had succeeded, we would need trust that we identified the foundation correctly and that the foundationalist structure reliably gives us truth. And this is no less the case if the foundation is certain.²⁵ We would still need trust in the connection between the state of certainty and truth.

We get the same conclusion no matter what notion of evidence or reasons we use. Evidence can be understood as something internal to the mind—generally, a phenomenal experience or a belief. In contrast, evidence is sometimes understood as public property, the sort of thing to which scientists or lawyers can point in the common project of attempting to answer questions. If reasons are public, they can be either objects, such as fingerprints, or facts (true propositions).²⁶ Alternatively, reasons could be some

25. I am not denying that there might be intrinsic value in having a foundationalist structure that is transparent to the mind. My point applies to the connection between such a structure and the satisfaction of the desire for truth.

26. See Kelly (2006) for an excellent summary of the various senses in which people speak of evidence.

combination of the public and private, such as facts known by the subject. There are many variations, but in every case, trust is needed. If a reason for belief is internal to the mind, the need for trust in the connection between a reason and something external to the mind is clear. If instead a reason is defined from an external perspective, what in fact indicates truth, that means that what we do when we are attempting to get truth may not be having a reason in the sense defined. That does not remove the need for trust, it just backs it up a step to trust in the link between what we do when we are trying to get truth and having a reason from an external perspective. The same point applies to evidence. Whether or not we define evidence in a way that builds a reliable connection to the truth into the concept, we need to trust the connection between (a) what we do when we make a fully conscious effort to use our faculties the best way we can to get truth and (b) success in reaching truth. Using an externalist notion of reasons or evidence therefore does not remove the need for trust in the connection between our faculties and getting the truth.

The fundamental reason we trust evidence or reasons is that looking for evidence is what we do when we are self-reflective, and we trust that. I call the quality of using our faculties to the best of our ability in order to get the truth *epistemic conscientiousness*. I think of this quality as the self-reflective version of the natural desire for truth. It is a natural desire brought to self-reflective consciousness and accompanied by the attempt to satisfy it with all of one's powers. I have argued that we need trust that there is a connection between the natural desire for truth and the satisfaction of that desire using the faculties that any person has, reflective or prereflective, but once a person becomes reflective, she thinks that her trustworthiness is greater if she summons her powers in a fully conscious and careful way, and exercises them to the best of her ability. What I am calling conscientiousness is the state or disposition to do that.²⁷ Conscientiousness is important because we do not think that we are equally trustworthy at all times. We trust that there is a connection between trying and succeeding, and the reflective person thinks that there is a closer connection between trying with the full reflective use of one's powers, and succeeding. Conscientiousness comes in degrees. There is a probably a degree of conscientiousness operating most of the time since we have some awareness of ourselves and the exercise of our

27. Note that as I define conscientiousness, it does not have any relation to duty.

powers most of the time. But higher degrees of conscientiousness require considerable self-awareness and self-monitoring.

A conscientious person has evidence that she is more likely to get the truth when she is conscientious, but she trusts evidence in virtue of her trust in herself when she is conscientious, not conversely. Her trust in herself is more basic than her trust in evidence, and that includes evidence of reliability. The identification of evidence, the identification of the way to handle and evaluate evidence, and the resolution of conflicting evidence all depend upon the more basic property of epistemic conscientiousness. I think, then, that evidence is what we take to be indicative of truth when we are conscientious, and we trust that that is identical with what is indicative of truth.²⁸ Norms of reasoning such as the rules of probability are tools for helping us figure out what is most conscientious to believe.²⁹ Likewise, I think that intellectual virtues are qualities that arise out of epistemic conscientiousness. These qualities are those that epistemically conscientious persons endorse and attempt to acquire. But we would not treat them as virtues unless we thought that our cognitive and sensory faculties are generally trustworthy because these qualities are useless in a being whose faculties are not naturally conducive to reaching their end.³⁰

It follows from what I have argued in this chapter that there are two levels of self-trust, both of which are more basic than any reasons or evidence we can identify. First, there is the general trust in our faculties that I argued in section 4 is the most rational response to epistemic circularity. Second, there is the particular trust we have in our faculties when we are conscientious—

28. Since I am proceeding from the first-person viewpoint, we could say that evidence is ultimately what *I* take to be indicative of truth when I am conscientious, but as we will see in chapter 3, when I am conscientious, I commit myself to trusting conscientiousness in others, so evidence includes what I conscientiously believe conscientious others take to be indicative of truth.

29. For this reason, trusting my faculties as a whole cannot mean trusting that, taken as a whole, using my faculties makes it more probable than not that I get the truth. That would make it too easy to convince myself that most of my beliefs are true. Believing a faculty is trustworthy does not include making a judgment of probability. My position is that judgments of probability depend upon a prior belief in the trustworthiness of my faculties.

30. In Zagzebski (2009: 82) I argue that qualities like intellectual attentiveness, carefulness, thoroughness, and openness to new evidence are forms of epistemic conscientiousness. But it does not do us any good to be careful, thorough, etc. unless our faculties put us generally on the right track. So we assume the general trustworthiness of our faculties when we treat these qualities as virtues.

exercising our truth-seeking faculties in the best way we can. Our identification of reasons for belief, norms of reasoning, and the qualities we think are intellectually virtuous are all derivative from what we do when we are epistemically conscientious. My judgment that the evidence supports some proposition p is not trustworthy without trust in my faculties, in particular, the conscientious use of my faculties. That means that trust in my faculties is always more basic than any judgment about the evidence and what it supports. Trust in myself is more basic than trust in my judgment of the reliability of myself or anyone else. This point will become important in later chapters.

The line of reasoning of this chapter has the consequence that ultimately our only test that a belief is true is that it survives future conscientious reflection, including reflection on future experiences, and future judgments about the past and present. The role of future conscientious self-reflection means that there is an important way in which the future justifies the present, and I will have more to say about that in later chapters. Here I am pointing out that there is a connection between what I take to be reasons for a belief, and what I take the truth to be. The way I identify what points to the truth is the same as the way I identify what the truth is. All I can do is to use my faculties the best way I can, that is, to be conscientious, and since future conscientious reflection includes reflection on the products of past conscientious reflection, survival of reflection in the future is my only way to tell that my past reflections have gotten me to the truth. Of course, I am not proposing that truth is what survives conscientious self-reflection. If it were, we would not need trust in conscientious self-reflection. But conscientious self-reflection is basic. There is nothing more fundamental in our psychic economy.

At the beginning of this chapter I proposed that the experience of dissonance unconsciously resolved gives us our first model of what rationality is. Given what I have argued in the rest of this chapter, the experience of the conscientious self resolving dissonance gives us our second model of what rationality is. It is rational to trust when it is needed to resolve dissonance, and epistemic self-trust is rational in this sense. But I think now that we can see that self-trust is a reason for belief, but a reason of a distinct kind. If a reason to believe p is something in virtue of which a conscientious person thinks some proposition p is true, then self-trust is a reason because it is in virtue of self-trust that I conscientiously think that what I take to be reasons for believing p are truth-indicators. It is in virtue of self-trust that I believe

everything I believe—that there are icicles on our little waterfall, that Obama is president, that sunlight is not good for the human skin, and so on for all of my other beliefs. Self-trust is an interesting kind of reason because not only does it have an affective component, it is first personal; it is a reason only I can have. I will return to the difference between first-personal and third-personal reasons in later chapters. My purpose in this chapter is to argue that given the place of reasons in our psychic structure, either self-trust is in the category of reasons or there are no reasons. That is why it is rational to have self-trust. Self-trust is the foundation of what we take rationality to be.

Epistemic Trust in Others

1. EPISTEMIC EGOISM

Trust in the self is natural and survives critical self-reflection, but what about trust in others? Credulity with respect to the beliefs of other people is also natural, but it is evident that there has been resistance to it in the modern era. Everyone agrees that if we did not depend upon others, we would have far less knowledge because we would have far fewer beliefs. Nonetheless, trust in others is escapable. The issue I want to raise now is whether it is rationally escapable. In this chapter I will argue that if we trust ourselves epistemically, we cannot consistently fail to trust others. We must not only believe they are trustworthy, but unless we accept *ethical* egoism, we must actually treat them as trustworthy, and in many cases we must treat them with the same trust we give ourselves. Epistemic self-reliance is not a coherent ideal.

The most extreme form of epistemic self-reliance is one in which someone refuses to take the fact that someone else has a given belief as a reason to believe it herself. Let us call this position *extreme epistemic egoism*. According to the extreme egoist, I have reason to believe p only when the direct exercise of my faculties gives me reasons for p . The fact that another person has a belief p gives me no reason to believe it. The beliefs of others are irrelevant to my reasons for having a belief because their beliefs are the outcome of the exercise of *their* faculties. As a self-reliant person, I aim to make my beliefs rest on my own directly acquired reasons. If someone wants to convince me that her belief is true, she needs to give me an argument to that conclusion from premises I believe and using forms of reasoning I accept.

It may be objected that it is impossible to separate the beliefs I form by the direct exercise of my own powers from those that I get from the exercise of the faculties of others. After all, we acquire language, concepts, and the tools for making judgments from other persons. Is it even possible to adopt a policy of acquiring and justifying one's beliefs only by the use of one's own powers?

I think the answer is yes. The important issue is not whether we acquired beliefs in our prereflective state from other people, but whether we can retreat to an extreme egoist position once we start reflecting. Compare extreme epistemic egoism about moral beliefs. An agent may in childhood acquire both moral beliefs and the conceptual resources to make moral decisions from parents and other adults, and then in adulthood take the position that she will retain no moral belief or acquire a new one unless she can determine the justification for the belief herself. That is a possible position to have, even a common one. There is nothing preventing a person from being an extreme epistemic egoist about moral beliefs. What I mean by the extreme epistemic egoist is a person who has the same position about all of her beliefs.

Here is an interesting argument for extreme epistemic egoism. Suppose I want to settle for myself whether p . When I reflect about it, I consider reasons relevant to the truth of p . But once I acquire the belief p , I do not treat myself as having acquired a new reason for p . That is, I do not say to myself that I now have more reason to believe p than I did before I made up my mind whether p . If so, the same point should apply to the beliefs of other persons. When I find out that somebody else believes p , her belief is not a reason to believe p . Her *reasons* for believing p can count as reasons for me to believe p , but if the fact that I believe p is not one of my reasons to believe p , then the fact that she believes p is not either. I think this argument is correct about parity between self and others, but it makes an interesting mistake, and I will get to that presently.

I have already mentioned the problem that extreme egoism is unrealistic for anybody who wants more than a paltry amount of knowledge. For this reason, philosophers tempted by extreme egoism generally back off of the position I have just described. For instance, we saw in chapter 1 that Elizabeth Fricker (2006b) toys with the ideal of extreme epistemic egoism (what she calls "intellectual autonomy"), but immediately acknowledges that

living according to its norm would be severely limiting.¹ Alternatively, someone might think that the extreme egoistic picture of a justified belief structure is attainable in principle and its hypothetical attainment is what justifies actual beliefs. So he could adopt beliefs on the testimony of others, but only because he believes upon reflection that those beliefs could be justified in principle by his own faculties even though he does not actually have the beliefs he would need in order to justify the beliefs he acquires from others. Modern scientific practice is sometimes idealized as capable of justification in roughly this way. An empirical scientist may accept the testimony of another scientist working in the same field, but think that that is justifiable only because the testimony of the source could in principle be verified by the direct use of his own faculties. Notice that even the hypothetical form of extreme egoism is a norm of nontrust in others.

There may not be many extreme epistemic egoists *tout court*, but as I have said, it is quite common for philosophers to be extreme egoists about moral beliefs, and I think also about religious beliefs. The objection that extreme egoism is impractical probably does not apply to those domains, and those who find extreme egoism an ideal are not forced by the demands of practicality to weaken their position in those areas. So philosophers who reject the ideal of extreme egoism *tout court* often accept it in the domains of value or religion. We will look at moral extreme epistemic egoism in chapter 8, and religious extreme epistemic egoism in chapter 9. In this chapter we will examine general forms of epistemic egoism.

There is a more moderate form of epistemic egoism that is more common than the extreme version. Those who hold this position maintain that the fact that another person has a certain belief can give him grounds for believing it, but only if, by using his own faculties, he has grounds for believing the source is reliable. Let us call this form of egoism *standard epistemic egoism*. The standard egoist agrees with the extreme egoist that no belief is justified unless it is ultimately based on reasons he has acquired by the direct use of his own faculties. The difference is that the standard egoist is willing to use his faculties to determine that he has reason to believe that another person is a reliable source of true beliefs. The reliability of the source must be justified

1. In addition to limiting our knowledge, extreme epistemic egoism is not apt to win any friends. Think how annoying it is when another person will not take your word for something that you are in a better position to know than he is because it does not seem probable on his own evidence.

by beliefs he has using his own perceptual faculties, memory, and reasoning.²

Both standard and extreme epistemic egoism are norms of belief that arise out of the value of epistemic self-reliance. As we saw in chapter 1, that value is hard to clarify and the story of the way it came to capture the modern mind is complicated. I offered the hypothesis that self-reliance gets much of its plausibility from its confusion with autonomy, and it has a romantic attraction, vividly described in the famous essay by Ralph Waldo Emerson. But in spite of the intuitive appeal of self-reliance, I think that epistemic egoism is false in both its extreme and standard forms. The issue is what we commit ourselves to do when we trust ourselves.

2. THE NEED FOR TRUST IN OTHERS

2.1. Why Epistemic Egoism Is Unreasonable

In chapter 2 I argued that I have no noncircular reason for thinking I am trustworthy as a whole. That means I have no noncircular reason to think that as a whole I am more trustworthy than other people are as a whole. But I might think that I have a reasonable defense of treating the faculties of others differently from my own. Let us consider whether I do.

As a self-reflective person, I trust the beliefs I form when I am epistemically conscientious, which I defined as using my faculties in the best way I can in order to satisfy the desire for truth. When I am conscientious I will come to believe that other normal, mature humans have the same natural desire for truth and the same general powers and capacities that I have. If I have a general trust in myself and I accept the principle that I should treat like cases alike, I am rationally committed to having a general trust in them also. Of course, in some cases I may have reason to think that some individual has defective or undeveloped epistemic powers, and if so, my belief in the general trustworthiness of that person would be defeated. In such cases I have good (ultimately circular) reasons to think that I am more generally trustworthy

2. The distinction I am drawing between extreme epistemic egoism and standard epistemic egoism is derived from Richard Foley's (2001) distinction between "epistemic egotism" and "epistemic egoism." In "Self-Trust and the Diversity of Religions" (2006) I used Foley's terminology, but changed it in Zagzebski (2008).

than some other person, but if I am honest, I must admit that those reasons apply to a very limited range of cases. Insofar as I have a general trust in the connection between my natural faculties and desire for truth, on the one hand, and success in reaching truth, on the other, then I should trust the same connection in other persons. When I see no relevant difference between others and myself, then given that I trust myself, I should trust them.

My reason for believing that other persons have the same natural faculties that I have is not *a priori* I since I do not know *a priori* I that there are other human beings who belong to the same natural kind as myself and who have the same general sensory faculties, memory, powers of reasoning, and desires. But the principle that I should treat like cases alike is *a priori*.³ Trust in myself means that there is a presumption in favor of the output of my faculties, so consistency requires me to have a presumption in favor of the output of the faculties of others who share my faculties, which is to say, virtually all other human beings. The default position is trust, not distrust.

I have said that epistemic trust in myself has three components: (a) I believe my faculties are generally trustworthy for the purpose of getting the truth, (b) I have a feeling of trust towards my faculties in that respect, and (c) I treat my faculties as trustworthy in that way. I am arguing now that under the assumption that I see no relevant difference between my general epistemic faculties and those of others, I ought to have the same attitude towards their faculties as I have towards my own. I should believe their faculties are generally trustworthy, feel trusting of their faculties, and treat them as trustworthy. That means that the deliverance of the faculties of another person regarding some proposition *p* counts positively in my deliberations about whether *p*. I cannot consistently trust my own faculties but not those of others. Extreme epistemic egoism must be rejected.

I also argued in the last chapter that there is a second level of epistemic self-trust that is my basis for distinguishing those cases in which I am trustworthy from those cases in which I am not. In addition to trusting my epistemic faculties in general, I have a particular, self-reflective trust in my faculties when I am exercising them conscientiously. But when I am conscientious and because I am conscientious, I not only acquire beliefs about the general faculties and desires of all other persons, I also acquire beliefs about

3. Cf. Tyler Burge's (1993) argument that we have *a priori* I entitlement to believe what others tell us.

particular persons. When I am believing conscientiously, I come to believe that many of them are just as conscientious as I am when I am as conscientious as I can be. At least, many of them are as conscientious as I am in certain respects or with respect to certain domains of belief. It follows that because I place particular trust in myself when I am conscientious, I must place particular trust in others whose conscientiousness I discover when I am being conscientious. The general principle is that insofar as I trust myself in virtue of having certain properties, I owe the same trust to others whose possession of those properties is something I discover when I am behaving in a way I trust.

One of the things I do when I am conscientious is to look for reasons or evidence for my beliefs. A conscientious person might not expect evidence for every belief, but to the extent that I have evidence, the evidence cannot be conscientiously ignored. When I am behaving conscientiously, I will recognize that I have plenty of evidence, in whatever sense of evidence you want, that there are other persons who are as conscientious as I am with regard to numerous beliefs. I also have evidence that some individuals are more conscientious than I am, and I have evidence that some individuals are less so.

Since my identification of evidence is derivative from what I do when I am epistemically conscientious, it is reasonable to trust myself when I am conscientious in advance of evidence of my reliability. That was one of the conclusions of chapter 2. So if I conscientiously believe that many other persons are as conscientious as I am in certain respects, I must grant them *prima facie* trust in advance of evidence of their reliability in those respects. I am committed to that by the principle that I should treat like cases alike. When I have evidence of someone's conscientiousness, I have reason to trust her faculties in the same way I trust my own when I am conscientious. Of course, in the situations of real life, it may be just as easy or even easier to get evidence of her reliability than of her conscientiousness. But the crucial point is that if, in believing in a way I trust, I come to believe that other persons have the property in virtue of which I trust myself, then I am committed to trusting them also.⁴ It follows that standard epistemic egoism is false. It is

4. It might be objected that there is evidence from the social psychology literature that people are often not very conscientious in a number of identifiable contexts. But this gives me reason to distrust both others and myself in such contexts. It does not support an egoistic bias.

not reasonable to expect evidence of the reliability of another person in advance of trusting her epistemically.

This argument leads to the consequence that there is a presumption in favor of the output of the faculties of others whose conscientiousness I conscientiously discover. The fact that a conscientious other person believes *p* gives me a *prima facie* reason to believe *p* myself. When I say a *prima facie* reason, I do not mean an apparent reason. I mean one that has genuine weight, although it can be defeated or outweighed by other reasons. Having a *prima facie* reason to believe *p* is not a decisive reason to believe *p*. Nonetheless, it is a genuine reason.⁵

Another point we have learned about self-trust applies to trust in others. The fact that someone else conscientiously believes *p* would not give me a *prima facie* reason to believe *p* unless the fact that someone else believes *p* is already a *prima facie* reason to believe *p*. That is to say, a person would not be trustworthy when she is using her faculties as well as she can unless those faculties were generally trustworthy. So we are committed to the following principle:

Epistemic universalism

The fact that another person has a certain belief always gives me *prima facie* reason to believe it.

This is a weak form of epistemic universalism.⁶ It means that the fact that another person has a belief *p* is a mark in favor of *p*. I have also argued that the fact that another person conscientiously believes *p* is a stronger mark in favor of *p*—a stronger *prima facie* reason. This stronger reason would presumably not apply to all other persons since we do not conscientiously judge that all other persons are conscientious, at least not all of the time.

5. Shelly Kagan distinguishes a *pro tanto* reason from a *prima facie* reason on the basis that the latter is something that only appears to be a reason, but upon further consideration may turn out not to be a reason at all, whereas the former is a genuine reason (1989: 17). Kagan says that what W. D. Ross called *prima facie* duties are really *pro tanto* duties. I prefer the term "*prima facie*," and will sometimes refer to a *prima facie* case for something, or will say that something is *prima facie* trustworthy. The use of "*pro tanto*" in some of those contexts is awkward, but I do not insist that my usage should be preferred by the reader.

6. Richard Foley (2001: 105–7) supports weak epistemic universalism. He also says that consensus adds credibility (121).

I want to stress that this point is not about trust in testimony. It is about the reasonable response to conscientiously believing that other persons are relevantly like myself, that they have whatever property I have that I trust in myself, and that the outputs of their faculties are relevantly like the outputs of my faculties. It does not matter whether they tell me anything. In fact, it does not matter whether they tell anybody what they believe. I have not yet raised the issue of how they relate to me and I to them. I am only talking about how I should think of their faculties and epistemic capacities in comparison to my own.

The standard or extreme epistemic egoist might make the following objection. She might concede that the path from self-trust to trust in others succeeds in showing that she is committed to *believing* that all other persons are generally trustworthy, and that many other persons are just as trustworthy as herself. But the belief component of trust is only one of its components. It does not follow that she is committed to actually trusting any of these people. The idea here is that she has no obligation to treat everyone as trustworthy whom she believes to be trustworthy. If she prefers to trust herself and not others, or to trust herself more than others, why shouldn't she do that? Why can't she reasonably refuse to trust others even though she cannot reasonably refuse to think of them as trustworthy?

My response is that she cannot do so if she cares about truth. If she does not trust others, she must ignore her own evidence that other persons are trustworthy; she must ignore the beliefs to which she is led when she is conscientious—when she forms beliefs out of a reflective desire for truth. The only way to do that is to care about her own faculties and their outputs more than the truth; to care about her own evidence, not because she thinks it indicates the truth, but because it is hers; to care about her own conscientious beliefs, not because as a conscientious person she cares about truth, but because they are her beliefs. She must epistemically trust herself, not because of her conscientiousness or even because of her natural human faculties, but because she is herself. And that is very implausible. This problem arises whenever I trust myself more than others when my evidence or the conscientious use of my faculties indicates that they are as conscientious as I am. To the extent that I epistemically trust myself more than I trust them, the fact that my faculties are mine trumps the desire for truth, and so I must be valuing my own

faculties more than the truth. That is ethical egoism in the realm of the intellect.⁷

Assuming I do not want to be an ethical egoist, I am rationally committed to not only thinking of others as trustworthy, but to actually trusting them on the same grounds as I trust myself. I must have the same attitude of general defeasible trust in all others that I have towards myself because they have the same faculties I trust in myself, and I must acknowledge that the level of trust that I have in myself when I am conscientious applies to many other persons whose conscientiousness I conscientiously discover.

What about the feeling component of trust? In chapter 2 I argued that it is more rational to feel trust in the self than not to feel trust since it is rational to resolve or prevent dissonance in the self, and the feeling of trust leads to more harmony within the self than believing that one is trustworthy without the feeling of trust. If a feeling is appropriate in some set of circumstances, it is also appropriate in relevantly similar circumstances, so given the argument of this chapter, it is appropriate to feel epistemic trust in others. The feeling of trust in others is rational in the same sense in which the feeling of self-trust is rational. I agree that it is probably going too far to say that feeling trust in either oneself or others is a rational *demand*, in the sense that I can be blamed if I lack the feeling. As I remarked at the beginning of chapter 2, much of what we do to resolve dissonance in the self is unconscious, and even at the reflective level, our degree of conscious control is limited. Control over feelings is particularly difficult to achieve. But this point about lack of control over feelings does not weaken the parity between trust in the self and trust in others. We will focus on trust in emotions in chapter 4.

7. The egoist might bite the bullet and agree that he values his own faculties and their products more than the truth even when the faculties are epistemic. That is, in fact, what Emerson says. "On my saying, 'What have I to do with the sacredness of traditions, if I live wholly from within?' my friend suggested—'But these impulses may be from below, not from above.' I replied, 'They do not seem to me to be such; but if I am the Devil's child, I will live then from the Devil.' No law can be sacred to me but that of my nature. The only right is what is after my constitution; the only wrong what is against it" (1903: 50–51). Here Emerson explicitly makes self-reliance more important than getting the truth.

2.2. Epistemic Egocentrism

Even though epistemic egoism in either its extreme or standard forms is mistaken, there are ways in which epistemic egocentrism is inescapable. Trust in oneself and trust in others are not symmetrical. For one thing, the latter depends upon the former, not conversely. I am committed to thinking of other persons as epistemically trustworthy if I think of myself as epistemically trustworthy, but my trust in myself is basic. I have argued that I must have a general trust in myself and my faculties, and since I am reasonable in believing that other people have the same faculties and the same natural desire, I am reasonable in trusting them and their faculties for the same reason I trust my own. Similarly, the reflective trust I have in others who are conscientious derives from my trust in myself when I am believing conscientiously. My trust in others depends upon my beliefs that they are like me and that I should treat like cases alike. My trust in myself obviously does not depend upon such beliefs.

Another difference is that my mode of access to my conscientiousness differs from my mode of access to the conscientiousness of others. It is possible that my conscientiousness is transparent to my mind, or at least, it is something I can discover through introspection. Of course, self-deception is possible, but in general, I know whether I am using my faculties the best way I can by reflection, and my evidence is probably nonpropositional. In contrast, my reason to believe other persons are conscientious depends upon observation of them and inferences about their inner efforts and abilities from their external behavior. Of course, my judgment of my general abilities relative to other people is also dependent upon observation of others, but there is still a difference between my evidence of the most significant property I trust in myself—my conscientiousness—and my evidence of the conscientiousness of others. The former is close to the best evidence I can have, whereas the latter is weaker.

Notice that this difference is not sufficient to justify epistemic egoism. Even under the assumption that my grounds for trusting myself are better than my grounds for trusting others, my grounds for trusting others are still very good. That is, I will inevitably have excellent evidence of the conscientiousness of others. In fact, I will inevitably have excellent evidence that many others have whatever property I trust in myself, whether it is conscientiousness or natural abilities.

It is possible that I lack good evidence of the conscientiousness of others because I do not go to the trouble of observing them carefully. But I cannot do that if I am conscientious. One of the things conscientiousness demands of me is to pay attention to the people around me since they are potential sources of the satisfaction of my desire for truth, and so conscientiousness demands that I attempt to get evidence of the conscientiousness of others. When I attempt to get evidence relevant to their conscientiousness, I will get plenty of evidence for it. So even if I frequently lack much evidence of their conscientiousness, that does not support epistemic egoism.

A third way in which self-trust and trust in others is asymmetrical is that I must trust myself unless I have reason to think I am untrustworthy on some occasion, but since I must trust myself in order to determine that something is a reason to think I am untrustworthy, trust in myself is a condition for counting anything as such a reason. If it appeared to me that I had reason to think that I am generally untrustworthy, that would threaten to destroy the self.⁸ In contrast, it is possible that my basic trust in some other person is defeated because I have reason to think that the person is generally untrustworthy in his epistemic functioning. But I always have a *prima facie* case for epistemic trust in others, so this difference does not support epistemic egoism.

A fourth difference is that I am stuck with the faculties I have and my degree of experience, background knowledge, intellectual talent, and perceptual acuity. Conscientiousness will lead me to develop my intellectual skills and virtues, but it takes time to develop them, and at any one time, all I can do is to conscientiously use the faculties I have and the skills I have developed up to that point in time. Other persons differ from me and from each other in talent, experience, knowledge, and intellectual virtue, and my trust in them will be determined both by my conscientious judgment of their conscientiousness and by my conscientious judgment of the acuity of their faculties and the intellectual traits they have developed. This difference

8. The movie *Gaslight* (1944) illustrates the problem with thinking that one has evidence that one is generally untrustworthy. The character played by Ingrid Bergman is manipulated by her nefarious husband (Charles Boyer) into believing that she cannot trust her memory or rational faculties. Losing epistemic self-trust leads her to think she is going crazy. For application of the term to actual cases, see Gass and Nichols (1988) and Stern (2007). I thank Juli Eflin for referring me to the use of the term "gaslighting" for the practice of undermining a person's self-trust.

also does not support epistemic egoism. On the contrary, it gives me more than one reason to trust others. When I am conscientious, not only will I believe that many other persons are conscientious, but I will also believe that many other persons have epistemic faculties, skills, and abilities equal to mine or better.

Epistemic egocentrism in the ways I have described in this section follows from the function of the executive self. I exercise my faculties in governing myself, and that includes having the control of an agent over my belief states. I am not maintaining that my belief states are voluntary in the sense in which many overt acts are voluntary. Nonetheless, my ability to reflect upon and to manage my belief states is one of the things I do in managing my self. I have no such management control over the belief states of others. The trust I owe others is an attitude I should have towards their faculties and the deliverances of their faculties if I am consistent, but *I* am the one who manages my trust in others as well as my trust in myself.

3. TRUST IN OTHERS AND THE TWO KINDS OF REASONS

3.1. The Distinction between Deliberative and Theoretical Reasons

I would now like to return to the brief discussion on the nature of reasons at the end of chapter 2. When combined with the argument of this chapter, there is an interesting consequence for the way we ought to treat trust in others as a reason for belief. In chapter 2 I pointed out that there are different senses of reasons or evidence, but every sense requires trust in the connection between what I do when I am using my faculties in the best way I can to reach the truth, and success in reaching the truth.

This situation highlights a distinction in two kinds of reasons that I believe to be important. This is the distinction between what I will call theoretical, or third-person reasons, and what I will call deliberative, or first-person reasons.⁹ What I mean by *theoretical reasons* for believing *p* are

9. In "First Person and Third Person Reasons and Religious Epistemology" (2011b), I describe the different characteristics of these two kinds of reasons and argue that keeping them distinct is helpful in addressing a host of problems in epistemology in general and in religious epistemology in particular.

facts that are logically or probabilistically connected to the truth of p . They are facts (or true propositions) about states of the world or experiences that, taken together, give a cumulative case for or against the fact that p (or the truth of p). They are not intrinsically connected to believing. We call them reasons because a reasonable person who comes to believe them and grasps their logical relations to p will see them as reasons for p . They can be shared with others—laid out on the table, so they are third personal. They are the reasons to which we refer in communicating with others. They are relevant from anyone's point of view. The connections between theoretical reasons and what they are reasons for are among the facts of the universe. Theoretical reasons aggregate and can be used in calculations of probability. What we call evidence is most naturally put in the category of those theoretical reasons we can identify.

In contrast, what I mean by *deliberative reasons* have an essential connection to *me and only to me* in my deliberations about whether p . Deliberative reasons connect *me* to the truth of p , whereas theoretical reasons connect facts about the world with the truth of p . Deliberative reasons do not simply provide me a weightier reason for p than they provide others. They are not reasons for other persons at all. They are irreducibly first personal.

Experience is an example that illustrates the difference between deliberative and theoretical reasons. If you have an experience, the *fact* that you have it is a theoretical reason for believing a variety of propositions. You can tell me about your experience, and if I believe what you tell me, I can then refer to the fact that you had the experience as a reason to believe whatever it supports. You and I can both refer to the fact that you had the experience as a reason to believe something, and so can anybody else who is aware of the fact that you had the experience. The fact that the experience occurred is therefore a theoretical reason. It is on the table for all to consider, and all can consider its logical and probabilistic connections to other facts about the world.

However, you are in a different position than I am with respect to your experience because you not only grasp the fact that you had the experience; in addition, you and you alone *had* the experience. That experience affects many of your reasoning processes, emotional responses, and the way you come to have or give up certain beliefs directly, and that is rational. In contrast, the fact that you had the experience is something you and I and many other people can come to believe. Anybody can form the belief that you had

the experience, thereby accessing that fact, but nobody but you can have your experience. The fact that you had the experience is a theoretical reason for certain beliefs. The experience itself is a deliberative reason for you to form certain beliefs.

Intuitions and emotions can also be deliberative reasons. Intuition in one of its senses is something internal to the mind that responds with an answer to a question, often as a response to a concrete case. My intuition that it is wrong to directly kill an innocent person to save five others gives me a reason to believe that the act is wrong, but you cannot have that reason because you do not have my intuition. But like experience, the fact that I have an intuition can be put out on the table. I can tell you what my intuition is, and you and others can take that into account in drawing conclusions about whether an act is right or wrong. The same point applies to my emotions. I will argue in chapter 4 that emotions can be grounds for beliefs, but like an experience and an intuition, an emotion is a reason only for the person who has the emotion. Emotions are deliberative reasons. But again, the fact that I have a certain emotion can be offered in support of the truth of some proposition, and that is a theoretical reason in support of the proposition.

In chapter 2 I argued that what we call reasons for p would not be indicators of the truth of p unless there was a close connection between what we do when we try our best to reach the truth and success at reaching the truth. We trust that the reasons we identify as reasons for p are in fact indicators of the truth of p because of a more basic trust in ourselves when we are conscientious. What I mean by an epistemic reason to believe p is something on the basis of which it is reasonable to think that p is true. If so, then trust in the self is an epistemic reason because it is in virtue of self-trust that it is reasonable for me to believe anything I believe. Trust is a state in a person on the basis of which she takes her beliefs to be true. This is a deliberative reason. It is a first-person reason because it is epistemically relevant only to the person who has it. Given what I have argued in this chapter, trust in others is also a deliberative reason for belief. Trust in someone else gives me a deliberative reason to believe some proposition p because my reason is based on their similarity to *me* and my trust in myself. This is not a reason that others can use in determining whether p . We have identified, then, several categories of deliberative reasons: experience, intuition, emotion, self-trust, trust in others. In fact, any psychic state in myself is potentially a deliberative epistemic reason for me to believe something, including my own states of believing.

Although the notion of evidence is multiply ambiguous, I have said that I think it is most naturally put in the category of third-person reasons. I argued in chapter 2 that we need trust that what we take to be indicative of truth is in fact indicative of truth, and so the evidence for p we think we can identify is never as basic as trust in the self, and given the argument of this chapter, it is not as basic as trust in others as a reason for believing p . What we call evidence is not only derivative from trust in the self, it is also derivative from trust in others upon whom I rely in identifying the evidence. It follows that trust is a first-person, deliberative reason for belief that is more basic than anything I take to be third-person reasons.

Since no one has figured out how to combine the first-person and third-person perspectives into a single viewpoint, deliberative and theoretical reasons do not aggregate. There is no system of adding together both kinds of reasons for believing p to give a summary verdict on the reasonableness of believing p . Third-person evidence for p does not exhaust all of the reasons for believing p , and in fact, does not even include the most basic kind of reason for believing p .

3.2. The Two Kinds of Reasons and Parity between Self and Others

We are now ready to return to the argument of the extreme epistemic egoist I presented near the beginning of this chapter. When I think about the reasons for believing p , I think about whether p is true, and so I consider reasons relevant to the truth of p . But once I consider and weigh the reasons for and against p and come to a conclusion that p , I do not then treat the fact that I drew that conclusion as another reason in favor of p . My reasons for believing p do not increase once I conclude my process of reasoning and came to believe p . But if that is correct, it should not make any difference to me that somebody else concluded her process of reasoning whether p in a certain way either. If the fact that *I* believe p is not a reason for me to believe p , then the fact that *you* believe p should not be a reason for me to believe p . In both cases, coming to believe something should not count as a reason in favor of the content of what is believed.

The standard epistemic egoist will immediately identify one defect of this argument. It focuses only on first-order theoretical reasons and ignores

second-order theoretical reasons, reasons for thinking that a certain believer is reliable, and I will return to those reasons in chapter 10. But the argument calls attention to something that I think is right: I should treat your believing as on a par with my believing, and this is the case no matter what kind of reasons we are using. If we want to know whether your believing p is relevant to whether I should believe p , we need to be clear about the kind of reasons we are considering. Suppose first that I am thinking only of first-order theoretical reasons. When I look at this class of reasons, I settle the question whether p by identifying and evaluating reasons that can be taken together to support p . In this sense of reasons, the extreme egoist is right that since the fact that I believe p is not a reason to believe p , then the fact that you believe p is not a reason to believe it either.

But when I consider whether p , I might do so by looking at the process by which a belief p is formed. If I know that you believe p and I conscientiously trust the way you acquired the belief p , then the fact that you believe p is relevant to whether I should believe p . Your belief gives me a deliberative reason to think p is true. But in this sense of reasons, the fact that I believe p is also relevant to whether I should (continue to) believe p or to believe p again if I forgot for a while that I used to believe it. Once I reflect upon the fact that I believe p or believed p in the past, then if I trust my past conscientious reflection whether p , I have a reason to believe p that I did not have before I reflected on the way I acquired the belief. My awareness that I believed as the result of a process I trust gives me an additional reason to believe—a deliberative reason.

My reason for believing based on trust in another person is the same kind of reason I have when it is based on trust in myself. Trust in the self and trust in others are deliberative, first-person reasons. These reasons are distinct from the reasons we share with others when we present the case whether p . So my trust in your believing process and my trust in my believing process give me first-person reasons to have the belief the process produces. In contrast, neither the fact that I believe p nor the fact that you believe p gives me or anybody else a first-order, third-person reason supporting the truth of p . Either way, there is parity between self and others, as the extreme egoist noticed.

Both kinds of reasons can be attacked. If I say that talking on a cell phone decreases peripheral vision, you might point to a study that indicates it does not. But if you want to attack my trust in the people who conducted the

study, you will do something quite different. You will attack their trustworthiness. And you can attack my own trustworthiness in a similar way. Empirical studies of mistakes in inductive reasoning attack trust in this sort of way.¹⁰ That is also the strategy of debunking arguments from evolutionary biology.¹¹ The idea of this kind of attack on reasons for p is to undermine trust in the process that produces the belief p , not to undermine the theoretical reasons to which we point to support a case for p . The attack is on first-person, deliberative reasons, not on third-person, theoretical reasons. We will return to the distinction between theoretical and deliberative reasons in chapter 10 when I address the problem of disagreement between myself and another trustworthy person.

The argument of this section leads us to a principle that I will use repeatedly in this book, the *Principle of Epistemic Trust in Others*:

In any case in which, by believing in a way I trust in myself, I am led to believe that others have the same property I trust in myself (to the same degree as I have myself), I have a *prima facie* reason to trust them as much as I trust myself.

The fact that someone has a belief p gives me a *prima facie* reason to believe p . The fact that someone whom I conscientiously believe is conscientious believes p gives me a stronger *prima facie* reason to believe p . If I have *prima facie* reason to trust a person epistemically in some domain or in situations of a certain kind, then I have a *prima facie* reason to believe what they believe in that domain or in situations of that kind. What the Principle of Epistemic Trust in Others shows is that I cannot treat other persons as simply sources of evidence for me like a computer or a clock. If I choose to treat them that way, I must treat myself that way. But I cannot treat myself that way. Reflection on trust in my own faculties forces me to extend my trust to all those persons relevantly like myself. Trust is a first-person reason for belief.

10. Two prominent researchers in this area are Daniel Kahneman and Amos Tversky. For a collection of many of their essays, as well as others discussing the various mistakes of reasoning people commonly make, see Kahneman, Tversky, and Slovic (1982).

11. A defense and exploration of this view can be found in Pinker (1997). For an example of how evolutionary psychology can be thought to debunk some specific phenomena, see Bulbulia (2004). I will discuss debunking arguments in chapter 11.

4. EPISTEMIC UNIVERSALISM AND COMMON CONSENT ARGUMENTS

Let us take a closer look at epistemic universalism. As I have argued, the general trust I have in myself commits me to a presumption in favor of the belief of any other person. This is a modest universalism because in virtually every real-life case of finding out in a way I trust that somebody has a certain belief, I also get information about the believer that affects the credibility of the belief. Furthermore, the content of the belief is often something to which I assign a high or low prior probability, given other things I believe, or other psychic states I trust, such as emotions or attitudes, so the kind of universalism I have defended may have little if any practical effect in many cases. Nonetheless the *prima facie* credibility of the belief of another person is important because not only can it affect the balance of total reasons, it affects my epistemic attitude towards other people even when it does not lead me to change my beliefs. Many things change when the default position is trust rather than distrust.

To illustrate the universalist thesis, let me try to construct a pure case, one in which there is no information about the source that is a potential defeater of the belief, and where the content of the belief is something about which I am neutral at the outset.

Suppose I find out that somebody somewhere of whom I know nothing believes that Poland was invaded by the Tartars in 1241, and suppose that my background knowledge is so vague that I have no reason to believe or disbelieve it in advance, and suppose that neither believing nor disbelieving that the Tartars invaded Poland would have any effect on anything else I trust. According to the principle of epistemic universalism, I have a *prima facie* reason to believe that Poland was invaded by the Tartars in 1241. This seems to me to be intuitively correct. Of course, even this case is not a pure case. I cannot find out that someone believes the Tartars invaded Poland that year without finding out many other things about the believer of the proposition, and I do have some background knowledge about medieval Poland and the Tartars, however limited. So in practice my defeasible reason for believing may be defeated easily, either because I have evidence that the believer is untrustworthy, or because I already have a belief I trust that conflicts with the belief about the Tartars. Nonetheless, it is worth noting that once I become aware of the fact that somebody believes that Poland was

invaded by the Tartars in 1241, I have a *prima facie* reason to believe it. The reason is not decisive; it is insufficient by itself to make it reasonable for me to acquire the belief. But the belief of the other person counts in favor of it. It is a mark in favor of its credibility.

Suppose now that I find out that a large number of people have the same belief. Maybe I learn that thousands of people believe that the Tartars invaded Poland. They tell narratives about it, teach it in their schools, and mark the event with ceremonies.¹² Again, we need to imagine that I lack any other information about these people or the content of their belief, and that is even harder to do when there are many believers. Nonetheless, it seems to me that if there is a large number of believers, that strengthens my defeasible reason for sharing their belief. The degree of trust I should have in their belief increases. Of course, it might turn out that most of these people acquired their belief from one or a few other people, and probably most beliefs that are shared by large numbers of people are spread through the community by testimony. So we should say that other things equal, a belief independently acquired by a great many people is more trustworthy than the belief of many people that can be traced back to the belief of only a few.¹³

However, even if many people acquire a belief from a single person, it would be a mistake to think that the trustworthiness of the belief is no greater than it would be if only the one person believed it. Suppose twenty people believe the outcome of one person's addition on the testimony of the person who added up the figures. Their belief does not count as much as it would if each of them had independently made the calculation, but it counts more than the case in which only the one who added the figures believes it. That is because the fact that twenty people have the same belief is evidence that nineteen people regard the source from which they acquired the belief as trustworthy, and that gives me a *prima facie* reason to treat the source as

12. This is not too far from the truth because the event is commemorated every fifteen minutes in Kraków when a bugler plays a plaintive tune from alternating directions at the top of the tower of St. Mary's Church in the Market Square, and then abruptly stops, in remembrance of the bugler who was allegedly shot through the throat in midtone while he was attempting to warn the people of the attack. However, they are not simultaneously reminded of the date. Of course, the legend is probably too good to be quite accurate.

13. Kelly (2010, 148) expresses the most common position when he says that consensus means little in the absence of independence.

trustworthy also. I have a defeasible reason to trust persons whom persons I trust, trust. If many thousands of people share a belief, that gives me a greater reason to trust the belief than if only one or a few persons believe it, even if the beliefs were not acquired independently.

But I am willing to leave this point aside, and look instead at common consent when there is independence. There is common consent that common consent gives us a reason for belief, and it is not only the philosophically uninitiated who treat it as a reason. Philosophers known for their carefulness in identifying and defending their grounds for making an assertion sometimes refer to common agreement without defense. Hume does this in his essay "Of Miracles." After giving an argument against the reasonableness of believing on testimony that a miracle occurred, he gives an example of something it would be reasonable to believe that has some of the features of a miracle. He imagines that we have the testimony of a multitude of persons that darkness fell over the earth for a period of eight days on a date long in the past. Hume thinks that there are features of the situation that increase the reasonableness of believing their testimony, in fact, making it reasonable to adopt the belief. The most important is widespread agreement among independent testifiers (*Enquiry*, Sec. X, Pt. II).

Many other philosophers have referred to common consent as a mark in favor of the truth of a belief. Aristotle's methodology in the *Nicomachean Ethics*, the *Politics*, and the *Metaphysics* is to begin with what everyone accepts, an approach that has the advantage of making what follows seem to be a development of what is already accepted by his audience, while putting objectors on the defensive. Many other philosophers give substantial weight to the closely related notion of common sense, which differs from common agreement primarily in that a writer who calls a belief common sense typically shares the belief, whereas a writer who refers to common agreement may not. Thomas Reid is undoubtedly the most influential philosophical supporter of common sense, and he often treats the denial of common sense with scorn,¹⁴ but I find it revealing that John Stuart Mill defends the *prima*

14. Here is a choice line from Reid: "To what purpose is it for philosophy to decide against common sense in this or any other matter? The belief of a material world is older, and of more authority, than any principles of philosophy. It declines the tribunal of reason, and laughs at all the artillery of the logician" (Reid 1997: 127).

facie credibility of a belief held by a large number of people even when it conflicts with one's own:

For, whatever opinion a person may adopt on any subject that admits of controversy, his assurance if he be a cautious thinker cannot be complete unless he is able to account for the existence of the opposite opinion. To ascribe it to the weakness of the human understanding is an explanation which cannot be sufficient for such a thinker, for he will be slow to assume that he has himself a less share of that infirmity than the rest of mankind and that error is more likely to be on the other side than on his own. In his examination of evidence, the persuasion of others, perhaps of mankind in general, is one of the data of the case—one of the phenomena to be accounted for. As the human intellect though weak is not essentially perverted, there is a certain presumption of the truth of any opinion held by many human minds, requiring to be rebutted by assigning some other real or possible cause for its prevalence. (Mill 1963, 430).

It is interesting that in spite of the long history of common consent arguments, contemporary books of informal logic typically treat it as fallacious, giving it the name *ad populum* fallacy. Copi's classic logic book offers as an illustration of the fallacy the use of the fact that the Golden Rule is a basic moral principle in many systems of ethics in widely differing cultures as a reason to believe the Golden Rule.¹⁵ On the contrary, it seems to me that that is a fairly strong *prima facie* reason to believe it, requiring a special reason to defeat it, as Mill maintained.

Part of the problem in debating the alleged fallacy of common consent arguments is that they do not all have the same form. The *ad populum* fallacy is usually described as the argument, "Everybody believes it, so it must be true," and it is then quite rightly pointed out that what all or most people believe is not necessarily the case, usually accompanied by counterexamples ("The earth is flat," "Heavy objects fall faster than light ones," etc). This observation is hardly sufficient to show that common consent has no rational weight, and it is not very illuminating. The argument endorsed by Mill in the above passage is an argument to the best explanation. Paul Edwards (1967) distinguishes two traditional variations of this form of argument. In one ver-

15. Copi (1986: 50). An interesting, although not uncontroversial, account of similar moral codes in a variety of religious traditions can be found in the "Appendix: Illustrations of the Tao" from Lewis (1974).

sion, the universality of a belief is taken to be evidence that it is instinctive, and the best explanation of that is that it is true. He calls this the “biological version” (148–50). In the second version, it is argued that either the best explanation of the universality of a belief is its truth or we must be skeptical of human intellectual faculties. Edwards calls this “the anti-skeptical dilemma” (150–51). I hope it is clear that my argument has neither form for two reasons. First, it proceeds from the commonality of a belief, not its universality, but more importantly, it is not an argument to the best explanation. My argument does not begin with a datum that needs to be explained—the fact that people commonly believe p , where various explanations of the datum are weighed against each other from the standpoint of a neutral arbiter. In that form of argument, no connection is assumed between the arbiter’s trust in the way she goes about evaluating alternative explanations and her trust in the beliefs of the persons evaluated. In contrast, the argument I have proposed links trust in the beliefs of others with self-trust. The rationality of trust in others is derivative from the rationality of trust in the self. I think that for that reason it is a stronger argument than the argument to the best explanation.

The argument of this chapter is both good news and bad news for the reasonableness of our own beliefs. Agreement with those we trust supports our beliefs, but it also makes disagreement with those we trust a problem. Rational self-trust forces us to trust many of those with whom we disagree. I have argued that we cannot ignore them without succumbing to ethical egoism. Disagreement with people we trust is a problem that arises in its deepest form within the aspects of the self I trust. On the one hand, I trust a certain belief I have. On the other hand, I trust my conscientiously formed belief that someone else’s conflicting belief is as conscientious as mine. If I am committed to a similar attitude towards his belief as towards my own, that forces me to engage in greater reflection on the structure of the self and the relative trustworthiness of the states of self I am able to manage. We will return to the issue of resolving conflicts within self-trust in chapter 10.

In chapter 1 I argued that the modern rejection of epistemic authority derives from two modern values, either of which makes epistemic authority problematic, but which are in tension with each other. These values are intellectual egalitarianism and a notion of intellectual autonomy that equates it with epistemic self-reliance. In the last two chapters I have argued that a

consistent self-reflective person is rationally required to have both epistemic self-trust and epistemic trust in others. The kind of egalitarianism I have defended is weak, but it is strong enough to make epistemic self-reliance, interpreted as either extreme or standard epistemic egoism, incoherent. I believe that self-reliance in some of the nonepistemic aspects of the self is incoherent also. I will turn to that next.

Trust in Emotions

1. THE RATIONAL INESCAPABILITY OF EMOTIONAL SELF-TRUST

The argument for the need for trust in our epistemic faculties applies to other parts of the self. In this chapter I will argue that we need trust in our emotion dispositions, and I will then focus on trust in the emotion of admiration, an emotion that gives us another ground for epistemic trust in others. Epistemic and emotional self-trust lead to a network of trust in communities and traditions that aid us in interpreting our experience and evaluating what we trust in ourselves, but I will not say much about communities until chapter 7. The purpose of this chapter is to show that the task of the conscientious person extends well beyond the attempt to make her beliefs fit their objects.

So far I have been focusing on trust in the relation between belief and truth, but that relation is an instance of a more general kind of relation. The issue of trust arises whenever we are in a state that is directed towards an object in such a way that it can be fitting or unfitting, right or wrong. We have many such states in addition to beliefs. Desires form one category. We desire objects that may or may not be desirable. To desire the undesirable is to make a certain kind of mistake. We aim to desire the desirable, and when we desire something that is not desirable, we have failed. The nature of the failure is not important for my point. Something might be undesirable *simpliciter*, or it might only be undesirable as it would be judged by the agent herself—an ideal or future self, and of course, there are other possibilities. But as long as (1) desires aim at the desirable, (2) there is a distinction between the desirable and the undesirable, and (3) the fact that I desire something is not sufficient to make it desirable, it follows that desires can succeed or fail just as

beliefs can succeed or fail. The possibility of failure raises the issue of whether we have reasons for thinking that what we desire is desirable, whether our reasons are circular, and whether they bottom out in trust in a disposition.

The same issue arises for the more complex and interesting category of emotions. I will not try to adjudicate between competing accounts of emotion here, but I want to show why it is appropriate to raise the question of trusting or doubting an emotion, given certain assumptions about the nature of emotion. First, I will assume that emotions have intentional objects and that having an intentional object is roughly what distinguishes an emotion from a sensation or a mood. We fear something, hope for something, pity someone, love someone, feel indignation at some state of affairs, feel sympathy with someone's plight, feel angry at someone, and so on. When I reflect, I can reasonably ask whether it is fitting or appropriate that anger is directed at the object of my anger, whether what I fear is really fearsome, whether the object of my indignation deserves that emotion, and so on. Some emotions may be pure reactions, and we do not think that the issue of fit arises. For instance, feeling irritated at something is more like finding something nauseating than feeling angry at it. I need not ask myself whether it is appropriate for me to feel irritated at Microsoft Word, even while I know that most people do not find it irritating. There are also emotions that can be appropriate or inappropriate, but where we think the determination of fit is up to me. For instance, I determine whom to love and what to hope for. Even so, I can make a mistake. I imagine that the range of fitting objects of love or hope varies more from person to person than the range of fitting objects of anger or fear, but there are still objects beyond the range. I may later judge that I should not have loved my roses so much, or that it was unwise to hope for a windfall profit on the stock market, and it seems to me that my later judgment can be correct. It survives critical self-reflection better than my earlier emotion. So with some qualifications, emotions can fit or not fit their objects, and we think that an emotion ought to fit its object.

Given that emotions may or may not appropriately fit the object, we can raise the same sort of questions about emotions as about beliefs and desires. That is, we can reasonably ask ourselves whether we have a noncircular way to tell that our emotion dispositions in conjunction with our other faculties reliably produce a state that fits its object. Similarly, we can ask whether we have noncircular reasons for a particular emotion, just as we can ask whether we have noncircular reasons for a particular belief.

One other feature of emotion that I assume for this argument is that an emotion has a cognitive component. I think that it also has an affective component, but I will focus on the former. In a state of emotion, something *appears* to the agent to be a certain way, a way that is distinctive of the emotion type and that is not purely descriptive. In a state of fear, the object of fear appears fearsome; in a state of pity, the object of pity appears pitiable; in a state of love, the object of love appears lovable; and so on.¹ I am not suggesting that an emotion includes a judgment or belief. Something can appear fearsome to the agent when she does not judge that it is fearsome. In fact, she may judge that it is not fearsome. But the category of appearances raises the question of fit. If the light appears to me to be green when it is red, there is a misalignment between the world and my faculties even if I do not judge that the light is green. There is a misstep of some kind—a lack of fit between my faculties and my environment. The faculty or disposition through which something appears to me is misrepresenting the object.² Similarly, if someone appears pitiful when she is not, there is a misalignment between the world and the emotion disposition operating in that situation even if I do not judge that she is pitiful. For philosophers who take the position that emotions are judgments or have judgments as components, the conclusion that we can be mistaken in our emotions is straightforward.³ My point here is that there can be error without judgment, and we need not think that emotions include judgments to think that an emotion can be in error. So one way that the possibility of error arises for emotions is that emotions involve appearances and an appearance can misrepresent the object.

We sometimes have a noncircular way to tell that an emotion is unfitting. That is due to another feature of emotion: the appearances that are components of emotion depend upon particular descriptive features. For instance, fearsomeness depends upon the fact that the object can harm me. Pitifulness

1. In Zagzebski (2003) and (2004a) I call concepts that we apply to the objects of emotion—pitiable, fearsome, contemptible, rude, etc.—“thick affective concepts.” I think these concepts are not purely descriptive, but are affectively laden. They are not understandable apart from the disposition to feel emotions of the relevant kind.

2. For a general account of appearing in perception, see Alston (1999).

3. See, for example, Nussbaum (2001) and Solomon (1980) and (1984). For a detailed account of emotions that interprets them as having cognitive content without judgment, see Roberts (2003).

depends upon the fact that the object is suffering. If the descriptive feature does not apply, the emotion is unfitting, and the subject can find out that it is unfitting by finding out that the descriptive feature does not apply.

Consider the case of feeling offended at something that appears rude. Philippa Foot argued many years ago that "rude" is appropriately applied to behavior only when it expresses lack of respect (1978: 96–109). I am not sure that that is precisely right, but I agree that a necessary condition for something's being rude is that it either expresses a certain attitude or is behavior of a type that usually expresses that attitude—disrespect or something similar. In any case, "rude" applies to only a certain range of behavior.⁴ Foot argues that a vast area of descriptive situations are ruled out of the range of behavior that can be rude—for instance, a man walking slowly up to the front door.

We could show that feeling offended at such behavior is unjustified, that is, does not fit the circumstances, by saying something like, "All he was doing was walking up to the door. He meant no offense, did not have a disrespectful attitude, and was violating no social norms dealing with the manner in which one should walk up to the door." If someone takes offense at the man walking slowly up to the front door, we could convince her that her emotion is unfitting in this way. If she believes us, she will conclude that the man's behavior is not rude and, typically, her feeling of offense will disappear effortlessly. If it does not, there is dissonance between her emotion and her beliefs, and if she trusts the beliefs more than the emotion, she will judge that her emotion is unfitting. There are other possibilities, such as the interesting case in which she trusts the emotion more than the beliefs, but my point here is that an emotion fits the circumstances only when certain descriptive facts obtain. If the agent's beliefs about those facts are false, the emotion does not fit, and this is something it is possible for an agent to see herself. So we can sometimes see that an emotion is unfitting or inappropriate in a noncircular way, that is, a way that does not refer to emotion dispositions.

Having an accurate grasp of the descriptive features of a situation is necessary for the right emotion, but I want to argue now that it is not sufficient. The move from the output of our perceptual and epistemic faculties to an

4. Stocker and Hegeman (1996: 74–77) offer objections to Foot's claim about the relation between rudeness and giving offense.

emotion state requires the operation of an emotion faculty or disposition. We have evidence that there is a distinct faculty or disposition in the production of emotion because of the cases of individuals who have all the right beliefs about the descriptive facts but lack the ability to feel emotions because of damage to sections of the prefrontal cortex region of the brain.⁵ Perceptual and cognitive processes are not sufficient to produce an emotion, much less an emotion of any particular type. A person without an emotion disposition will not have an emotion no matter what her descriptive beliefs, and absent a particular emotion disposition, there are indefinitely many affective states one could have that are compatible with any set of descriptive beliefs.

Imagine beings just like us in perceptual and cognitive abilities, but with different emotion dispositions. For instance, suppose that there are beings otherwise like us, but who reverse the human emotions of contempt and offense. In a situation in which a human being typically feels offended, these beings feel contempt, and in a situation in which a human being typically feels contempt, they feel offended. We could not convince the emotion-reversers that it is not appropriate to feel offended when someone is doing what we see as contemptible, nor could they convince us of the contrary, because it takes an emotion disposition to move from the belief that a situation has certain descriptive features to an emotion such as offense or contempt, and we differ in those dispositions.

In chapter 2 I compared our system of epistemic faculties to an air-conditioning system that can read itself, and I pointed out that the system would have some ways to tell that it is unreliable, but it could not tell in a noncircular way that it is reliable. The same analogy applies to our system of emotion dispositions. We would know that something has gone wrong with the system that produces our emotions if, for instance, the system is "set" to produce the feeling of offense only when we believe we have been treated with disrespect, but we often feel offended in other circumstances. But if it regularly produces the feeling it is set to produce and does not produce that feeling in other circumstances, we have no way to tell that the setting itself is appropriate, just as the air-conditioning system that reads itself cannot tell

5. For example, see Damasio (1994). For other examples of how damage to the prefrontal cortex affects the emotional dispositions of persons, see Berlin, Kischka, and Rolls (2004: 1108–26), and Driscoll (2009).

that as long as it goes on when it reads 73 degrees and goes off when it reads 72 degrees, that it is keeping the room temperature at 72 degrees. For that reason, we must trust the connection between the emotion disposition and the fittingness of what it produces to its object.⁶

Suppose, then, that I have settled for myself the relevant descriptive facts in some situation. How can I tell that my emotion is fitting? To put the same question another way, suppose that two people agree on all the descriptive facts but differ in emotion. You and I agree that someone has expressed disrespect for us, but I see it as rude and you do not. How could we settle the disagreement? My point is not that emotion disagreements need to be settled. Usually they do not. I am raising the issue of settling disagreement as a way to show how an emotion is justified. I do not see how it can be done without reference to other instances of the same emotion type in relevantly similar circumstances. So if I feel offended and you do not, I would point out that the situation is like other situations in which you felt offended or thought the feeling of offense was appropriate. If you are not disposed to feel offended in circumstances of the relevant kind, it does no good to point out to you that it is of the relevant kind.

The justification of fear is also circular. Suppose you and I both observe the presence of something that can harm us. We agree on all the relevant descriptive features of the situation—that the object or animal is harmful, the degree of harm it can inflict upon us, and the probability that it will in fact harm us. But I feel fear and you do not. Once we settle on that and become aware of the differences in our emotional response, at least one of us will be puzzled. Certainly I will be puzzled. I might actually admire you because I might believe that fear, while appropriate in the circumstances, sometimes prevents a person from making the most rational response to a dangerous situation. You might be better able to figure out a way for us to escape than I am in the grip of my fear. But that does not make your emotion more fitting than mine in the sense I mean. Another possibility is that

6. I need to make a qualification on this point. When I gave the argument for epistemic circularity in chapter 2, I distinguished Alston's claim that there is no noncircular justification for each of our basic epistemic faculties—perception, memory, etc.—from Foley's more cautious claim that there is no noncircular way to justify our epistemic faculties taken as a whole. Similarly, it might turn out that it is possible to have noncircular justification for a particular emotion type—regret, pity, compassion, etc., even though there is no noncircular justification for our emotion dispositions taken as a whole.

I would admire you if I believed that your ability to avoid fear is due to discipline, that in the past you felt fear in such situations, but you learned through careful training and practice not to be afraid. That also does not show that the emotion of fear is inappropriate. It shows that there can be reasons not to have an appropriate emotion. But if your lack of fear is due to a lack of the disposition to feel fear, I would think there is something wrong with you. I do not see any way to justify the judgment that fear is appropriate and lack of fear is not without reference to other instances in which the disposition to fear operates. If you lack the disposition or if your disposition is radically different from mine, I would be at a loss to justify the appropriateness of any particular output of my own disposition. My judgment that fear is appropriate in this case is correct only under the assumption that my disposition to fear usually produces an emotion appropriate to the circumstances.

The appropriateness of the emotion of disgust has received some attention in the recent literature. The research of Jonathan Haidt and colleagues (1993) shows that people who refer to disgust as a reason not to do something cannot justify their judgment any further than the experience of the emotion of disgust itself. Haidt's examples include eating one's pet dog that was killed by a car, cutting up an American flag and using it for rags to clean the bathroom, and having sex with a dead chicken. As far as I know, Haidt does not take a stand on the appropriateness of disgust in these cases. The focus of most of the discussion of his research has been on whether finding an act disgusting is a good reason to say you should not do it.⁷

But I am not discussing here the issue of whether an emotion is a reason to act in certain ways. My topic is the appropriateness of the emotion to the circumstances. Martha Nussbaum (2006) has given an extended argument that disgust should be irrelevant to the law, and she clearly thinks that disgust is often an inappropriate response to a situation, quite apart from its legal ramifications. She argues that disgust can be attributed to a fear of our

7. Kwame Anthony Appiah (2008) says in reference to the Haidt research, "If 'You can't have sex with that, it's a dead chicken' is (as I rather suspect) a bad reason, we'll want to be able to say why" (141). Appiah's discussion is about disgust as a moral sentiment. Given the context, I cannot tell if Appiah means that he suspects that the fact that something is a chicken does not give one a good *moral* reason not to have sex with it, or if he thinks it is not a good reason *simpliciter*. Steven Pinker argues that "taboos" such as those investigated by the Haidt group lack justification. Taboos include, but are not limited to, acts that people who adhere to the taboo find disgusting. But see MacIntyre (1990: 177ff.) for a subtle discussion of the role of taboo in a system of other concepts, including religious ones.

humanity; in particular, a rejection of our bodily nature. She argues further that disgust expresses “magical ideas of contamination” (14), and aspirations to purity that are not realistic for human life as we know it. It appears to me that Nussbaum’s position is not only that one should not use disgust as a ground for a legal judgment; she thinks that disgust is often a mistake. What appears disgusting is not disgusting. The emotion misrepresents the object. But how could we settle a disagreement between someone who finds sex with a dead chicken disgusting and someone who does not? They might agree on all the descriptive features of the act and its consequences, including the fact that the act is harmless. Each disputant might compare her emotion of disgust with previous situations in which she felt disgust, as well as similar situations in which she did not. Once she has done that, there is not much that can be done to resolve the disagreement except by reference to cases in which both parties take the feeling of disgust to be appropriate. That is, the two parties together can only do what each of them would do to settle for herself whether the emotion is appropriate. They compare one instance of the emotion with another. Sometimes one or the other will find anomalies in her emotion, and that will help them settle the disagreement. But if they simply have different emotion dispositions, I do not see how it is possible to adjudicate differences in the dispositions.⁸ What happens, I think, is that there are people who find that their emotion of disgust changes over time and with reflection, and possibly disappears entirely, whereas the emotions of other persons do not change, even with similar experiences and similar reflections. But we cannot expect anyone to change an emotion because *we* have changed it in similar circumstances. We need to admit that once we agree with each other on the descriptive features of a situation, then if there is a difference of emotional response, neither one of us can reasonably judge that we are right and the other is wrong without the assumption that our own emotion dispositions are more trustworthy than the dispositions of the other person. It is hard to see how we can produce such an argument if the emotions of both persons survive their own critical reflection on their total set of beliefs and emotions without dissonance.

Emotion circularity is not the deepest reason we need trust in our emotion dispositions. In chapter 2 I argued that even if we could complete the

8. Haidt et al. (2008, 1107) present one way a person could be led to see disgust as inappropriate, and then argue that if people ignore all feelings in making judgments, they may have little else to go on.

search for reasons for our beliefs in a noncircular way, we would still need to trust that there is any connection at all between what we do when we try to get the truth and success in reaching the truth. Similarly, the need for trust in our emotion dispositions is not primarily due to the fact that we cannot complete the task of justifying an emotion in a noncircular way. The problem is that we have no way of telling that there is any connection at all between what we do to justify the fittingness of an emotion and its actual fittingness. To the extent that we rely upon our emotions in the conduct of our lives and believe that the emotions upon which we rely are fitting, we need basic trust in the tendency of our emotion dispositions to produce fitting emotions for the same reason we need basic trust in the tendency of our epistemic faculties to produce true beliefs.

Someone might accept my assumption that emotions have cognitive content, but reject my argument that they can fit or not fit their objects on the grounds that what makes an emotion appropriate is not that it fits its object, but just that it leads to action that is right for circumstances with certain descriptive features. I have heard the claim that as long as a person acts appropriately, any emotion from which the act arises is appropriate, as well as no emotion at all. On this view, we gain nothing in explanatory power by saying that the cognitive content of emotions can be fitting or unfitting, nor is it accurate to call the feeling aspect of an emotion appropriate except in the sense just noted. If a person responds to a suffering person by helping her, whatever emotion leads her to do so is appropriate. If she runs from danger, whatever emotion leads her to do so is fitting just because it leads her to run. Otherwise, it makes no difference that she feels one thing rather than another. Inner states mean nothing.

An adequate answer to this objection would require a full defense of my view of the nature of emotions, but I want to offer one observation about it here. As long as an emotion is a response to the world and it has cognitive content, it is hard to see how someone can think an emotion state is not the sort of state that can fit or fail to fit its object without having the same view about beliefs. If the appropriateness of an emotion is determined by its power to cause a person to act in a way that is right for the situation, the appropriateness of a belief ought to be determined the same way. According to this position, when I see a suffering person, it ought not to matter whether I believe that I see a suffering person or whether I believe that $2 + 2 = 4$ or that the sky is blue, as long as I help the person. What matters is that the

belief state leads to the right act. When I am in danger, it ought not to matter whether I believe I am in danger or instead believe that Benjamin Franklin invented the lightning conductor, as long as I run away. But, of course, we do think that a belief can fit or fail to fit its object independently of the acts to which it leads, and we think that because we think a belief state is a state of taking the world to be a certain way. My position is that an emotion is also, in part, a state of taking the world to be a certain way. It differs from beliefs in several respects, but as long as it includes an aspect of appearance, its appropriateness cannot be simply a matter of leading to appropriate external behavior. We must ask ourselves, then, how we can tell that an emotion is appropriate. When we do so, we encounter the same problem that we encounter with our other faculties. Upon reflection we find that we need basic trust in our emotion dispositions, just as we need basic trust in our sensory and epistemic faculties.

2. TRUSTWORTHY AND UNTRUSTWORTHY EMOTIONS

There are important differences between our emotions and some of our other faculties that lead some people to become emotion skeptics. Our reasons for trusting our emotions are circular, but some of our emotions lack circular justification. For the most part, our sensory faculties produce consistent outputs, but many of our emotion dispositions do not, and we later judge that the emotion was a mistake. An emotion may change when our beliefs about the descriptive facts do not change, or we judge the emotion to be inappropriate after reflection. We have experiences of falling in and out of love with the same person, becoming angry and then losing anger at the same incident, feeling contempt that turns to pity, feeling disgust that changes into compassion, and so on in a multitude of cases that illustrate the instability and lack of consistency of many of our emotions. These experiences happen often enough that most of us are much less trusting of our emotion dispositions than of our epistemic faculties.

An emotion skeptic will say that these considerations show that we lack even circular justification for our emotions. She will say that our emotions are unstable, do not survive critical self-reflection, and therefore should not be trusted. But general emotion skepticism is disingenuous since there are

many emotions that we all trust, and many emotions that we need to trust in order to live a normal life.

One is the emotion of sympathy. David Hume was surely one of the most rigorously untrusting of the great philosophers, but he trusted sympathy, and he trusted it enough to make it the centerpiece of his moral philosophy.⁹ But how could we justify the appropriateness of sympathy? Perhaps there is an argument that consistency requires us to feel about others in a way that is (very roughly) similar to the way we feel about ourselves, but it is doubtful such an argument would get very far. At least, it is doubtful that it will make headway against anybody who does not already have a natural inclination to sympathy. We could not prove that the particular emotion we call sympathy is appropriately directed at *anything*, much less that it is appropriately directed towards particular other persons. It seems to me that the most honest thing to say is simply that we trust it.

There are many other emotions that we trust. In my experience, even philosophers with the greatest proclivity to moral skepticism trust their emotion of indignation. Unlike sympathy, indignation is an emotion that quite obviously is not always appropriate because it is dependent upon or sensitive to moral judgments, usually judgments of injustice. But curiously, people will often trust their emotion of indignation more than the associated moral judgment, and sometimes use the emotion as grounds for the judgment.¹⁰ Whether this is justified is an intricate question, but I use it only as evidence that there are emotions we trust without noncircular defense.

Most of us have considerable trust in some epistemic emotions as well—emotions directed at beliefs. One is the emotion that Thomas Reid calls “epistemic ridicule.”¹¹ We think, “That’s ridiculous!” or “How absurd!” when we hear something we take to be epistemically outrageous. Reid says we have this emotion when faced with someone who denies a First Principle of common sense, and he clearly thinks it is appropriate in such cases.¹² What is

9. See *A Treatise of Human Nature*, esp. Vol. II, Pt. I, sec. XI; Vol. III, Pt. II, sec. II.

10. The same point applies to outrage, which I interpret as a stronger form of indignation. I find it puzzling that some people trust outrage in themselves and others simply because it is a moral emotion.

11. In my vocabulary, “ridicule” is the name of a form of behavior, not an emotion. But since it is hard to know what to call the emotion that accompanies that behavior, I am following Reid in calling it ridicule.

12. Thomas Reid, “Essay Six: Of Judgment,” (1983: 256).

particularly interesting about Reid's point for my purpose here is that he does not use the fact that someone violates common sense as a way to justify the emotion of ridicule. He argues the other way around. Trust in the emotion of ridicule grounds his claim that we have reason to think that the object of ridicule is epistemically unjustified. Again, my point is not that there is no way we can be mistaken in this emotion, but that we place a great deal of trust in it, and there is no justification for it that does not assume the general trustworthiness of the disposition that produces the emotion.

It appears, then, that we trust some emotions even though we do not trust others. If trust in an emotion survives conscientious reflection, we have the same type of grounds for trusting it as we have for trusting beliefs. Perhaps we think that our past emotions have a poorer record of survival of conscientious reflection than our past beliefs. If so, we have reason to think that our emotion dispositions in general are less trustworthy than our epistemic dispositions, and so we ought to be cautious, but I do not see that we have reason for general emotion skepticism.

I suggest that a conscientious person should treat emotion dispositions the same way she treats her epistemic faculties. She does not have a noncircular justification for the reliability of either kind of faculty, but the outputs of both kinds of faculty can survive conscientious self-reflection. In both cases she needs to rely upon the faculty to lead a normal life, and in both cases, she cannot do that without placing basic trust in the faculty or disposition itself. She monitors the faculty, adjusting or attempting to adjust its outputs in response to reflective judgment. What I have called epistemic conscientiousness is the self-conscious attempt to make our beliefs fit the truth. I want to suggest now that there is a form of conscientiousness in which we make a self-conscious attempt to make our emotions fit their objects. Like epistemic conscientiousness, emotional conscientiousness would not be trustworthy unless our dispositions were already basically trustworthy. There is no point in trusting ourselves when we use a faculty to the best of our ability unless the faculty is generally suited for success. The trustworthiness of emotional conscientiousness depends upon the general trustworthiness of our emotion dispositions, and trust in ourselves when we are emotionally conscientious depends upon a more basic trust in our emotion dispositions in conjunction with our other faculties.

Could an epistemically conscientious person opt out of trusting any of her emotions? I have not said how many emotions survive conscientious self-

reflection for any particular person. I think it is highly unlikely, perhaps impossible, that none of them does. The consequences of total lack of trust in emotions would be drastic. I have argued in another place that moral judgments depend upon emotions, so if I am right about that, skepticism about emotions would lead to skepticism about moral judgments.¹³ I have not defended that position here, so my claim that radical emotion skepticism is not viable does not depend upon it, but I am relying upon the intuitive force of examples of emotions that are generally trusted by conscientious persons, and which are furthermore used as partial justifications for action. If a conscientious person trusts her decision to act in a certain way, she needs to trust the emotions that form at least a partial basis for the decision. In any case, I think it is fair to say that we often act by depending upon the appropriateness of certain emotions. If the emotion is not appropriate, that needs to be established by the failure of the emotion to satisfy the demands of conscientious self-reflection. I do not see that any general skeptical considerations are apt to succeed.

3. ADMIRATION AND TRUST IN EXEMPLARS

I want now to focus on a particular emotion that I think we trust as much as we trust sympathy, indignation and the sense of the ridiculous. That is the emotion of admiration. I am focusing on this emotion because it is one of the emotions trust in which commits us to epistemic and nonepistemic trust in other persons. I think that admiration plays a crucial role in learning moral and intellectual virtues, and it is an important element in the justification of authority.

Let me begin my discussion of admiration with a story. On January 17, 1912, Sir Robert Falcon Scott reached the South Pole. There followed a harrowing attempt to get out of the Antarctic peninsula, and he was the last of his party to die, leaving farewell letters under his body that were discovered months later. He and the members of his party encountered horrific conditions during their attempt to walk out. The weather was extreme, with temperatures among the coldest on record. Frostbite was so serious it took them several hours to put on their boots in the morning. Day and night they

13. Zagzebski (2003a) and (2004a).

battled hunger, dysentery, bleeding gums, exhaustion, mental confusion, and despair.¹⁴ In one letter, Scott writes that a member of their party, Captain Oates, sacrificed himself because his ruined feet were slowing down the march. According to Scott, Oates stepped outside the tent in a blizzard one night, saying to the others, "I am just going outside and may be some time" (22).

To me it is obvious that what Captain Oates did was admirable. I admire it and I trust that emotion. Even if I became convinced that Oates committed an immoral act of suicide, I would still trust my admiration for him; I would then think that the moral and the admirable can come apart. Certainly, the admirable can conflict with what we judge to be the best thing to do, all things considered. Scott provides an example of that also. He refused to bring dogs on the expedition, and declared that eating dogs was inhumane. In one letter he says that the goal was not just to reach the Pole, but to reach it in a certain way. When it was reached unaided, he wrote, "the conquest is more nobly and splendidly won." I find that admirable also, even though it seems to me it would have been prudent to take sled dogs along.¹⁵

We admire qualities as well as individual acts. We find courage and compassion and generosity admirable, and we feel the opposite towards arrogance, venality, and cowardice.¹⁶ Perhaps it is harder to trust our admiration for qualities than for acts, but I think that is only because it is hard to be confident we can identify qualities. Can we be sure that Captain Oates was a courageous person or that Scott had a sense of dignity? Perhaps not, but I think we are still confident that we admire courage and the sense of dignity, whether or not we know of anybody who has either in the fullest sense that we admire.

We also admire persons. Many individuals are admirable for one or two qualities of intellect or character, but there are also persons who are admirable *as persons*. It is difficult to say what it takes to be admirable as a person. Does she have to be admirable in every respect? It is very unlikely that there is anyone like that, but there are persons who are admirable in the important

14. Related by Annie Dillard (1982: 29). One current account of Scott's journey is Crane (2005).

15. As you would expect, there are debunking books about Scott such as Huntford (1985). Among the books that help to restore Scott's reputation: Fiennes (2003) and Solomon (2001).

16. I am not sure what emotion is the opposite of admiration. I assume it is contempt.

respects. The most interesting cases are those in which we admire a person as a person in advance of identifying her admirable qualities, or admiring her in a way that is not reducible to admiring qualities we can identify.

I think that the emotion of admiration is *sui generis*. We cannot reduce it to a more basic emotion, but we can say some things about it. I think of the admirable as something like the imitably attractive. To admire someone is to see the person as attractive in the sense of a model or an exemplar, and to feel a desire to imitate the person. The desire is highly defeasible, of course. There are numerous reasons why we would not want to imitate an admirable person, and certainly not every admirable act. I have no desire at all to imitate Sir Robert Scott's expedition to the South Pole, but I do find that my admiration for his sense of dignity includes an impulse to have that sense myself. To admire a person is to see the person *as* admirable, and to see her as admirable is to see her as imitable—as worth imitating. Since an emotion need not include a belief, I can see a person as admirable without judging that she is admirable, but if I trust the emotion of admiration, I trust the way I see the object of my admiration. In that case I trust that she is admirable and worth imitating. When I do that, given that trust includes belief, I believe that she is admirable and worth imitating.¹⁷

Humans, like other animals, imitate without reflection and often without conscious awareness that we are imitating. Much of our behavior is learned through imitation, from simple things like how to use a knife and fork, to much more difficult things like how to play the piano or to do philosophy. We also learn through imitation some things we wish we had not learned, like a parent's bad habit or the mannerism of a teacher. When we become consciously reflective about what we are imitating, we need a way to distinguish what is worth imitating from what is not. I suggest that we imitate pre-reflectively, but we rely upon the admiration we have upon reflection, the admiration we trust, as the most basic way to make the distinction between what is worth imitating and what is not. In other work I have described a way that admiration for exemplars can be the foundation for a comprehensive ethical theory in addition to playing a central role in moral training.¹⁸

17. For an empirical study of the urge to imitate the admirable, see Thrash and Elliot (2004).

18. I argue for this in Zagzebski (2004a: Part I), and give an overview of this kind of theory in (2010a).

My interest here is in the way that trust in admiration commits us to trust in others.¹⁹

Epistemic admiration plays an important role in acquiring intellectual virtues, learning the norms of inquiry, and assessing our beliefs. We find certain persons and certain epistemic behaviors admirable and we learn to think critically and carefully, to be open-minded, intellectually fair, and persevering by imitating persons who have those traits. More importantly for my point in this chapter, we learn to value traits like open-mindedness, intellectual fairness, and carefulness by admiring those traits and trusting the emotion of admiration. At some point we may conclude upon reflection that these traits are truth-conducive, but I think it is unlikely that our valuing these traits depends upon prior evidence that they are truth-conducive.²⁰

To be imitable, an epistemic exemplar need not be epistemically perfect. As long as someone is better than I am, imitation will lead to my epistemic improvement. But the more I trust my epistemic admiration, the more I trust that the admired person is admirable, and the more I trust that she is admirable, the more I will trust the particular outputs of her epistemic faculties. Self-trust in my emotion of epistemic admiration thus leads to epistemic trust in admired others, and it gives me a reason to trust the admired person's epistemic outputs, which is to say, her beliefs.

In chapter 3, I argued that trust in others is a commitment of self-trust because of the principle that I am rationally committed to treating like cases alike. My argument in this chapter does not depend upon that principle. The argument here is more direct. Since trust in epistemic admiration includes the desire to imitate the exemplar in her epistemically admirable beliefs, trust in admiration leads to trust in the desire to behave epistemically in particular ways, including the adoption of particular beliefs. Trust in my emo-

19. There can be an excess of admiration and imitation. I know a woman who feels threatened by the behavior of others because she thinks that their behavior, including such things as their choice of furniture and time of rising in the morning, makes an implicit demand on her to do the same thing.

20. James Montmarquet thinks it is a mistake to characterize intellectual virtues as truth-conducive in the sense of reliably obtaining truth. On the one hand, it is conceivable that there are epistemically virtuous agents whose beliefs are almost entirely false in (say) Cartesian evil demon worlds. On the other hand, there could be intellectually lazy agents who have mostly true beliefs. Reliability, then, should not be a necessary component of an intellectual virtue. See Montmarquet (1993: 32). I discuss the connection between intellectual virtue and reliability in Zagzebski (1996: 176–94).

tion of epistemic admiration gives me a *prima facie* reason to trust the desire to epistemically imitate another person by believing what she believes. In fact, insofar as I epistemically admire an exemplar more than myself with respect to a certain belief and I trust that admiration, I have a *prima facie* reason to trust her belief more than my own even when our beliefs conflict. Of course, there are other things to consider besides the relative admirability of another person and myself with respect to a given belief, and there are many ways in which my *prima facie* trust in the other person's belief can be defeated by aspects of myself I trust more. However, I think that trust in my emotion of epistemic admiration makes the problem of disagreement with others particularly acute. The problem is not just that other persons are as reasonable as I am, but that some of them are more epistemically admirable than I am, and it is self-trust that leads me to trust their superior admirability. I will return to the problem of disagreement in chapter 10.

In chapter 2 I argued that to trust myself when I am conscientious is the bottom line. There is no noncircular defense of the connection between epistemic conscientiousness and truth, and there is no noncircular defense of the connection between the conscientious use of my nonepistemic faculties and success in reaching the end of those faculties. But there are reasons for trusting conscientiousness that refer to nonepistemic faculties, and we can now see one of the most important reasons: I admire epistemic conscientiousness and I trust my admiration for it. I admire conscientiousness in others and trust that emotion. That also gives me a reason to trust others epistemically that does not go through the principle that I ought to treat like cases alike.

Many people take for granted that there is no connection between being epistemically admirable and being morally admirable or admirable in one's practical life. The attempt to identify epistemic norms and epistemically admirable behavior is cleaner if there is no more than an accidental connection between epistemic and moral admirability. But if I am right that there are persons who are admirable as persons, all things considered, they would have both intellectual and moral virtues, and I think that a study of these individuals may reveal ways in which their morally or practically admirable features illuminate the epistemically admirable. It is difficult to articulate or even to be fully conscious of the quality that makes persons admirable in this way. Traditionally, that quality has been called wisdom. For a very long time there was practically no scholarly work on wisdom or on wise persons,

either because scholars feared that it takes one to know one, or because they were skeptical that there is any such quality. Lately that has changed, and there is beginning to be substantial research on wisdom.²¹ It is interesting that persons deemed to be wise by others seem to combine exceptional personal development and what has been called “self-transcendence” with impressive cognitive abilities and a vast range of knowledge.²² Of course, there are some differences in the conclusions reached by one research group and another, and I am not endorsing the work of a particular group, but I want to call attention to a conclusion that many researchers have in common—that wise persons have advanced cognitive skills and a wide range of knowledge. There is clearly something epistemically admirable and hence imitable about wise persons in addition to their admirability and imitability in their moral and practical qualities.

If there are persons who have impressive epistemic accomplishments along with other features that make them admirable as persons, they may be helpful in getting us to see epistemic values and norms that are hard to identify if we focus our attention exclusively on those persons whose admirability is limited to the epistemic domain. For instance, one of the issues that has received attention from epistemologists recently is the value of true belief. A number of philosophers have pointed out that not every true belief is valuable, and hence, it is unlikely that every instance of knowledge is worth much attention.²³ There are differences between instances of knowing that are really valuable and those that are not. But if so, it is unlikely that we will learn the difference by imitating those whose admirability is limited to their ability to acquire knowledge or true beliefs. That suggests that one of the things we desire—to get valuable instances of knowledge, is something we can learn by imitating those persons whose admirability includes the epistemic but extends beyond it. If I am right about that, wisdom is an important object of study for reasons that overlap the aims of epistemology. We will

21. See, for example, many works by Robert J. Sternberg, such as (1998), (2000), (2003); Rowley and Slack (2009), Ardelit (2004), and Baltes and Smith (2008). A program at the University of Chicago has embarked upon an interdisciplinary approach to defining and researching wisdom. Its website can be found at <http://wisdomresearch.org/>. For an overview of wisdom by a philosopher, see “Wisdom” by Dennis Whitcomb (2011). See also Tiberius (2005) and (2008).

22. See Orwall and Perlmutter (1990).

23. Sosa (2003: 156); Zagzebski (2003b), Baehr (2009).

look at the ways in which self-trust leads us to treating wise persons as authorities in chapter 7.

In this section I have looked at the ways in which trusting our emotion of admiration leads to epistemic trust in particular others, both for their ability to get knowledge or true belief, and for their ability to identify epistemic values other than truth or knowledge. If we trust our ability to identify persons who are epistemically exemplary, we have a reason to imitate their exemplary behavior and their exemplary qualities. We have a *prima facie* reason to imitate their beliefs and, more importantly, to imitate the ways in which they acquire beliefs. We have a reason to think that their judgments of what is worth pursuing epistemically are trustworthy, and so we have a reason to make those judgments ourselves. Clearly, all these reasons are defeasible, but they can be important in guiding our epistemic lives.

4. TRUST IN THE EMOTIONS OF OTHERS

I have argued that I have the same general grounds for trust in my emotion dispositions as I have for trusting my epistemic faculties. The difference is that I have strong circular reasons to trust my epistemic faculties, but some of my emotion dispositions fail the test of circular justification. Nonetheless, some emotions pass the test. I then argued that trust in the emotion of epistemic admiration commits us to trust in epistemic exemplars. So there is more than one feature of the self trust in which rationally commits us to epistemic trust in others.

In chapter 3 I argued that epistemic self-trust commits me to epistemic trust in others. In this section I want to look at whether emotional self-trust commits me to trust in the emotions of others. Very roughly, I think the answer is yes. If I trust the disposition that produces a certain emotion, say admiration, and if I conscientiously believe that certain other individuals have the same quality of emotional conscientiousness that grounds my trust in my emotion disposition, then it follows from the principle that I ought to treat like cases alike that I ought to trust the emotion disposition of the other person. If I believe that Sam is as trustworthy as I am in the disposition that produces emotions of admiration and contempt, and Sam feels admiration for Sarah, I have a reason to admire Sarah. Of course, emotion has a feeling component, and coming to believe that someone else has a given emotion is

rarely sufficient to cause me to have the same emotion (leaving aside cases of empathy). But the question I find interesting is whether I should judge that I have reason to admire Sarah in such cases. I think the answer is yes. I have a *prima facie* reason to think that Sam's emotion fits its object, and so I have a *prima facie* reason to judge that Sarah is admirable even if I do not feel admiration. If a trusted friend finds eating meat repugnant, that gives me a *prima facie* reason to find it repugnant too even though I do not. If someone I admire feels indignation at a politician's behavior, that gives me a *prima facie* reason to feel indignation at the behavior. If someone whose disposition to compassion I find trustworthy feels compassion for someone whom others treat with contempt because of her crimes, that gives me a reason to feel compassion.²⁴

In discussing reasons for believing what trusted others believe in chapter 3, I said that the fact that someone whom I epistemically trust has a belief gives me a *prima facie* reason to believe it, but the reason may not be strong enough to give me a reason to form the belief. My claim was only that the fact that she believes it counts in favor of the belief. Similarly, my claim here is that the fact that a trusted other has a certain emotion counts in favor of the emotion, not that it alone is sufficient to make it reasonable for me to acquire the emotion.

My claim does not apply to all emotions. There are many emotions the fittingness of which does not properly transfer from person to person. I may find someone else's emotion trustworthy, but that does not give me a reason to have the same emotion unless I am in the same circumstances. Clearly, a student's fear of an upcoming exam does not give me any reason to fear it too. There are also emotions (Peter Strawson calls them "reactive attitudes") that depend upon a previous interpersonal transaction. The appropriateness of the emotion depends upon participation in the transaction. Gratitude, resentment, and guilt are in that category.²⁵ But even if we leave these emo-

24. See Festinger (1954) on social comparison theory, about comparing our emotional reactions with those of others.

25. However, Strawson (2008) says that people outside the transaction can have a generalized or analogous emotion on their behalf. "The generalized or vicarious analogues of the personal reactive attitudes rest on, and reflect, exactly the same expectation or demand in a generalized form; they rest on, or reflect, that is, the demand for the manifestation of a reasonable degree of goodwill or regard, on the part of others, not simply towards oneself, but towards all those on whose behalf moral indignation may be felt, i.e., as we now think, towards all men" (15–16).

tions aside, there are numerous emotions that, if appropriate for the trusted other, would be appropriate for me because there are no relevant differences in our circumstances, and the emotions do not depend upon a previous personal transaction to which I am not a party. My conjecture is that in addition to admiration, this would apply to many instances of contempt, compassion, indignation, sympathy, and reverence, as well as some instances of fear and love.

In my experience, we are much less likely to trust the emotions of others than to trust their beliefs, and I find that interesting. In cases in which we trust our own disposition to have a certain emotion, we do not generally extend the same trust to the emotions of other persons, even when we have the same grounds for trusting the other as ourselves, and there are no relevant differences in our circumstances. But at the same time, we often unconsciously adopt the emotions of others, particularly those whom we trust and admire. We unreflectively pick up responses of disgust, fear, or admiration from those around us, especially when faced with repeated exposure to the same emotion type in similar circumstances. If we already do it unreflectively, it seems to me, it would be better if we did it reflectively.²⁶

The phenomenon of imitation of emotion can lead to a change in our attitudes in deep and morally important ways. It seems to me that this was the primary mechanism in one of the most important social revolutions of modern times: the change in attitudes towards persons from different racial and ethnic groups. Of course, there were persons who gave arguments using premises that they thought were already accepted within their society, but most of the reasons to which reformers pointed were not new. In any case, those reasons were easily rejected by individuals who did not feel the same basic respect for persons of other races. The arguments only worked to the extent that underlying emotions changed. People acquired a feeling of respect for other races, and a feeling of disgust for those who did not respect them. Some persons might have directed a change in their emotions voluntarily, but I suspect that the most significant explanation for the change in emotions was imitation of the emotions of trusted others, particularly those whose moral wisdom was admired. People did not just pick up arguments,

26. In the psychological literature, this effect is often called emotional contagion. There are studies of this effect in the workplace and in parent-child relationships. See, for instance, Hatfield, Cacioppo, and Rapson (1994); Doherty (1998); and Barsade (2002).

they picked up emotions, and in a few generations the emotions, along with a large number of beliefs, were vastly different.

In my judgment unconscious imitation of the emotions of admired others in this way was reasonable. Certainly, there are cases in which we judge someone to be admirable whom we later judge not to be admirable, and I am not denying that. I am raising the question, "What should be our reflective judgment about the trustworthiness of the emotion of an admired person, one whose admirability applies to the domain of the emotion, and one whose admirability is something we continue to trust upon reflection?" We often trust the admired person's emotion unconsciously, and unconsciously adopt it. I am suggesting that there should also be conscious reflective trust. That is, we should judge that if the admired person's emotion is admirable, it is imitable. We have a reason to imitate it.

5. EXPANDING THE RANGE OF TRUST

To conclude this chapter, let us look at some ways that the range of trust expands because of trust in emotions. At the end of chapter 3 I claimed that it is reasonable to trust those who are trusted by persons we trust. Another way to get the same conclusion is to look at our own state of trust as the object of self-reflective scrutiny. I have argued that trust is a state that is a hybrid of the epistemic, the emotional, and the behavioral. We can raise the same questions about trusting trust that we raise about trusting emotions like admiration and epistemic states like beliefs. Why should we trust the state of trust in ourselves? What does that have to do with trusting the trust others have? We have seen reasons for extending our trust both to the epistemic states of other persons and to their emotions. I want to extend the same line of argument to trusting their trust.

Although we have no noncircular way to prove that what we trust is trustworthy, we do have circular ways of justifying any given instance of trust, using the test of what we trust when we are conscientious. I would not want to trust something that is not trustworthy, but I would also not want to fail to trust something that *is* trustworthy. My trust can fail to fit the circumstances either by trusting when I should not or by failing to trust when I should, by trusting too much or by trusting too little. Like beliefs, desires,

and other emotions, trust is a state that can succeed or fail at fitting the circumstances. Trusting what I trust when I am trying my best to trust the trustworthy is all I can do. Nobody could expect me to do more. But I can be expected not to do less.

Trusting my states of trust when I am conscientious in my trusting commits me to trusting others when they are conscientious in their trusting for the same reason that trusting my conscientious beliefs and emotions commits me to trusting their conscientious beliefs and emotions. Since trust, like belief, has an object, trust in the state of trust includes trusting that the object is appropriate to the emotion. This expands the range of what I conscientiously trust:

1. I trust that what I trust when I am epistemically and emotionally conscientious is trustworthy.
2. I have as much reason to trust what others trust when they are epistemically and emotionally conscientious as I have for trusting what I trust when I am conscientious.
3. Therefore, I should trust the things that are conscientiously trusted by others whom I conscientiously trust.

An important object of their trust is other persons. So trusting the trust of those I trust leads to the following principle:

Expansion of Trust Principle: I have reason to trust those who are conscientiously trusted by those I conscientiously trust.

The argument I have given in this chapter for the Expansion of Trust Principle supports the common consent argument I gave in chapter 3. There I argued that even though I have a stronger reason to believe what is believed by many others when the beliefs are independently acquired, it is still the case that when the beliefs are acquired from a single source, I have a defeasible reason to trust their trust in the source. So there can be chains of trust: A trusts B's trust in C's belief p . But the Expansion of Trust Principle also explains how epistemic trust can form a network. Epistemic trust can inhere in a community of persons, not simply a chain of individuals leading back to a single source. The bonding of a community through mutual epistemic trust is an important element in the conscientious spread of beliefs. The community may have a loose structure. What is important for the rationality

of believing is the dependence of members of the community on their trust in the other members of the community.

We can see the importance of trust in giving a person a reason to believe what others in her community believe by looking at the case in which trust is lacking, and at the worst, when there is distrust rather than trust. This phenomenon occurs with unnerving frequency in any context in which a belief is either political or becomes politicized. Take the example of the belief that there is human-caused global warming. Most academics of my acquaintance believe it, although they typically lack sufficient evidence to believe it. They believe it because they trust the experts who have come to that conclusion—good academics like themselves. Even if the climate change experts are in a different academic field, there is a loosely structured community in which we participate—the community of scientific and university scholars. My neighbors are skeptical of human-caused global warming for a parallel reason. They distrust the community of academics because they believe that academics tend to be driven by a political agenda, one they do not share. For both sides, epistemic trust or distrust in others is the most significant reason for the belief. I think that means that the ways in which trust is acquired and lost within a community and among the members of the human race is of the utmost importance in the common human pursuit of knowledge. A state that is a hybrid of the cognitive and the affective has a pivotal role in epistemic justification. I think that we should give it more attention in scholarly work in epistemology as well as in public discourse.

Trust and Epistemic Authority

1. AUTHORITY IN THE REALM OF BELIEF

1.1. Can We Believe on Authority?

Philosophers generally assume that authority applies only to the political domain and do not even mention epistemic authority.¹ In chapter 1 I suggested that that is because authority in general has been in disrepute in the modern era, and it is only because political authority is inescapable that philosophers make the attempt to justify it. Theorists recognize the tension between authority and autonomy, but the tradition of modern liberalism maintains that political authority can be derived from the authority of the self over the self. We can have authority without sacrifice of autonomy. In this chapter I will develop the line of argument of the last three chapters and will argue that epistemic authority can be justified in the same way. Authority in the epistemic domain is justifiable even if we begin with the natural authority of the self.

So far I have argued that the natural belief that the natural desire for truth is satisfiable makes epistemic self-trust rationally inescapable, and consistent epistemic self-trust and trust in the emotion of admiration commit us to trust in many other persons, in particular, to trust in epistemic exemplars. Given this conclusion, we have the resources to see how trust in certain others gives us the grounds for taking beliefs on their authority. I will argue that epistemic authority satisfies principles modeled on the general principles of authority proposed by Joseph Raz. In this chapter the main focus will be on the author-

1. For instance, the article "Authority" in the *Stanford Encyclopedia of Philosophy* (<http://plato.stanford.edu/entries/authority/>) addresses only political authority.

ity of another person's beliefs. We will look at the defense of stronger kinds of authority in chapters 6 and 7: the authority of testimony and its application to epistemic communities.

I will continue to approach the issue primarily from the standpoint of the subject, but let me begin with a question from outside the subject's perspective, which I brought up briefly in chapter 1. Is anyone ever justified in using coercion over the beliefs of others? This is a peculiar question since it is hard to see how it is even possible to successfully coerce someone else's belief, much less to do so in a justified manner, but it is also hard to get clear on the reason. In chapter 1 I mentioned one answer given by John Locke: "It is absurd that things should be enjoined by laws which are not in men's power to perform. And to believe this or that to be true does not depend upon our will" (2009: 33). On the next page, however, Locke offers a different and incompatible reason: "Nobody is obliged in that matter to yield obedience unto the admonitions or injunctions of another, further than he himself is persuaded. Every man in that has the supreme and absolute authority of judging for himself. And the reason is that nobody else is concerned in it, nor can receive any prejudice from his conduct therein" (35).

Neither reason stands up to scrutiny. Let us start with the second. Each person has, Locke says, "absolute authority" to judge for himself. Clearly, your beliefs and judgments are within your own private space. No one can do your believing for you; your beliefs would not be yours if you were not the one who forms them. And you are the one who can reflect upon them in a way that may lead to change. But Locke's claim that your authority over your own beliefs comes from the fact that no one else is concerned in it is surely mistaken. Obviously, other people are affected by your beliefs insofar as they lead to acts and your acts affect them, and even apart from the acts to which beliefs lead, a belief about another person can involve a wrong to that person. I take it that that is the reason for the moral injunction against vicious gossip.²

But we can leave Locke's second claim aside because his more influential reason for claiming that there is no duty to believe by obedience to authority

2. The idea that unfounded beliefs about another person can constitute a moral wrong to that person is the point of W. K. Clifford's example of the island religion (1999: 71–72). But not everyone agrees that "vicious gossip" is morally wrong. There is evidence that the spread of information about defectors in Prisoner's Dilemma situations pressures people to cooperate. See Nowak and Sigmund (2005).

is that belief does not depend upon the will. This position has received a lot of attention in recent philosophy, and I think it involves some confusions. In the preceding chapters I have assumed that we can exercise reflective self-control over our beliefs. Our beliefs are not like pains or passing thoughts. There are no norms for pains or passing thoughts, but there are norms for beliefs. We teach students those norms and remind ourselves to follow them. If there were no norms, there would be no point in inquiring about the rational way to respond to the beliefs of others or to evidence contrary to our beliefs. This does not mean, of course, that we are continuously exercising reflective control over our epistemic states. We not only acquire most of our beliefs without reflection, but as I stressed in chapter 2, we often resolve conflict between one belief and another or between a belief and some other part of our psyche unconsciously. Most of the time we do not monitor our beliefs because we have processes that adjust our beliefs with little or no conscious awareness. But on some occasions we do not resolve conflict automatically and have to figure out a way to do it. In other cases it is the process of reflection that generates conflict, and then it has to be resolved. Or maybe there is no conflict, but reflection shows us the reasonableness of adopting or eliminating certain beliefs because of our trust in certain norms. It is uncontroversial that we can exercise control over our beliefs and the processes by which we change our beliefs in these ways and others. Believing p because S told me that p is true and I ought to believe it is no harder in principle than believing p because p seems to me to be probable on evidence E . In both cases we can do it if we have adopted norms that make it a reasonable thing to do. I have not argued yet that there are such norms, but if there are no norms, that is not because belief is not under the control of the will.

But can we believe on command? When we consider that possibility, we tend to imagine an agent of the sovereign proclaiming, "All subjects take heed. From now on all of you must believe the following by order of the King," followed by the announcement of a list of propositions. To believe in this scenario seems impossible (although, as we will see in chapter 7, Hobbes thought that that is exactly what ought to happen). The problem with obeying such an order is not that one cannot obey on command, but that one cannot obey a command to believe when one has no reason to think that what is commanded is true. If the authority is political, then clearly there is no reason to think that what is commanded is true, but if the authority is epistemic, and the subject recognizes that what the authority tells us to

believe is likely to be true, it is no harder to believe on command than to believe ordinary testimony, as I pointed out in chapter 1. In any case, there is plenty of evidence that people *can* believe on the authority of an epistemic authority. For instance, the Catholic view of “obedience of the judgment” in assenting to Church teachings is not only psychologically possible, but is something actual people have done for centuries.³

However, my defense of epistemic authority in this chapter does not assume that we can believe on command. Perhaps this seems to undermine my aim since some philosophers maintain that intrinsic to authority is a right on the part of the person in authority to command, and a corresponding duty on the part of the subject to obey.⁴ But even if no one can legitimately command a belief, it does not follow that there is no epistemic authority. Authority is typically exercised in the practical domain by telling somebody to do something, but the way it is typically exercised is not essential to it. What is essential to authority is that it is a normative power that generates reasons for others to do or to believe something preemptively.

The feature of preemption is a distinguishing feature of authority from the subject’s perspective, and it will get considerable attention in this chapter. A preemptive reason is a reason that replaces other reasons the subject has. Believing what another person believes or tells me preemptively is parallel to doing what he tells me to do preemptively. In both cases what the authority does gives me a reason to believe or do something that replaces my other reasons relevant to the belief or act. The kind of reason authority gives me is what is essential to it. The right to command is not necessary.⁵

I am not suggesting that taking a reason to be preemptive is sufficient for either acting or believing on authority. If you love someone, you might take the fact that he or she asks you to do something as a preemptive reason to do it, but when you do so you are not treating the beloved as an authority. Similarly, it is possible (although less likely), that you will take the fact that the

3. See Vatican II (1964: pars. 12, 25) and discussion by Francis A. Sullivan (1983: 162–64).

4. This claim can be found in numerous places, for instance Wolff (1998: 4–9).

5. If an authority can give me a preemptive reason to believe or do something even though the authority does not have the right to command, that means that the right to command acts involves more than giving the subject a preemptive reason to act in a certain way, but I will not discuss practical authority in this work.

loved one has a certain belief as a preemptive reason to believe it. If so, believing preemptively is not sufficient for believing on authority. We usually do not think that the people we love have a normative power to give us preemptive reasons just because we love them, even if we choose to take their wishes as giving us preemptive reasons. In contrast, authority is such a power.⁶

Since I am approaching the issue of authority from the point of view of the subject, I will assume that authority is a power that the subject can recognize and to which she can rationally respond. It is not necessary for my account that the authority sees herself as an authority for a particular subject; indeed, for this chapter it is not necessary that she even be aware of her subject's existence.

One more preliminary matter is the ambiguity in the term "authority." Sometimes the word "authority" is used to refer to the person or institution that has authority, and sometimes it is used to refer to the normative power that person or institution has. I will use the term "authority" in both ways and will rely upon context to disambiguate my usage. So I will sometimes speak of a subject doing something because an authority told him to do so, but I will also speak of the conditions of authority or the similarities between epistemic and political authority. In the latter contexts authority is the normative power possessed by the bearer of authority.

Given my focus in this book on the subject's perspective, the question I want to raise, then, is the following: If I am a conscientiously self-reflective person, should I ever treat another person as having a kind of normative epistemic power that gives me a reason to take a belief preemptively on the grounds that the other person believes it? I will argue that the answer is yes.

1.2. The Epistemic Authorities

Let us go back to our conclusions in previous chapters. When I am conscientious, I will recognize that the fact that someone else believes p gives me a *prima facie* reason to believe p myself, and I have a stronger reason when

6. It is interesting that in Robert Adams's (2002) version of Divine Command theory, a human subject's obligation to do what God commands is connected to the reason a person has to do what a loved one requires. There is a sense, then, in which love *does* have authority in Adams's view. See his chapter 10 on obligation and chapter 11 on divine commands.

I conscientiously judge that she has the same qualities I trust in myself, or when my trust in her is based on trust in my emotion of epistemic admiration for her in some circumstances. My conclusions so far have been weak. I have not claimed that the reasons for believing what another person believes override other reasons, much less preempt them. In fact, even when there are no other reasons, my reason to believe what another person believes may be too weak to make it reasonable to believe *p*. I argued in chapter 3 only that the fact that another person believes *p* counts in favor of the credibility of *p*.

But the reasons I have for believing what someone else believes become more interesting when they are strong enough to be sufficient for belief. There are a number of possible cases of this kind. Maybe I judge that someone else is at least as conscientious as I am and her faculties just as good, and so her reasons are as good as mine would be if I had any, and hers are pretty good. So someone whom I conscientiously believe to be as epistemically competent and perceptually acute as myself looks out a window and reflects aloud on the scenery. Her belief that there are two cardinals in the ash tree can be a sufficient reason for me to believe the same thing. Or I might not be in a position to judge her degree of conscientiousness because I do not know her, but I trust her because she is trusted by persons I trust. The woman looking out the window might be in that category. In that case also I might have a sufficient reason to believe what she believes. Or maybe I lack even that much reason to trust her. The woman looking out the window might just be another human being. Given some assumptions about human perception of ordinary objects in ordinary conditions, even then it can be reasonable to believe what she believes because she believes it.

To take another case, suppose a stranger at the train station in Florence has jotted down directions for himself on how to get to the Uffizi, and I read his notes. It can be reasonable to take his general similarity to me as a sufficient reason to believe what he believes about the way to the Uffizi. I am not denying that in all of these cases I may detect contextual factors that make it prudent to get further evidence before adopting the other person's belief, particularly if I am going to act on it, but there are still many cases of these types in which it is not necessary to get such confirmation. It can be reasonable to adopt the other person's belief in the absence of further reasons for or against the proposition he believes.

There are harder cases in which I have independent reasons for or against the belief of a person I trust. In these cases I have to decide what I trust more

when I am doing my conscientious best. The weighing of reasons under conditions of stronger or weaker trust is complicated and I am not going to propose formulas for balancing reasons. I am interested in the special case in which I have reason to believe that another person is more likely to get the truth about some question than I would be if I attempted to get the answer to the question by a direct use of my faculties.

Our investigations in chapters 3 and 4 lead us to expect that this can happen. I may conscientiously believe that another person is doing a better job of what I am trying to do when I am conscientious, that is, when I am trying to get the truth in some domain. The domain might be very narrow. Maybe I am an accomplished cook, but I trust the people who test bakeware for *Cook's Illustrated* more than myself on this issue. It is my attempt to get the truth about baking that leads me to believe that they are doing a better job of it in this area. The domain can also be much broader. If I am not a cook, I might trust everything *Cook's Illustrated* says about cooking and cooking equipment more than myself. And the domain can be broader still. Scientific disciplines like biology and physics are large domains whose practitioners are not equally expert at all subfields within the discipline, but I might reasonably judge that they are more likely to get the truth about anything in the discipline than I am.

I have argued that another route to trust is epistemic admiration. I may epistemically admire someone more than myself in some domain, and if I trust that person more than I trust the way I get beliefs in that domain, I may trust whatever beliefs he gets in that domain more than I trust my own. In all these cases I trust *the way in which he gets* his belief more than the way in which I would get the belief. In cases of these kinds the conscientious thing to do is to let the other person stand in for me in my attempt to get the truth in that domain and to adopt his belief. This in broad outline is what I mean by epistemic authority.

2. THE CONTOURS OF EPISTEMIC AUTHORITY: THE PRINCIPLES OF JOSEPH RAZ

Before the modern era in the Western world, the legitimacy of authority was thought to derive from God, the ultimate authority. Of course, many people accept the same view today and I have no objections to it, nor do I object to

positions that base authority on natural law and the natural need for human beings to live in societies. Those approaches give us other models for justifying epistemic authority, but they are widely disputed in modern liberal societies, and so they do not serve my purposes in this book. What I want to do instead is to look at an influential account of the nature of authority that is consciously proposed within the framework of political liberalism, and which justifies authority from assumptions that any rational person would accept. The account I will use is that of Joseph Raz (1988). I have no position on whether Raz's account is adequate for all the purposes required of political authority. It is Raz's account of authority *as such* that I am interested in. I believe he has identified conditions strong enough that a subject who recognizes that the conditions obtain has a preemptive reason to follow the authority. Further, I believe that all of his theses of authority can be satisfied by authority in the epistemic domain.

The first condition of authority is *content-independence*, a condition Raz takes from H. L. A. Hart (Raz 1988: 35). An authoritative utterance gives the subject a reason to follow the directive which is such that there is no direct connection between the reason and the action for which it is a reason. The authority might have directed any number of different actions, and if he had directed a different action, the subject would have had a reason to perform that other action instead.⁷ So the authority might order citizens to drive on the right side of the road, but if the authority had ordered them to drive on the left, the subjects would have had reason to drive on the left. The feature of content-independence is compatible with the need for reasons to accept the authority as an authority, and we will return to that, but under the assumption that the subject has reason to accept the authority as legitimate, the subject has reason to do what the authority says that is not dependent upon the content of what the authority says.⁸

7. Raz (1988) gives his theses in the third person, although he gives some of his examples in the first person, and I assume that the first-person versions of his theses are straightforward. Notice that the third-person version justifies a stronger kind of authority than the first-person version since the third-person version does not require that the subject recognize that the authority satisfies the justifying conditions. It only needs to be true. We will briefly look at third-person authority at the end of chapter 6 and in the last chapter.

8. Raz argues that while content-independence is necessary for authority, it is not sufficient because threats and requests are also content-independent (36). I assume that epistemic threats and epistemic requests are unusual and will leave them aside for my purposes here.

There are cases of epistemic authority that satisfy the condition of content-independence. That is, there are cases in which an authoritative person or community's belief gives the subject a content-independent reason for belief. If the epistemic authority had believed a different proposition, the subject would have had reason to believe the other proposition instead. So if I am justified in believing what the climate scientist judges about the emission of greenhouse gases, I would be justified in the same way if the climate scientist had had a different belief about the emissions. Given that trust is directed towards the way in which the authority gets her belief, that is what we would expect. Beliefs taken from another person can therefore satisfy the condition of content-independence.

The second feature of epistemic authority and the one I have said is its most important feature is the *Preemption Thesis*. As Raz describes it, the Preemption Thesis states that the fact that an authority requires performance of an action is a reason for its performance that replaces other relevant reasons and is not simply added to them (42, 57–59).

The epistemic analogue of Raz's Preemption Thesis, formulated in the first person, is as follows:

Preemption Thesis for epistemic authority

The fact that the authority has a belief p is a reason for me to believe p that replaces my other reasons relevant to believing p and is not simply added to them.

This thesis says that in certain cases the authority stands in for me in determining whether p . I have reasons for thinking that the authority is more trustworthy than myself in some range of beliefs, based on aspects of myself that I trust, but that is compatible with letting the authority's determination of the truth within that range replace my own determination of the truth within the range.

What would justify me in believing what someone else believes in this preemptive way? To begin answering that, let us return to Raz's proposals.

Raz argues that in addition to the Preemption Thesis, authority ought to be understood in terms of two other theses: the Dependency Thesis and the Normal Justification Thesis (1988: chap. 3). The first states the general character of the considerations that should guide those who give authoritative directives. The second concerns the type of argument that would justify an attribution of authority.

Raz's *Dependency Thesis* states that all authoritative directives should be based on reasons that already independently apply to the subjects of the directives and are relevant to their action in the circumstances covered by the directive (47). Raz denies that the directive must correctly reflect the reasons upon which it depends, but it must be intended to reflect those reasons.

The *Normal Justification Thesis* (NJ thesis) states that the normal way to establish that a person has authority over another person is to show that the alleged subject is likely better to comply with reasons that apply to him if he accepts the directives of the alleged authority as authoritatively binding and tries to follow them, rather than by trying to follow the reasons that apply to him directly (53).

Although Raz is interested in political authority, the features of authority he identifies in these two theses do not depend upon the fact that the authority at issue is authority over actions, nor do they depend upon the fact that the authority issues directives that are intended to apply to particular subjects. Let us look at the relevance of the Dependency Thesis to the epistemic domain and frame it from the subject's perspective.⁹ A reasonably close epistemic analogue of that thesis says that an epistemic authority's belief is authoritative for me only if her reasons for believing what she believes reflect the reasons I would have if I were forming the belief myself. I think that this thesis is on the right track, but is probably false. Of course, it is not clear what reasons count as appropriate for me to use, given that my level of understanding and development of intellectual virtues is inferior to that of an exemplar. In many situations, the exemplar is not just a person who has better access to evidence and a better grasp of it than I have. She is a person who has more of the qualities I trust in myself insofar as I am epistemically conscientious. She may also have deeper understanding than I, and I trust that. She may have special insights that I trust, and in many cases I would not have those insights if I were forming the belief independently. The general point is that an epistemic authority is someone

9. Raz (1988) mentions epistemic authority only briefly, and he does so for an interesting purpose. He says that the Dependency Thesis is more plausible in the epistemic domain than in the domain of political authority, and he uses the plausibility of the former as a reason for accepting the latter. He does not discuss epistemic parallels to the Normal Justification Thesis or to the Preemption Thesis, nor does he discuss the content independence of epistemic authority. However, in Raz (2009b: 155–57) he discusses the application of his Normal Justification Thesis to the epistemic domain.

who does what I would do if I were more conscientious or better than I am at satisfying the aim of conscientiousness—getting the truth. But since it is due to self-trust that I trust the exemplar, the ground of my trust in her authority depends on myself, and her authority satisfies a version of the Dependency Thesis.

There is more than one epistemic version of Raz's Dependency Thesis broad enough to include the cases of taking a belief on the authority of an exemplar. I propose the following:

Dependency Thesis for the authority of another's belief

If the belief p of a putative epistemic authority is authoritative for me, it should be formed in a way that I would conscientiously believe is deserving of emulation.

In defending his Dependency Thesis of authority, Raz says that it is sometimes rejected because it is confused with what he calls the No Difference Thesis, which is in fact false. "The no difference thesis asserts that the exercise of authority should make no difference to what its subjects ought to do for it ought to direct them to do what they ought to do in any event" (48). Raz argues that there are many ways in which an authority's directive can make a difference to what the subjects ought to do while respecting the Dependency Thesis (chapters 3 and 4, *passim*). The epistemic analogue of the No Difference Thesis would be the thesis that the fact that an epistemic authority believes a certain proposition p should make no difference to what I ought to believe since the fact that the authority believes p is reason to believe what I ought to believe in any event. The falsehood of this thesis follows from what I have said about the Dependency Thesis. In fact, the thesis is very implausible. It would mean that if I know that someone with far greater intellectual virtues and background knowledge than I have believes p , I ought to believe p only if it was already the case that I ought to believe p . But my access to the relevant reasons for believing p might be exceedingly remote, and my ability to draw conclusions from the reasons I have might be inferior to the ability of the authority.

Let us now turn to the epistemic analogues of the Normal Justification Thesis. There are first-order reasons for believing as well as for acting. For instance, I have a reason to believe what is best justified on the evidence. This immediately raises the issue of whose evidence is relevant. Presumably

it is not just my own evidence, but the evidence available to my epistemic community, or to the human race. It is possible to be in a situation in which I conscientiously believe that I am more likely to believe something that is justified by the evidence available to my community or to the community of humans if I believe what a certain other person believes than if I try to figure it out myself—identifying and weighing the evidence myself. That person's epistemic authority could be defended by a principle like the following:

The authority of another person's belief for me is justified by my conscientious judgment that I am more likely to form a belief justified by the available evidence (suitably qualified) if I believe what the authority believes than if I try to figure out what to believe on the evidence myself.

But given what I argued in chapter 2, believing on the evidence is not my only epistemic aim, and not even the most important one. The evidence we can identify is derivative from what conscientious persons do in attempting to get the truth. Evidence is that which indicates truth for a conscientious person. So I care about the evidence because I care about the truth. But again, it is possible to be in a situation in which I conscientiously believe that I am more likely to believe the truth if I believe what a certain other person believes than if I try to figure it out for myself. Such a person's authority can be justified by an epistemic form of Raz's NJ thesis that I will use repeatedly in this book:

Justification Thesis 1 for the Authority of Belief (JAB 1)

The authority of another person's belief for me is justified by my conscientious judgment that I am more likely to form a true belief and avoid a false belief if I believe what the authority believes than if I try to figure out what to believe myself.

In chapter 2 I argued that what we ultimately take to be the truth is what survives conscientious self-reflection, and we always need to trust the connection between conscientious self-reflection and success in reaching the truth. That suggests another version of the normal justification thesis:

Justification Thesis 2 for the Authority of Belief (JAB 2)

The authority of another person's belief for me is justified by my conscientious judgment that I am more likely to form a belief that survives my con-

scientious self-reflection if I believe what the authority believes than if I try to figure out what to believe myself.

The JAB theses are not sufficient to justify taking a belief on epistemic authority without qualification. For one thing, a small difference between myself and the putative authority is not likely to be sufficient to ground authority. In addition, I might judge that even though the putative authority is more likely to get the truth whether p than I, the authority is not very likely to get the truth either, and in many such cases I should forego having a belief whether p . But if the authority is in a better position to get the truth than I, the authority is presumably also in a better position to judge whether she should believe or withhold belief, and I can still conscientiously judge that I should follow the authority in that case. There are also cases of competing authorities, cases in which the authority is more likely to get the truth than I am, but so are several other persons. There is no reason for me to trust this particular authority more than others. A partial response to this situation is the same as to the first. It is likely that an authority would be aware of competing authorities, in which case the authority's own response to the presence of competitors is relevant to what I should do if I am conscientious. The case of competing authorities is similar but not identical to the case in which I am an authority myself. We will return to cases of this kind in chapter 10.

One feature of JAB 2 that I particularly like is that it makes it clear that what justifies authority is not just the fact that the authority can help me to believe like *other* conscientious persons. Rather, the authority can help me to believe as I would believe myself, given my desires, emotions, and other beliefs, and given that I aim to have a belief that will ultimately survive my own conscientious self-reflection. The point of epistemic authority is to help me in believing conscientiously.

The topic of this chapter cuts across a number of popular topics in epistemology, and it is easy to confuse it with one of those issues. The epistemic authority of a belief is related to the topic of testimony, but the question for this chapter is not the issue of when it is reasonable to take a belief on testimony, nor is it about the way knowledge is transferred from person to person. The authority of testimony will be the topic of chapter 6. The parameters of the problem of this chapter also differ from that of the currently popular problem of reasonable disagreement. In that literature we are to imagine that

the agent starts with a belief, finds out that an epistemic peer disagrees with him, and then has to decide what to do. I am raising an issue about taking a belief on authority that arises whether or not I have already formed a belief on the matter. If I come to trust another person in the way she forms a belief whether p more than I trust my own way of forming a belief whether p , I have a reason to take her belief on authority. However, if I already have a belief whether p , that affects my judgment whether taking a belief on authority will survive my conscientious self-reflection better than forming the belief on my own.

To take a simple example, I used to think that it was a good idea to use sprays containing zinc to prevent head colds or to shorten their duration. Then I read that the Food and Drug Administration says these products can lead to the loss of the sense of smell and should be withdrawn from sale. I defer to the FDA in this matter since I trust it more than myself in this domain and I believe the conscientious thing to do is to let the FDA's belief preempt my other reasons for believing that zinc is a good thing to use (until the FDA comes out with another pronouncement, which it probably will). Taking my belief from the FDA preemptively is justified by both JAB theses. But the existence of a previous belief can affect my conscientious judgment that the FDA is more likely to give me a belief that will survive conscientious self-reflection than a belief I get on my own. If I already have a belief whether p , then presumably it has already survived whatever degree of self-reflection I devoted to it, which could be anything from very little to an enormous amount. So having a previous belief generally makes it less likely that I will conscientiously judge that taking a belief from someone else will survive conscientious self-reflection better than the belief I already have, but the principle is the same: Trust my conscientious judgment about what will survive conscientious self-reflection.

Trust in conscientiousness explains why a conscientious person ought to adopt a belief on authority more readily when the belief has little effect on her other beliefs and emotions. I am much more likely to make the conscientious judgment that another person's belief will survive my conscientious self-reflection better than a belief I get on my own when the belief is not relevant to most of my other beliefs. When I found out that one of my sons went to a wine symposium and now enthusiastically believes that the shape of the glass affects the taste of the wine, it was not difficult to adopt the belief on his authority because it was easy for the

belief to survive conscientious reflection on my total set of beliefs. But if I find out that someone believes that Doris and Walter Trinka were not my real parents, that would be a different matter. The fact that I would be upset shows I give it *prima facie* credibility, but the conflict between that belief and many other beliefs I have would make it far less likely I would adopt the other person's belief even if I have reason to believe the person is reliable. My point here is not the trivial one that it would be psychologically more difficult. The fact that an epistemically conscientious person adopts beliefs that are likely to survive conscientious self-reflection explains why it would be less reasonable to adopt the belief about my parentage than the belief about the shape of wine glasses.

3. PREEMPTION AND EVIDENCE

Let us look now at what it means to take someone else's belief as a preemptive reason to believe the same thing. It might appear that this requires doing something psychologically impossible: ignoring our own reasons for and against the belief. But letting a reason preempt my other reasons does not require ignoring my other reasons. In fact, it is because I am not ignoring them that I see that the belief of the authority has a certain status vis-à-vis my other reasons. If I stop at a red light because that is what the law requires, I let the fact that the law requires it be my reason for stopping. I do not ignore the fact that I would prefer not to stop because I am in a hurry, or that I believe it is generally safer to stop, or that I do not want to take the chance of getting a high-priced traffic citation. But if I stop because the law says to do so, that reason has the status of being my reason for stopping. It can be the reason even though I am quite capable of reciting many reasons for and against stopping.

The Preemption Thesis also does not require that I *decide* to let one kind of reason preempt my other reasons. I am not committed to the position that the executive self controls states of the self by decision or force of will. As I have said repeatedly, my position is that rationality is doing reflectively what we do automatically and sometimes unconsciously when we adjust states of the self to resolve dissonance. The degree and scope of our control is not at issue in the Preemption Thesis. The latter is a thesis about what I ought to do, what a rational person would do if conscientious.

Still, the Preemption Thesis will be resisted by those who insist that a reasonable agent treats an authority's belief as evidence of the truth of the belief that ought to be added to her other evidence. Why isn't it more reasonable to add my other reasons to the balance of reasons, perhaps weighing the authority's belief more heavily than my other reasons? Isn't the authority's belief just one more piece of evidence that I put into the mix of my total evidence?

This approach is tempting for those who think of all reasons as theoretical. I have said that theoretical reasons aggregate with each other, but they do not aggregate with first-person reasons. It is much easier to see how preemption works if the preempting reason is a deliberative, first-person reason like trust in the self or trust in others. If I trust someone else's superiority to myself, that is a first-person reason to take the fact that she has a belief as a reason for me to have it also. That is a different kind of reason than the third-person evidence that can be marshaled in support of the proposition believed. The fact that it is a different kind of reason makes it easier to see how it can replace another kind of reason.

But even if we are considering only third-person reasons, there is another problem with treating the authority's belief that *p* as just one reason among others to believe *p*. If I do so, I will worsen my track record in getting the truth. Joseph Raz anticipates this objection for practical authority and gives a response that also applies to epistemic authority. Raz says:

Consider the case in a general way. Suppose I can identify a range of cases in which I am wrong more than the putative authority. Suppose I decide because of this to tilt the balance in all those cases in favour of its solution. That is, in every case I will first make up my own mind independently of the "authority's" verdict, and then, in those cases in which my judgment differs from its, I will add a certain weight to the solution favoured by it, on the ground that it, the authority, knows better than I. This procedure will reverse my independent judgment in a certain proportion of the cases. Sometimes even after giving the argument favoured by the authority an extra weight it will not win. On other occasions the additional weight will make all the difference. How will I fare under this procedure? If, as we are assuming, there is no other relevant information available, then we can expect that in the cases in which I endorse the authority's judgment my rate of mistakes declines and equals that of the authority. In the cases in which even now I contradict the authority's judgment the rate of my mistakes

remains unchanged, i.e. greater than that of the authority. This shows that only by allowing the authority's judgment to pre-empt mine altogether will I succeed in improving my performance and bringing it to the level of the authority. Of course sometimes I do have additional information showing that the authority is better than me in some areas and not in others. This may be sufficient to show that it lacks authority over me in those other areas. The argument about the pre-emptiveness of authoritative decrees does not apply to such cases. (68–69)

Notice that if this argument is sound, it does not matter whether the authority's judgment is about what to do or what to believe, and it seems to me that it is sound.

There are amusing empirical studies that support the position that pre-emption leads to better results and is quickly learned by rats, but humans resist it and are outperformed by some other animals. Animals like rats and pigeons maximize. If the animal discerns that one choice is better the majority of the time, it chooses that option all the time. In contrast, humans attempt to match probabilities. For instance, if humans are trying to predict whether a red rather than a green light will flash, they proportion their choices to match the probability of the mechanism. So if the light has flashed green 75 percent of the time, humans will typically predict green 75 percent of the time, whereas in similar situations, rats will *always* choose the option that appears 75 percent of the time. The rats are right 75 percent of the time. The humans do worse!¹⁰ It follows that we are better off preempting, but it is hard for us to accept that. It is especially hard for us to accept that when the authority is human.¹¹

10. See Mlodinow (2008: 5–6). There are a number of studies that indicate that animals maximize whereas humans attempt to probability match. In a study that pitted Yale undergraduates against rats, the humans and rats had to predict where the food was in a T-maze. The rats did better. See Gallistel (1990: chap. 11). There is also a study that indicates that preemption is favored by the right side of the brain. See Wolford, Miller, and Gazzaniga (2000: 1–4). I thank Marian David for pointing me to these studies.

11. It is not hard to find anecdotal evidence that humans dislike preemption. We are very unforgiving of an authority who makes a mistake even if following the authority gives us a significantly better track record. It is possible (at least it used to be possible) that there are financial experts who really are better at investing our money than most of us are. Once we identify such an expert, we should follow that person's investment advice all of the time, as Raz argues. This does not require us to make the assumption that the expert is perfect. But woe to the financial expert who makes even one mistake! The wrath of his client will be upon him.

There is another source of resistance to preemption plus content independence. Suppose the authority's belief is something outrageous. Can't that count as a defeater of your belief that it is an authority? Yes it can, but that does not count against preemption. Suppose your physician tells you to take 4,000 pills an hour for the rest of your life.¹² I assume that you trust your belief that you should not take so many pills more than you trust your judgment that your physician is an authoritative guide to your health. To determine whether the physician is a better guide to your health than you are, you have a right to take into consideration anything that you find trustworthy when you are conscientious. But as long as you conscientiously think the physician is a better guide, you have reason to take the physician's directive as one that preempts your own decision about what you should do in that domain. And, of course, the same point applies to your *belief* that you ought to take 4,000 pills an hour. Epistemic authority has the consequence that trust in ourselves in some domain is replaced by trust in the authority, but it remains the case that a general trust in ourselves leads us to trust the authority, and the judgment that someone is an authority can be withdrawn.

Adam Elga makes a similar observation about the condition I have called content-independence:

Only in highly idealized circumstances is it reasonable to defer to someone's opinion *absolutely whatever* that opinion might be. For example, upon finding out that my forecaster is confident that it will snow tomorrow, I will follow suit. But upon finding out that my forecaster is confident that it will rain eggplants tomorrow, I will not follow suit. I will conclude that my forecaster is crazy. The same goes for the news that I myself will believe that it will rain eggplants tomorrow. In realistic cases, one reasonably discounts opinions that fall outside an appropriate range.

In addition, not even a perfect advisor deserves absolute trust, since one should be less than certain of one's own ability to *identify* good advisors. (2007: 483)

12. My example is adapted from an example by Thomas May (1998: 145). In May's scenario the physician tells you to take 46,000 pills an hour for the rest of your life. I am not sure it is even physically possible to follow those directions, and that makes the point unclear. Presumably the case we want is one in which, even if it were possible to do what the physician directs, you should not do so because the directive is absurd, and you trust your judgment of its absurdity.

I think Elga's point here is well-taken and it does not dispute preemption, although it requires a qualification of content independence. My judgment that someone is an authority in some domain is dependent upon a judgment whether his beliefs will survive my own conscientious self-reflection. If I conscientiously judge that a belief he has in the relevant domain clearly does not survive present self-reflection and is not likely to do so in the foreseeable future, I must decide whether to change my judgment that he is an authority or instead change my judgment about whether his beliefs will survive my future self-reflection. In some cases I will do the former. This is not always easy or straightforward because it means that my past judgment was mistaken, and I will need a judgment about how that happened that itself satisfies my conscientious self-reflection, and which reduces the likelihood that I will make the same mistake again. But the general point is that all I can do as a conscientious person is to make my psychic states as harmonious as I can while simultaneously attempting to make them fit their objects. Doing so often makes taking beliefs on authority the best strategy.

4. THE VALUE OF TRUTH VERSUS THE VALUE OF SELF-RELIANCE

The argument of this chapter has an interesting consequence for the value of truth in relation to the value of epistemic self-reliance. The person who values self-reliance presumably also wants true beliefs, but he values them more if he gets them a certain way—by using his own powers. One way to express this desire is that we want to get credit for the truths we acquire; we want the truths we get to be attributed to us as agents. It seems to me that this desire does not rule out getting beliefs from other persons. After all, if we acquire true beliefs on authority justified by JAB 1, then even though we are getting to the truth indirectly, it is the result of an intelligent and conscientious strategy that is a better strategy for getting the truth than a direct strategy would have been. So it seems to me that we get credit for *the truth* when we acquire it on authority. However, it is tempting to think that we would have deserved even more credit had we reached the truth by depending upon our own skills, virtues, reasoning powers, and background knowledge directly. But this raises a challenge for our epistemic values. Would we think that the subject who gets to the truth through the direct use of her own

powers deserves more credit for reaching the truth when that method is a less reliable strategy than taking the belief on authority? If you think the answer is yes, you must think that the value of believing by relying upon ourselves and not others is not just a constraint on the aim of true belief; it is an independent value. And it is not only independent, it trumps the value of true belief.

I am among those epistemologists who argue that what distinguishes knowledge from mere true belief is something like getting credit for the truth acquired.¹³ Roughly, the knower merits the truth rather than merely acquiring the truth. If that is right, then we should ask ourselves whether she gets more credit the more she relies upon her own powers. If she does, then the aim of knowledge and the aim of true belief can come apart. The aim of truth would lead her to adopt the strategy of taking many beliefs from others, whereas the aim of knowledge would lead her not to do so even when she is more likely to get the truth by relying upon others. That would be a very peculiar consequence. As I have said, I do not think this is a problem with the credit account of knowledge since my position is that the subject gets credit for getting the truth when she uses the best strategy available to her for getting the truth even when it involves relying upon others, but it is interesting to notice that if epistemic credit arises from self-reliance, and epistemic credit distinguishes knowledge from mere true belief, the aim of knowledge can oppose the aim of true belief.

This brings us back to chapter 1 and my unsuccessful search for a defense of the value of epistemic self-reliance. I concluded that it might simply be a basic value for some people, not a value that derives from the value of autonomy or the desire for truth. One conclusion of this chapter is that the value of self-reliance is partially opposed to the value of truth. If epistemic self-reliance is a good, it is not a good that pertains to truth; it can even oppose it. Perhaps JAB 1 has the consequence that we should take more beliefs on authority than many people will tolerate. That forces them to make a decision about which value is more important: truth or self-reliance. If someone values epistemic self-reliance more than truth, he should say so. Furthermore, if the value of knowledge arises in part from the value of self-reliance as well as the value of truth, he is going to have to decide what to say

13. See Sosa (1991: 277); Riggs (2002: 95) and (2009); Greco (2007: 116) and (2010); Zagzebski (1996), (2007: 151), (2003b), (2010a: chap. 5, sec. 4).

about knowledge in those cases in which self-reliance and the best strategy for getting the truth are opposed.

The defender of self-reliance should acknowledge something else. If he values epistemic self-reliance more than getting truth, the two forms of the JAB theses pull apart. He can be in a situation in which taking a belief on authority is justified by JAB 1 but not by JAB 2. Since I think that JAB 2 is more basic, I am not offering this as an objection to the position, only pointing out that it goes against the very strong presumption of the value of truth in contemporary epistemology.

5. STANDARD EPISTEMIC EGOISM AND EPISTEMIC AUTHORITY

Extreme epistemic egoists will reject the argument of this chapter, but notice that with a modification, JAB 1 should be adopted by standard epistemic egoists. For them, authority is justified by my *evidence* that I am more likely to get the truth if I believe what the authority believes than if I try to figure out what to believe myself. The egoist will not consider what I conscientiously trust that is not also evidence in whatever sense the egoist accepts, but I think epistemic authority is a commitment of standard egoism on the basis of a modified JAB 1. The argument of this chapter has the consequence that even the standard egoist should take beliefs preemptively from persons who satisfy the conditions given in JAB 1, modified as I have just said.

The kind of authority I have defended in this chapter is weak. For one thing, it applies to inanimate objects like GPS systems, thermometers, and other instruments. Egoists often find it easier to accept authority from an inanimate object than from another person. I have argued that they have just as much reason to treat other persons as authorities as GPS systems. This kind of authority is weak because the external authority is merely a meter of truth. The authority owes the subject nothing and the subject owes the authority nothing. Authority is more interesting when it involves interpersonal relationships. In the next two chapters we will look at authoritative testimony and authority within epistemic communities. That will permit us to investigate a stronger kind of epistemic authority involving a relationship of trust between authority and subject, and which will illustrate another side to epistemic trust—epistemic responsibility and the function of epistemic cooperation. This kind of authority is much closer to authority over acts.

The Authority of Testimony

1. CONSCIENTIOUS TESTIMONY

So far our investigation of trust in others has focused on trust in someone else's beliefs. But we cannot trust someone else's belief without finding out what it is, and we cannot do that except through the mediation of speech or writing or some other behavior. The behavior need not be verbal. For instance, I may want to attend a lecture but might not be sure where the lecture room is. I may then follow a group of people whom I know are going to the lecture. I assume that their behavior indicates their belief about the location of the room, and when I do so, it is their belief that I take to be authoritative.¹

Beliefs can be indicated by many other kinds of behavior, but the behavior that gets the most attention in philosophical discussion is verbal behavior, primarily oral, but also written. We need not agree on the hotly disputed topic of the nature of assertion to notice that a belief can be directly expressed in verbal behavior, and arguably, an assertion just is the direct verbal expression of a belief.² If so, the most direct way to find out what someone believes is to hear him assert it.

My view is that an assertion need not involve asserting something to someone else. If I overhear you reciting your creed, I hear you making assertions that express your beliefs. If I treat them as authoritative, I do so insofar as I think the beliefs the assertions express are authoritative. If I am right

1. I thank Elizabeth Fricker for this example.

2. Bernard Williams (2002: 74) calls an assertion the "direct expression" of a belief. That seems to me to be right. However, my view of assertion is tangential to the line of argument of this chapter.

about that, there is nothing authoritative about an assertion *per se*. What is authoritative is the belief the assertion expresses. If I had some other way to find out what you believe that was just as good, I would have the same reason to trust the belief that I have when you assert it out loud to yourself.

The situation changes when you tell *me* that *p*. In philosophy this is a species of the practice commonly called "testimony," a philosophical term of art that is intended to cover all cases in which a person A says that *p* to another person, B, who then believes *p* at least partly on the say-so of A.³ But there are many different things that can be going on in a testimonial interaction. Anything can be treated as evidence, and B might take A's saying that *p* as evidence for *p*, evidence that can be combined with any other evidence she has. In fact, the speaker A might intend for B to take her word as evidence. There are ways in which she can indicate that. A common way is to say something like "I believe that *p*," or "It's just my opinion, but I think that *p*," rather than to assert "*p*." By calling her hearer's attention to her own belief state, she is offering the fact that that state exists to the hearer as a reason to think that the content of her belief is true, but she is giving no assurance, and is signaling that she does not want to be held accountable if *p* turns out to be false. She can do the same thing by tone of voice or manner. She might simply say "*p*," but say it hesitantly or in some other way that indicates to the hearer that she does not want to be responsible to the hearer for its truth.

In contrast, the speaker might *tell* the hearer that *p*. When I tell you that *p*, I am expressing my belief that *p*, but I am not merely revealing my belief state to you. Suppose I tell you that a raccoon pulled down a bird feeder in our backyard, demolished the feeder, and scattered the parts (I am now telling you). I assert that the mischievous animal did the deed and thus express my belief that it did so, but I am asking you to believe it too because I said so. My intention with regard to you gives me responsibility. The practice of telling moves us from the private to the interpersonal domain. Belief has no interpersonal dimension, and if I am right that assertion need not presuppose a hearer, assertion can lack an interpersonal dimension as well. The practice of sharing my belief with someone whom I expect to use the fact that I have the belief as her own evidence has a minimal interpersonal dimension. In

3. In my experience, nonphilosophers do not understand what philosophers mean by testimony. They think testimony involves courts of law, not ordinary tellings. For a detailed discussion of the difference between legal and natural testimony, see Coady (1992: chap. 2).

that case my business is limited to telling her honestly what I believe, and I hand over to her the responsibility of making of my belief whatever she thinks it supports. In contrast, the practice of telling involves an implicit contract between speaker and hearer. When I tell you that p , I not only assert that p , but I intend that you accept p because I said so. I am asking you to trust me to give you the truth. When you believe what I tell you in that way, you accept my request and give me trust.

The model of telling I am using makes the paradigm case a face-to-face interaction, but as we will see in later chapters, the audience of a telling can be groups of people distant from the teller in time or space, and the teller need not be an individual. It can be a religious or scientific or political community. As long as A can ask B to believe p because A said so and B can accept the request, A and B can be parties to a telling. They need not be individuals, and they need not have direct acquaintance with each other.

My account of the practice of telling is similar to the "assurance model" of Richard Moran (2005) and the "second person model" of Benjamin McMyler (2011a). Moran makes the important observation that what the speaker intends in an act of telling is not extraneous to the hearer's reasons for believing what he says, an observation made some years ago by Paul Grice (1989). It is necessary to have an account of telling that explains this feature of the testimonial act. McMyler adds to Moran's model the idea that the hearer has the epistemic right to defer to the testifier when challenged by others. The testifier has the epistemic responsibility to make the belief justified in this way. So this model makes the justification of a belief acquired on testimony an interpersonal project.

The practice of telling as described by Moran and McMyler is what we would expect, given my argument that we cannot consistently treat other persons as simply sources of evidence. I have argued that it is because of trust in myself that I should trust them. If this is the reflective position of many people, and people reasonably grant each other *prima facie* trust because of shared qualities they reasonably trust, then we would expect to find practices that call upon people to trust in particular circumstances. We ask for and grant epistemic trust on particular occasions because we already have an expectation that we can assume the general trustworthiness of others and the particular trustworthiness of those who are conscientious. The additional trust that a recipient gives someone who tells her something

would not be reasonable were it not for a social context in which epistemic trust in other persons was *prima facie* reasonable.

The way of thinking about testimony I am defending requires responsibility and trust on both sides. First there are the responsibilities of the teller. When someone A tells B that *p*, A assumes responsibility to B for both the truth of *p* and her conscientiousness in believing *p*, and A acknowledges that B places epistemic trust in her if B believes what A tells him. If B takes *p* on the word of A, B has a valid complaint against A if A is either unconscientious in her belief or if the belief is false. If A is not acting in a conscientious fashion, and if it was reasonable for B to trust A in those circumstances, A has betrayed B's trust and it is appropriate for B to resent A. In fact, if *p* is later revealed to be false, then even if A was conscientious in telling B that *p*, A owes B something. She has a *prima facie* duty to retract her statement to B. Her false testimony would be understandable if it was done conscientiously, but it is appropriate for her to feel regret because she was unable to deliver on her contract. This is not to say that when A tells someone that *p*, she is guaranteeing the truth of *p*; but it is appropriate for her to take the heat if she is wrong.⁴

As Moran (2005) has pointed out, there is no explanation for feeling let down when the testimony is false if testimony is evidence. I think Moran is right about that. If testimony were evidence and the testimony turned out to be false, the testifier could reasonably say, "Well, you knew there was a certain probability that I was speaking falsely, so why are you mad at me? This situation is no different than the case in which it rains when you calculated the probability at 20 percent. You just got unlucky in your evidence." But I assume that the recipient of testimony would not accept that response, and justifiably so. We hold each other accountable for what we tell each other. If testimony merely provides evidence to the recipient, there is nobody for the recipient to hold accountable but herself.

When A tells B that *p*, B also has responsibilities. A requests B's trust, but B might reject it. That situation explains why Elizabeth Anscombe (1979) and J. L. Austin (1946) were right when they maintained (independently)

4. David Owens (2006) denies that responsibility in testimony involves responsibility to somebody. The argument is criticized by McMyler (2011: chap. 3). See also Fricker (2006a), which gives a detailed account of the speech act of telling as offering the audience a right to believe on the speaker's say-so, and which permits the hearer to rightfully complain if what the speaker says turns out not to be true.

that A justifiably feels insulted if A tells B that *p* and B does not believe her.⁵ I would add that A justifiably feels insulted if B *does* believe what A tells him, but only because B possesses evidence that A is reliable in this circumstance, and would not have believed what A says otherwise. Telling is a two-way street. The teller asks for trust and counts on the recipient to trust her. In return, she assumes the responsibility that goes with that trust, taking upon herself the epistemic burden of believing in a conscientious fashion, and doing so not only for herself, but for the recipient.

So the teller has certain expectations of the recipient in addition to the recipient's expectations of the teller. Telling bonds teller and recipient in such a way that both take some risk and both can be harmed by false expectations. The teller may count on the recipient to believe what she tells him because her future actions depend upon it. Or the teller and recipient may plan on joint action for a common purpose, which can be anything from planning a party to designing a rocket sled, and successful telling may be necessary for the achievement of those purposes. In chapter 2 I mentioned Annette Baier's definition of trust as acceptance of vulnerability to harm by another with the attendant judgment that the other person will not inflict the harm. On Baier's definition, the practice of telling typically involves trust on both sides of the telling relationship.

I think that part of the problem in adjudicating different models of testimony is that epistemologists try to force every case in which B takes a belief partly on the word of A into the same model. It follows from what I have said that there is a significant difference between informing someone that I believe *p*, on the one hand, and telling him that *p*, on the other. In the former case it is appropriate for the recipient to take the report of my belief as evidence of the truth of what I believe, evaluated his own way, whereas it is inappropriate for the recipient to take what I *tell* him merely as evidence of the truth of what I say.⁶ But sometimes

5. Austin draws a parallel between the speech acts of testifying and promising. He says, "If I have said I know or I promise, you insult me in a special way by refusing to accept it" (1946: 171). Elizabeth Anscombe (1979: 150) says, "It is an insult and it may be an injury not to be believed. At least it is an insult if one is oneself made aware of the refusal, and it may be an injury if others are." See also Edward Hinchman (2005: 565). Hinchman says the testifier is entitled to feel slighted if the hearer does not believe his testimony since the hearer was given an invitation to trust the testimony and it is rejected.

6. Jonathan Adler makes a similar point about the role Gricean conversational norms play in utterances designed to be merely taken as evidence versus utterances designed to be taken as direct testimony. See Adler (1997: 444–51).

the evidence model fits. Anything can be treated as evidence and there is nothing preventing me from taking the fact that someone else says p as evidence that p is true. But the practice of telling is importantly different, and it often leads to confusion to use the term “testimony” for both. In what follows I will use “testimony” for what I have described as a telling unless the context makes it clear that I intend to use “testimony” for both practices.

In some situations the two practices are combined in interesting ways, although again, the two ways need to be distinguished. For example, C might overhear A tell B that p . If C believes p , she has a very different kind of reason for believing p than B has. B believes p because of his trust in A. C believes p because she has evidence of the trustworthiness of A with regard to B. A and B are parties to a telling. C uses the telling as evidence. However, it is possible for C to become party to a telling indirectly. A, B, and C might have a relationship which is such that C knows that it is only an accident that A told B rather than herself. Maybe A is taking C and B to the airport and C hears A telling B she will arrive at 9:00. C might then take what A tells B as equivalent to what A would have told her. In that case she might reasonably regard herself as party to the contract. She could assume that A has taken responsibility for the truth of p and defer to A if asked to justify it, and A might accept that responsibility with regard to C as well as with regard to B. This requires special circumstances, however, and in general the person who overhears a telling is not a partner in the trust relationship.

Testimony can result in chains of epistemic responsibility in a more direct fashion. I may have a position of authority and trust my subordinates who report to me that certain events occurred. Because I trust them, I sign a letter to my superior containing the information passed on to me by them. I am responsible to the recipient of the letter for the claims I make, but I in turn defer to my informants who are responsible for their testimony to me. I defer to the testifier to satisfy the rules of conscientious cognitive function in forming his belief, but I am responsible to the recipient of my testimony for the fact that the belief is conscientiously formed. This is one of the ways that there is a division of epistemic labor in our practices of telling.

Since testimony involves responsibility to others, I think that in general, conscientious testimony requires a greater degree of conscientiousness than conscientious belief. Even when a belief is conscientiously acquired, the decision to express it to others brings with it an extra step—the decision to go on

record as having the belief in question.⁷ In general that brings with it the demand for an extra step of conscientious reflection. This is an ideal model and I do not deny that we are all expected to know that human beings are naturally garrulous. To some extent that counts as an excuse for failing to take the step of extra conscientiousness between what we believe and what we say, but I do not think it eliminates the responsibility and the conscientiousness required to exercise responsibility.⁸

The scope of trust in testimony is greater than the scope of trust in belief. The latter is merely trust in accuracy. The epistemic self-trust I argued we need in order to satisfy our natural desire for truth is trust in the general accuracy of our epistemic dispositions, and the particular accuracy of those dispositions when operating in conditions of epistemic conscientiousness. The trust I argued we owe to others when we are consistently conscientious is also trust in accuracy, and the arguments of chapters 3, 4, and 5 support taking beliefs from others in conditions in which we reasonably trust their conscientiousness, or admire them epistemically and trust our admiration for them. But trust in testimony involves trust in sincerity as well as trust in accuracy. When A tells B that *p*, A decides to tell B that *p*, which means that it is open to A to speak what she takes to be false or in some other way to mislead B. So if B trusts A's testimony, he must trust both A's accuracy and A's sincerity, and in telling B that *p*, A assumes responsibility to B for both her accuracy and her sincerity.⁹ She can be held accountable if she is either inaccurate or insincere.

The differences between trust in another person's belief and trust in another person's testimony suggest that a person's testimony can be less trustworthy than her belief, but it can also be the reverse. Sometimes it would be reasonable to take another person's belief that *p* as my own belief when it would not be reasonable to take her testimony pertaining to *p* because

7. Can one go on record as having a belief without testifying to the belief? I might record my beliefs in a journal that is not intended to be read by others. Perhaps then I am responsible to my future self but not to others.

8. I assume that an excuse does not lessen responsibility, only blame. For a related point see Frankfurt (2005).

9. For a discussion of the conditions of accuracy and sincerity, see Bernard Williams (2002: chaps. 5 and 6). The conditions of sincerity and accuracy were discussed over a century ago by A. F. Ravenshear (1899). See Moran (2005) for a discussion of the importance of the freedom of the testifier.

I reasonably trust her accuracy pertaining to p , but not the motives that would lead her to tell me that p . For instance, I might trust a physician's notes more than what he says to my face. On the other hand, even when I do not think that a person is trustworthy enough in her belief that p to give me sufficient reason to believe p on the grounds that she believes it, I might think so if she told me. That is because I might reasonably believe that she is more conscientious in her testimony than in her (mere) beliefs, and I might then trust her testimony enough to believe it when I would not have believed it based merely on the fact that she believes it. So I might trust a claim a scientist makes publicly more than the notes she writes for herself.

It follows from the differences between trust in testimony and trust in belief that reasonable belief on testimony depends upon nonepistemic trust. I am reasonable in taking a belief from another person, either in the preemptive way that is distinctive of authority or in weaker ways, only if I conscientiously trust both the sincerity of the source and her accuracy.

When does a conscientious person trust the sincerity of others? I think that an argument for a presumption of trust in the sincerity of others parallels the argument for trust in their accuracy that I gave in chapter 3. I have *prima facie* reason to think that others are as sincere as I am in what they say. If I naturally speak what I take to be the truth and do otherwise only for special reasons, then if I conscientiously believe that other people are basically similar to me in their natural dispositions, I have a *prima facie* reason to think that they naturally speak what they take to be the truth and do otherwise only for special reasons.¹⁰ In chapter 4 I proposed an Expansion of Trust Principle: I have reason to trust those who are conscientiously trusted by those I conscientiously trust. That principle applies to trust in the sincerity of others as well as to trust in their accuracy. If others conscientiously trust the sincerity of another person, that gives me *prima facie* reason to trust that person's sincerity also. I might also admire a person's character, and trust that admiration. In that case I might not have direct evidence of her sincerity, but perhaps I find her admirable in a way that would be incompatible with insincerity. Of course, in the particular case, I may have more direct

10. For a classic statement of this point, see Thomas Reid, "Essays on the Intellectual Powers of Man," (1983: 87; 94–95). For a more recent defense of the point, see Lyons (1997: 172–76).

evidence of the sincerity or lack of sincerity of a particular person. All of these reasons parallel reasons for believing in the accuracy of other persons' beliefs.

When another person tells me that *p*, my judgment of her sincerity is more complicated than my judgment of her accuracy in one respect. Her accuracy does not depend upon her motives towards me or her relationship towards me, whereas her sincerity does. However, if she obtained the information she tells me from other people, her accuracy depends upon features of her relationship with those people that may be even more difficult for me to judge than the relevant features of her relationship with me. So trust in the testifier's accuracy often includes trust in the sincerity of other persons towards the testifier. Even trust in the accuracy of someone else's *beliefs* often depends upon trust in the sincerity of other persons towards that person. That means that, broadly speaking, trusting what someone tells me is no more difficult or problematic than trusting the accuracy of what she believes. Of course, it is no less problematic either.

2. TESTIMONY AND DELIBERATIVE VERSUS THEORETICAL REASONS

My distinction between testimony as evidence and testimony as telling does not line up with the popular contrast between reductionist and nonreductionist views on testimony, but it lines up well with my distinction between theoretical and deliberative reasons. As Richard Moran (2005) and Benjamin McMyler (2011: chap. 3) have argued, reductionists and typical antireductionists agree that the testimonial event provides the hearer evidence on the basis of which she is justified in believing the testimony. The disagreement between them is over the way the evidence justifies the belief. Reductionists follow Locke and Hume and construe the evidence as inferential, whereas antireductionists generally take their cue from Reid and construe the evidence as noninferential. Either way, the recipient has the responsibility to make the belief justified. This view no doubt emerged in the early modern period because of the rise of a view of autonomy that stresses individual rather than corporate responsibility, together with the view that individual responsibility is a matter of properly handling one's own evidence. What is taken to be debatable is only the

nature of the relevant evidence, a debate that assumes the evidence model of testimony above.¹¹

The model of testimony as interpersonal trust is obviously not reductionist, but it is not Reidian antireductionism either. It differs from reductionism in a more dramatic way than the Reidian view because it is a model in which the recipient relies on the testifier for her epistemic justification. The testifier does not intend to offer evidence to the hearer, and the hearer does not take it to be evidence. Being invited to trust is not being offered evidence, and accepting it is not taking the testimony as evidence.

The distinction between theoretical, or third-person, reasons and deliberative, or first-person, reasons explains these contrasting models. The evidence model of testimony is the only model that makes sense if all epistemic reasons are third personal. In that way of looking at reasons, testimony is a process by which third-person reasons are passed around. They are either passed around directly—we acquire them as we see the world around us, or they are passed around indirectly by inductive inference. There is no other alternative.

I have said that anything can be treated as evidence. Whether or not the testifier intends to be offering the recipient evidence, and whether or not the recipient takes what the testifier offers simply as evidence, there is nothing preventing the recipient from making a good inductive inference that the testimony is probably true on the evidence. She can do that even if she accepts the testifier's request for trust. If that happens, the recipient has theoretical reasons for believing the testimony whether or not she also has deliberative reasons based on trust in the testifier. She might prefer to refer to her theoretical reasons for believing the testimony when communicating with other people, people who are not party to her relationship with the testifier. That, as I have said, is one of the functions of theoretical reasons. I think McMyler is right that she has the epistemic right to defer to the testifier for the justification of the belief, but there are many reasons why she might prefer not to exercise that right, and might prefer to offer theoretical reasons to other persons as grounds for the truth of the belief.

11. See McMyler (2011: chap 1) for a historical argument that Locke divorced testimony from authority and made it a species of inductive inference. McMyler argues that Locke's skepticism was about authority, not induction.

In contrast, the trust model of testimony is one in which telling gives the recipient a deliberative reason to believe what the speaker tells her. Trust is irreducibly first personal because it is a reason only for the person who has it. The speaker asks the hearer to give her trust and she may grant it or she may not. Her reason for granting it depends upon her relationship with the speaker. McMyler (2011: chap. 5) has defended the view that testimony gives the recipient second-person epistemic reasons for belief, modeled on Stephen Darwall's view of second-person reasons in the practical domain. I think McMyler is right that I am justified in believing what you tell me in a way that makes an irreducible reference to you as well as to me. When you tell me that p , you ask me to trust you, and if I accept your invitation to trust, I trust *you*. The reason I am justified is not that anyone of description F is justified in trusting anyone of description G and I happen to fit description F and you happen to fit description G. Trust is justified because of features of the reflective consciousness of persons, and those features are tied to the agency of the self. I have said repeatedly that my argument does not depend upon any particular view of the voluntariness of belief. However, it does depend upon the view that there is an executive self that is the agent of reflective consciousness. Reflective consciousness leads to changes in aspects of the self upon which the self reflects, including beliefs. There is a kind of control that we exercise over the self when we do these things. It is because of deliberative reasons that we exercise the control of an agent over our beliefs. We have no control at all over theoretical reasons. Theoretical reasons for p are facts that have logical or probabilistic relations to p , and I am often able to figure out what those facts and relations are. In contrast, deliberative reasons are features of my self that I use in deliberating about whether p . My trust in you is in that category. Notice that since my trust in you gives me a reason to believe something that it does not give anybody else and cannot give anybody else, that reason is first personal in my sense. What McMyler calls second-person reasons are a subclass of what I have called first-person, deliberative reasons. They are first personal because they make an essential reference to me. They are not third personal.

In chapter 3 I said that theoretical and deliberative epistemic reasons do not aggregate. Neither one reduces to the other and they do not reduce to some third kind of reasons. Deliberative reasons neither increase nor decrease the theoretical case for some proposition p . My deliberative reasons are not facts of the universe that affect the theoretical case for p that I present

to others, and I can grasp that even if *I* am the one giving the theoretical case. Similarly, my theoretical reasons are not added to my deliberative reasons to increase my deliberative reasons. So if you tell me that *p*, my belief that you have been right 90 percent of the time in a domain pertaining to *p* is not a reason to believe *p* that is added to my deliberative trust in you in the relevant domain.

The practice of testimony reveals the importance of human agency within the epistemic domain, and it raises important philosophical problems on the nature of the self. It is understandable that epistemologists want to leave those issues aside as much as possible, and if testimony can be reduced to a process of passing around theoretical reasons, investigation into the nature of the self can be prevented or at least delayed. But such a model of testimony makes invisible the epistemic bonds between individuals and the foundations of epistemic communities. Ironically, we live in an age of obsession with the self, but often as philosophers we do our best to ignore the self. Theoretical reasons appear to have the advantage of objectivity just because they are third personal, but in chapter 2 I argued that deliberative reasons are more basic for us than our ability to use any theoretical reasons. It is because of trust in ourselves that we take ourselves to be capable of identifying theoretical reasons, and it is because of trust in ourselves that we are committed to trusting the same features in others. Whether we think of testimony as evidence or as a contract of trust, the ultimate reason for taking beliefs on the testimony of others is deliberative.

3. PRINCIPLES OF THE AUTHORITY OF TESTIMONY

Let us look now at the relationship between testimony and epistemic authority. I have argued that someone's belief can be authoritative for me even if I do not find out what he believes by his testimony, so epistemic authority is not limited to testimony. There is epistemic authority that is not testimonial. There is also probably testimony that is not authoritative, although that remains to be seen. We will need to look at the principles of authority applied to testimony to see when they are satisfied. In discussing epistemic authority, I am particularly interested in the authority of testimony in the sense of tellings. Tellings that satisfy the conditions of epistemic authority are significant because authority in communities arises from tellings, not

evidence. Later we will consider whether the conditions for epistemic authority apply to cases in which a hearer takes what a speaker says as evidence.

As I have already pointed out, there are some differences between the trustworthiness of another person's belief and the trustworthiness of her testimony, but in many situations a case of authoritative belief will also be a case of authoritative testimony. I propose that there are cases in which the fact that someone else tells me that p gives me a content-independent reason for believing p that preempts my other reasons for and against believing p and that satisfies theses of epistemic authority that parallel the theses on the authority of another's belief. There are many variations of these theses, but I will propose the versions that seem to me most promising. Let me start with a proposal for the Dependency Thesis:

Dependency Thesis for authority of testimony

If the testimony that p of a putative epistemic authority is authoritative for me, then it should be formed in a way that I would conscientiously believe is deserving of emulation if I were testifying whether p myself.

I will not go through the different variations of this thesis that express Raz's insight and which are close to the Dependency Thesis for the authority of belief given in chapter 5. But I want to make a couple of remarks on the reason I chose this formulation.

I assume that the point of Raz's Dependency Thesis is that the authority is in a sense the subject's representative, and so the authority must take into account the subject's reasons for action in his directives. In formulating my epistemic versions of the Dependency Thesis, I made the authority closer to an exemplar than a representative. The authority does a better job of what I am trying to do in my epistemic behavior, and the authority's superiority is something I can recognize from my own point of view. That suggests that a constraint on the authority of testimony is that when I am conscientious, I would trust the way the putative authority produces the testimony she gives. I assume that means I would trust the accuracy and the sincerity of the authority (and probably also the clarity with which she expresses her testimony). If I were in a position to testify whether p myself, I might imitate the authority in the process leading to testifying that p . I would not necessarily do so since there is more

than one way that a trustworthy testifier could go about believing and testifying whether *p*, but it must be the case that what the authority does is one of the trustworthy ways of doing so. Since authority is constrained by what the subject takes to be trustworthy when she is conscientious, this thesis respects the primacy of the subject, so it is analogous to the Dependency Thesis for the authority of another's belief given in chapter 5, but it is not an exact analogue of Raz's Dependency Thesis.

The authority of testimony can be justified in the same way the authority of belief is justified. If my aim is to have a belief justified by the available evidence, and if I conscientiously judge that I am in a situation in which I am more likely to form a belief justified by the evidence if I believe what someone else tells me than if I try to figure it out myself, taking a belief on that person's authority is justified.

In chapter 2 I argued that what we take to be evidence is what we take to indicate the truth when we are conscientious, and the aim to get true belief is more basic than the aim to believe on evidence. I also argued that our ultimate test for truth is what survives conscientious self-reflection. Given the aims of truth and survival of conscientious self-reflection, there are justification theses for the authority of testimony that closely parallel the justification theses of chapter 5.

Justification Thesis 1 for the Authority of Testimony (JAT 1)

The authority of a person's testimony for me is justified by my conscientious judgment that I am more likely to satisfy my desire to get true beliefs and avoid false beliefs if I believe what the authority tells me than if I try to figure out what to believe myself.

Justification Thesis 2 for the Authority of Testimony (JAT 2)

The authority of another person's testimony for me is justified by my conscientious judgment that if I believe what the authority tells me, the result will survive my conscientious self-reflection better than if I try to figure out what to believe myself.

It seems to me that there are many cases that satisfy both JAT 1 and JAT 2. We can expect considerable overlap between the authority of another person's belief and the authority of her word, so situations that satisfy the JAB theses of chapter 5 will often also satisfy the JAT theses. But I think it is im-

portant that even in the many situations in which I take both the authority's belief and the authority's testimony regarding that belief as authoritative, there is a difference between taking the authority's belief because she believes it and taking the authority's word because she told me. The latter has the potential to be a stronger kind of authority and is closer to authority over actions. The testimony of an authority bonds authority and subject together in a way that permits complex epistemic and nonepistemic projects between them.

When one person conveys information to another, it is done for a reason—the good of the speaker, the good of the recipient, or a communal good in which both speaker and recipient participate, among other possibilities. Speakers do many other things when they tell another person something besides conveying information (or misinformation). They teach skills and acquaint their hearers with practices that have practical or symbolic importance. They speak in order to form emotional bonds with each other and with the communities to which they belong. They attempt to inspire hope or fear or respect or sympathy by the things they say. Sometimes the speaker is not attempting to speak the truth and expects the hearer to know that. The speaker may have no defect of sincerity or accuracy, but simply does not intend to be conveying information to the hearer. For instance, there are contexts in which it is more important to the speaker to say something witty than to say something true (and it seems to me that some guys think that most of the time). It is important for the hearer to be good at judging a speaker's reasons for speaking, since the reasons for accepting the speaker's testimony need to be understood against the background of the reasons the speaker has for saying it. The reasons typically arise from the context of a relationship between speaker and hearer, the relationships in which both of them participate, and the goals of their community.

I said that our investigation of the authority of another person's beliefs in chapter 5 indicates that there is epistemic authority that is not testimonial. Is there reasonable belief taken on testimony that is not authoritative? Presumably there is. I might take a belief from you on your word even though I do not judge that you are better than I would be at getting the truth in the domain of the belief. You only need to be good enough. In fact, I might conscientiously judge that I am better than you are in figuring out the truth in some domain, but I lack the time or opportunity or inclination to find out the truth directly. As long as I judge that you are trustworthy enough, it is

reasonable for me to take the belief from you. We do that all the time. I ask you what time it is and you glance at your watch. I am perfectly capable of looking at a timepiece myself, but choose not to. There is nothing unreasonable about taking a belief on testimony in that situation, but I am not treating you as an authority. The same point applies to testimony that I judge will survive my conscientious self-reflection. I can conscientiously believe on your testimony as long as I judge that doing so will survive my future self-reflection. You need not be better than I am at getting a belief that will survive my future reflection. You are not an authority for me, but believing on your testimony is reasonable.

Notice that there is a connection between conscientious belief on testimony and conscientious belief on testimonial authority. In both cases the goal of believing is respected, and both practices utilize an indirect strategy for reaching the goal—getting the truth and having beliefs that survive conscientious self-reflection. Skepticism about such an indirect strategy is hard to defend for anyone who takes those epistemic goals seriously. In fact, for anyone who seriously wants the truth, the avoidance of belief on authority is even less rational than the avoidance of belief on nonauthoritative testimony. The latter is rationally optional. The former is not.

4. TESTIMONY AS EVIDENCE AND THE AUTHORITY OF TESTIMONY

Let us consider a person who only recognizes what she takes to be theoretical reasons as reasons for belief. Is she justified in taking a belief on the testimony of an authority? Her model of the testimonial transaction is not the interpersonal model I have defended, but if she is a standard epistemic egoist, she should accept a modified version of JAT 1 for the same reason she should accept a modified version of JAB 1. The authority of testimony is justified for the egoist by her *evidence* that she is more likely to get the truth if she believes what the authority tells her than if she tries to figure out what to believe herself. Given my argument that she also has evidence that she will be worse off in getting the truth in that domain if she does not take a belief on the authority's testimony preemptively, the egoist should take a belief that satisfies the modified version of JAT 1 preemptively. The egoist is committed to the authority of testimonial beliefs in the appropriate conditions.

The same point applies to anyone who accepts the evidence model of testimony. It does not matter whether testimonial evidence gives inferential or noninferential support for the belief. Suppose I hear speaker *S* say that *p* and reasonably take that to be either direct or inductive evidence that *p*. I then believe *p* on *S*'s testimony. If I also have evidence that *S* is more likely to get the truth whether *p* than I am, then I ought to believe *p* on her testimony, and I ought to believe *p* preemptively for the reason I have already given. Defenders of the evidence model of testimony are therefore committed to taking beliefs on authority justified by the JAT 1 thesis, and they are committed to that whether or not they are epistemic egoists. Elizabeth Fricker recognizes this point. She uses the evidence model of testimony, but she argues that it is rationally mandatory to take a belief on testimony when it satisfies conditions very close to JAT 1.¹²

I cannot say whether defenders of testimony as evidence would accept JAT 2 as grounds for belief on authority. I do not see why they could not accept it, and my argument that survival of conscientious self-reflection is the basic ground for rational belief has the consequence that they should accept it, but I will leave that for them to consider. In general, I think it is fair to say that JAT 2 is independent of any particular view of testimony.

5. THE PARALLEL BETWEEN EPISTEMIC AND PRACTICAL AUTHORITY

The project of this book is undertaken primarily from the first-person standpoint: When, if ever, am I justified in taking a belief on the authority of another person? By approaching my project in this way, I am proceeding from the assumption that the ultimate authority over oneself is oneself. Whether I am attending to deliberative reasons or theoretical reasons, the conditions for taking a belief on authority depend upon reasons as I understand them. But authority in the sense that applies to acts does not depend upon features of the subject's consciousness. Raz's theses of authority are given in the third

12. Fricker says that a person properly accepts a proposition on the basis of testimony only if she correctly recognizes that the testifier is epistemically better placed than herself with respect to the proposition in question. In that case it is not only rationally permissible, it is rationally mandatory to defer to the other person (2006b: 232). Fricker therefore comes close to accepting the position for which I am arguing here, although she does not discuss preemption.

person, from a perspective outside that of the subject. His Normal Justification Thesis states that the normal way to establish that one person has authority over another is to show that the alleged subject is likely better to comply with reasons that apply to him if he accepts the directives of the alleged authority as authoritatively binding and tries to follow them, rather than by trying to follow the reasons that apply to him directly. Notice that this thesis does not say that the subject must be able to *see* that he is likely better to comply with reasons that apply to him if he tries to follow the authority. It is enough that it can be *shown* that that is the case, where the question of who shows it to whom is left vague. In any case, Raz does not require that the subject himself grasp the fact that the conditions justifying authority over him obtain. It is enough that they do obtain and that that can be shown to anybody asking for a defense of the authority's authority. The subject himself might be such a person, but even if the subject is shown that the conditions specified in the NJ thesis obtain, the authority's authority does not begin with the recognition of that fact by the subject. The normative power of the authority does not depend upon anything the subject actually recognizes.

If so, it is natural to wonder whether epistemic authority is strong enough to have this same feature. Here is an argument that if I accept JAT 1 or 2, I should also accept their Razian third-person analogues. When I think about why I accept authority justified by JAT 1 or 2, I realize that I trust my conscientious judgment that some person or community is more likely to get to my ends than I am myself because I think that the fact that my judgment is conscientious is the best indication I have that the judgment is true. If so, I must think that what really justifies me in taking someone as an authority is not my conscientious judgment that the person is more likely than I am to get me to my truth, but the *fact* that that is the case. If it is a fact that someone else is more likely to give me what I want than I am myself, then I *ought* to conscientiously judge that it is a fact, and if I ought to conscientiously judge that it is a fact, then authority is justified by a judgment I ought to make. But if so, it is but a small step to justification of epistemic authority theses that are even closer to Raz's Normal Justification Thesis:

Third-person JAT 1 thesis

The authority of a person's testimony for me is justified by the fact that I am more likely to satisfy my desire to get true beliefs and avoid false

beliefs if I believe what the authority tells me than if I try to figure out what to believe myself.

Third-person JAT 2 thesis

The authority of a person's testimony for me is justified by the fact that I am more likely to get a belief that will survive my future conscientious self-reflection if I believe what the authority tells me than if I try to figure out what to believe myself.

An indication that we implicitly accept the third-person version of the justification theses is the way in which we justify our beliefs on authority to other persons. If I wish to demonstrate to B that A is an authority for me, I will point out that I am more likely to get the truth and avoid falsehood in some domain if I take my belief from A. That is, I use the third-person version of JAT 1 to show that I am justified. I might not even mention to B that I made a conscientious judgment to that effect myself. If I do so, that is only to show B that the judgment is likely to be true. But for the same reason, B can point out to me when the conditions given by the third-person versions of JAT 1 or 2 are satisfied. B can point out to me that I ought to make the conscientious judgment that someone else satisfies the conditions of authority for me, and so I ought to take a belief on authority in that case.

The point here is that if I can justify to others my taking a belief on authority under certain conditions, they can justify to me that I should take a belief on authority under the same conditions. I do not deny that other people are not often in a very good position to tell when I should take a belief on authority, but the principle is the same. It follows that the third-person justification of epistemic authority is a natural consequence of the first-person justification. Notice that if the third-person versions of the JAT theses are true, the parallel between epistemic authority and practical authority is even closer.

Although the third-person theses make epistemic and practical authority close, there are still differences. It is crucial that political authority be justified apart from the perspective of the individual subject, but since there are no epistemic laws, epistemic courts, or epistemic punishments other than the punishment of failing to have a harmonious self, there are few practical consequences of failing to take beliefs on authority on grounds we do not recognize. But our issue here is what rationality requires of me, not what

“the authorities” can do to me. Of course, if someone wants to define authority as the right to coerce, and if it is impossible to coerce belief, then there is no epistemic authority, but I see no reason to define authority in terms of coercion.¹³ In any case, belief *can* be coerced in the sense that it is possible to impose sanctions on epistemic subjects who fail to take beliefs on authority.

In chapter 5 I followed Raz in claiming that the power to give preemptive reasons is sufficient for authority, and there is epistemic authority in that sense. In this chapter we have seen that there is epistemic authority in an even stronger sense of authority. When a teller satisfies the conditions of the JAT theses, she has the right to do something that is very close to commanding. Telling is not mere assertion, but includes an imperative to the recipient. When the recipient does what the speaker says to do and adopts the belief preemptively, she has done something parallel to obeying a command.¹⁴ Finally, if the third-person versions of the JAT theses above are true, the relation between an epistemic authority and a subject is even closer to the relation between a practical authority and the subject. There are differences between epistemic and practical authority, and I will highlight some of those differences in the next chapter. But I think we have seen in the last two chapters that epistemic authority is indeed a form of authority. It has all of the essential features of practical authority.

13. See Ripstein (2004) for a defense of the view that the state's claim to authority is inseparable from the right to coerce.

14. Raz says (2009b: 156) that theoretical authorities do not *have* authority over me even though they *are* authorities. But his reason for making this claim is that belief does not depend upon the will. As I have said, I do not see that someone's having authority over my beliefs entails that belief depends upon the will. But to get the strong kind of authority I am defending in this chapter, it must be the case that the authority can issue a belief imperative. Since ordinary telling does that anyway, there is no special requirement for the engagement of the will in believing what the authority tells me that is not already satisfied in ordinary testimony.

Epistemic Authority in Communities

1. EPISTEMIC AUTHORITY AND THE LIMITS OF THE POLITICAL MODEL

At the most abstract level, epistemic authority has the same features as political authority. However, there are differences between them that become significant when we look at the way the two forms of authority operate. Raz says he offers his work as a contribution to the literature on political freedom, and he aims to justify political authority within the framework of political liberalism (1988: 1). He moves from the political realm to the general realm of authority over action in order to present the contours of authority as he understands it, but he does so with an eye to applying his theses to the political domain and with eventual application to the authority of law. I imagine that this means he is operating with certain constraints.

One constraint is the desire to maximize political freedom and to minimize political authority. It is interesting that most modern political thought is motivated more by fear of bad authority than by desire for good authority. The idea is that it is more important to devise an account of authority that prevents tyranny than to give the bearer of authority the function of assisting the subjects in pursuing their individual and collective good.¹ With such an aim, it is reasonable to restrict authority as much as possible, compatible with having a tolerably smooth-functioning society.

A second constraint is that the account must be applicable to authority over large populations with no presumption of personal trust between authority and subject. The strong personal bonds that exist in small communi-

1. Yves Simon (1991) is an exception, but he is consciously antimodern. See also Jean Elshtain (2008).

ties such as the family or village cannot be assumed when the authority is distant from the subjects, they have very different aims, and there is no personal interaction between them.

These constraints do not apply to practical authority in many kinds of communities, and they are not very important when the issue is whether I should treat someone as an epistemic authority. When I take a belief on the grounds that another person believes it, I hardly need fear her power if she is not aware that I am taking my belief from her. And even if she is aware of it, as long as she thinks a certain proposition is true and knows that for that reason I think it is true also, she has done nothing to me except to give me what she thinks I want and which I do want: a true belief. She has not infringed upon my epistemic freedom any more than any other reason for me to believe something infringes upon my freedom. On the contrary, if I treat her as an authority because I am consistently self-trusting, my acceptance of her belief as the conscientious thing for me to do is an exercise of my epistemic freedom. There is no serious potential for epistemic tyranny.

The situation changes to some degree when we look at the authority of testimony. When a person tells me that *p*, he aims at getting me to believe *p* on his say-so. If he is sincere and has no ulterior motive, the situation is the same as in the previous case—there is no issue of power or infringement of freedom. He believes *p* and takes upon himself the epistemic responsibility of believing *p* conscientiously for both of us. But of course, he might not be sincere. His motives might be to manipulate me into having a certain belief for his own ends. If so, epistemic tyranny is a worry, but we need to identify the source of the danger to see how it should affect the way we think of epistemic authority.

The most obvious kind of epistemic oppression occurs when the epistemic authority is the same as the political authority. As I mentioned in chapter 1 (note 26), Rousseau argues that the Legislator should determine the articles of faith. Rousseau realizes that the state cannot compel a citizen to believe them, but it can banish whoever does not believe them as an antisocial being.²

An even more extreme view is that of Hobbes, who argues that religious authority derives from the Sovereign, who should also control the

2. Rousseau says:

Subjects have no duty to account to the sovereign for their beliefs except when those beliefs are important to the community. Now it is very important to the state that each

universities, known for their tendency to incite civil war.³ Both Hobbes and Rousseau think of epistemic authorities as subservient to the state. But in both cases it is the coercive power of the political authority that permits this kind of abuse. Since coercion is not at issue in epistemic authority in my sense, the flagrant oppression defended by Hobbes and Rousseau is not a problem, and it is not necessary to fashion a view of epistemic authority to avoid it.

These days nobody in a Western democracy would defend a view like that of Hobbes or Rousseau, but there can be a more subtle kind of epistemic oppression. Our beliefs can be manipulated in many different ways by others. Since nobody can get me to believe something against my judgment, manipulation of my beliefs only works if I am unaware of it. So epistemic oppression often operates by making it appear to the subject that she formed the belief through a rational process.⁴ Notice that this is not a case of

citizen should have a religion which makes him love his duty, but the dogmas of that religion are of interest neither to the state nor its members, except in so far as those dogmas concern morals and the duties which everyone who professes that religion is bound to perform towards others....

There is thus a profession of faith which is purely civil and of which it is the sovereign's function to determine the articles, not strictly as religious dogmas, but as expressions of social conscience, without which it is impossible to be either a good citizen or a loyal subject. Without being able to oblige anyone to believe these articles, the sovereign can banish from the state anyone who does not believe them; banish him not for impiety but as an antisocial being, as one unable sincerely to love law and justice, or to sacrifice, if need be, his life to his duty. If anyone, after having publicly acknowledged these same dogmas, behaves as if he did not believe in them, then let him be put to death, for he has committed the greatest crime, that of lying before the law. (*Social Contract*, Bk. IV, chap. 8, 185–86).

3. See *Behemoth*, Dialogue 1, cited in chapter 1, where Hobbes argues that the Church and the universities need to be disciplined so that they teach what the Sovereign wants. In the *Leviathan* Hobbes has this to say about the relationship between religious and political authority: "And whereas some men have pretended for their disobedience to their sovereign a new covenant made, not with men, but with God; this also is unjust: for there is no covenant with God, but by mediation of somebody that representeth God's person; which none doth but God's lieutenant, who hath the sovereignty under God. But this pretence of covenant with God, is so evident a lie, even in the pretenders' own consciences, that it is not only an act of an unjust, but also of a vile, and unmanly disposition" (*Leviathan*, Pt. II, chap. 18, par. 3).

4. The common theme of the invisibility of oppressive social structures appears in one of its first and most influential forms in J. S. Mill's argument in *The Subjection of Women* (1869). Mill argues that men have designed the whole education of women to enslave their minds in a way that women do not notice. I would add that one thing that made this harder for women to notice is that men did not notice it either. For an interesting recent theory of the way in which political oppression can occur via epistemic oppression, see M. Fricker (2007).

believing on authority because of a conscientious judgment that the source of the belief is authoritative. On the contrary, the believer is oppressed because she does *not* conscientiously judge that the belief comes from an epistemic authority. The belief is therefore heteronomous in the sense described at the end of chapter 1. Unlike the political case, we do not prevent this kind of oppression by restricting the range of epistemic authority. On the contrary, a subject can prevent it only by taking care in making the judgment that someone is an epistemic authority.

We obviously want to protect ourselves from believing falsehoods taken on the word of epistemic authorities who betray our trust. We also want to avoid taking beliefs from persons whose pronouncements are false even though sincerely given. So we want to know whom to trust epistemically. I suggested in chapter 1 that the motive for much of modern epistemology is a general distrust of others. In part we inherited that attitude from Descartes, who complained at the beginning of the *Meditations* that epistemic authorities had let him down. That danger is not the equivalent of political tyranny; it is the danger of being misled. So I might judge that I have been led astray by someone I took to be an epistemic authority more often than authority has helped me to get truths I would not have gotten without it. The political analogy to that situation is one in which a political authority makes wrong decisions more often than not. That is unfortunate, but it is not necessary to fashion restrictions on the justification of epistemic authority to prevent it. It is only necessary to find someone trustworthy.

I suggested that the other desideratum behind Raz's theses is to make authority applicable to the governance of large populations where the bonds of trust between authority and subject are minimal. Epistemic communities widely vary both in size and in the level of trust within them, but even when epistemic authority extends over a large population, it is not comparable to the political case. We must devise a system of governance for the political state that motivates the subjects to obey even when they do not trust those who govern, but there is no comparable problem in the case of epistemic authority, even when the putative epistemic authority claims authority over a large population, such as a governmental agency that issues weather alerts or warnings about health hazards. It is not necessary to devise a system to get people to believe an epistemic authority even when they do not trust it.

Raz's model of authority is helpful in understanding the general features of epistemic authority, but it hides the function of trust in authoritative relations

in communities small enough to have communal ends and interpersonal relationships that bind the community together. Such communities often include both practical and epistemic authority. Epistemic authority is important in teaching and training, as well as in enterprises that involve cooperative effort. The person in authority and the persons who are subject to the authority count on each other to do their part in a communal undertaking that includes authoritative testimony as a necessary component. The strength of the authority and its depth in the lives of the subjects vary enormously from one community to another because of the nature of the goals that hold the community together and the degree of trust between the members. I think that these differences can affect the way authority is justified in a community.

2. AUTHORITY IN SMALL COMMUNITIES

2.1. Justifying Authority in Small Communities

When we look at JAT 2, we see an important way in which it is not the epistemic analogue of Raz's Normal Justification Thesis. Raz's thesis anchors authority in the subject's reasons for action prior to and independent of reasons she acquires under the authority—an understandable constraint, given the aim of justifying political authority. JAT 2 anchors authority in the subject's reflective judgment, but not necessarily the subject's prior reflective judgment, or her judgment independent of trust in the authority. I think that the strategy of appealing to the subject's reflective judgment explains the justification of authority in small communities better than the NJ thesis. Since such communities contain both epistemic and nonepistemic authority, it is useful to look at some of their distinctive features.

Consider communities that exist for a well-defined purpose, such as building decks, performing orchestral works, or investigating homicides. These communities generally have leaders whose authority is defined by the purpose of the community. The subject of authority has reason to accept the authority's legitimacy because a condition for serving the purpose of the community is that there are subjects who obey the person in authority—the leader of the orchestra or homicide team, or the builder. The subject need not have reasons to act in a way that serves the community's purpose in advance of becoming a member of the community whose purpose she is

serving. Authority in the community is justified by reference to the communal purpose, not reasons the subject has apart from her participation in the community. Of course, if the community is voluntary, she has reasons for joining it, but the legitimacy of the authority over her is not dependent upon her having those reasons.

For instance, I may have reasons to become a homicide detective in a certain city. If I become a member of a homicide team, I now participate in a community that has communal purposes, and the legitimacy of the authority in the community is justified by reference to those purposes, not to the purposes I had in joining. If I lose my reason for joining the team, I am still subject to its authority as long as I am a member. The same point applies to building decks and performing music. I have reasons to join an orchestra or to get a job building decks, but the authority of the orchestra director or the builder who hires me is not contingent upon the existence of those reasons. The person in authority can lose legitimacy, but not because of his relation to my reasons for belonging to the community. It is his relation to the community's reasons that matters.

The same point applies to authority in some communities that are not voluntary, such as one's birth family. Authority in the family is not contingent upon reasons the child has for being a member of the family. In fact, one does not need reasons to be a member of a family, and it is not clear that it is even coherent to ask for reasons. But a family has communal purposes with reference to which the parents have legitimate authority.

This means that Raz's NJ thesis does not explain what justifies authority in certain communities. If authority in a community is justified by second-order reasons to think that obeying the authority makes it more likely that the members of the community will act on certain first-order reasons, it is often the *community's* first-order reasons that are most relevant. As a member of the community, the community's reasons for acting are my reasons.

There is a second way that Raz's NJ thesis is not sufficient to explain authority in small communities that contain a high degree of trust between authority and subject. It is sometimes reasonable for a subject to take the word of the authority as a reason to revise the first-order reasons for the sake of which she originally placed trust in the authority. This often happens in the relationship between teacher and student. The student believes that he will do a better job of reaching his goal of learning a subject if he follows the teacher's directions than if he attempts to learn on his own. So the teacher's

authority is justified by the Normal Justification Thesis. But one of the things the student can learn is that he should revise his understanding of the goals of study, the goals he was serving in accepting the teacher as an authority. This often happens in learning philosophy since the student's initial goal is typically modeled on learning in information-based fields. So he may undertake a study of philosophy because he wants philosophical information. He wants to find out what the great philosophers thought. As he learns more philosophy, if he trusts what he learns, then what he trusts when he is conscientious gradually changes, and his trust in the teacher can make it reasonable for him to let the teacher's testimony about the goals of philosophy change the reasons he initially had for studying philosophy and accepting the teacher's authority. In this way it can be reasonable for him to accept a revision of his first-order reasons on the teacher's authority even though it was trust in the teacher's ability to aid him in acting on those reasons that led him to accept the authority of the teacher.

The teacher is closer to an exemplar than to a leader of a team. There are communities in which the authority in the community is an exemplar who transmits a skill or an art or a *techné* in Plato's sense. An exemplar's authority can be justified either by the subject's admiration for the exemplar and her trust in that emotion, or by the subject's reflective judgment that the exemplar has the qualities she trusts in herself in a greater degree than she has herself. But if the exemplar is teaching the subject how to play the cello or baseball, or how to practice law or do philosophy, the justification for the exemplar's authority is ultimately success in the subject's learning of the skill or art, or acquiring competence in the practice. This is one way that the future justifies the present, a theme to which I will occasionally return.

The same point applies to a subject learning an entire way of life, such as a new member of a monastic community. The *Rule of St. Benedict* is not only a spiritual handbook, it is also a brilliant defense of Benedictine authority on grounds that a modern liberal would accept—a surprising fact, given that it was written 1,500 years ago. Benedict begins with an appeal to those who are longing for "days of real fulfillment" (2003: 11). Within the space of a few paragraphs, Benedict appeals to the monk's first-order reasons for living, and the monk's second-order reasons for thinking that living as a monk according to the *Rule* and under the direction of an abbot is a better way to live in the way he aims to live than to do it alone or to do it by becoming one of the other kinds of monks Benedict describes in chapter 1—a Sarabaiter or

gyrovague, who have no abbot or no rule. So it appears to me that Benedictine authority satisfies Raz's Normal Justification Thesis. But the Benedictine monk, like the philosophy student and most of us living outside a monastery, begins with only the vaguest idea of his ends and the first-order reasons upon which he should act. Benedict presents guidelines for Christian practice in chapter 4 of the *Rule*. Most of these rules are familiar to the monk and he has already accepted them, but some of them are new, and there would be no reason for Benedict to list them in detail and to prescribe that they be read aloud regularly if he did not think that the monk would not also accept these guidelines on the authority of the *Rule*. This is an example of an authority who identifies the individual's prior first-order ends, makes them specific, and adds to them in certain ways that the individual reasonably accepts on the word of the authority.

Benedictine authority illustrates a third limitation of the NJ thesis. What attracts the monk or potential monk to the *Rule* is its admirability and the admirability of its author. To imitate the *Rule* is most generally to try to become a certain kind of person. Putting oneself under the direction of the *Rule* is a better way to become that kind of person than to try to do it directly. But the end of becoming the kind of person proposed to the monk in the *Rule* need not be an end he has in advance of finding the *Rule* admirable. On the contrary, he might want to become that kind of person just because he finds the *Rule* admirable. So he could accept the authority of the *Rule* under an abbot because he sees that as the best way to become the kind of person that he wants to become, but he has that end because he trusts his admiration for the *Rule* that sets out for him that ideal. That indicates that trust in one's emotion of admiration can ground authority directly, aside from the fact that one has a second-order reason to think that obeying authority will serve one's first-order ends.⁵

The monastic life is also a good example of my point that authority in the learning of a practice is justified by success in learning the practice, a success that the subject could not acquire without it. The success of monasticism as a practice over many hundreds of years and in many parts of the world is a justification of the authority structure that successfully produces monks who live the life they hoped to live. But what really justifies it, I think, is not

5. I have investigated the way the *Rule of St. Benedict* satisfies Raz's theses of authority in Zagzebski (2010b).

that they succeed in living a life they set out to live when they were novices, but that they succeed in living a life that survives their own future conscientious self-reflection, a life of harmony within the self. I think that this means that even if we are talking about practical authority in small communities, not epistemic authority, a justification of authority thesis analogous to JAT 2 is better than Raz's Normal Justification Thesis.

I have suggested four ways the NJ thesis does not explain what justifies authority in small communities containing a high degree of trust:

1. The community has communal ends, and authority can be justified by reference to those ends rather than the ends of the individual members of the community.
2. A member of a community can have reason to modify his first-order ends on the word of the authority.
3. Authority can be justified by the subject's trust in his admiration for the authority as an exemplar rather than by the authority's ability to help him act on his first-order ends.
4. Authority over persons learning a practice can be justified by the subject's success in learning the practice.

Each of these features of authority can be handled by a generalization of JAT 2. I propose that this is the most general thesis justifying authority of all kinds:

General Justification of Authority Thesis (GJA)

The authority of another person is justified for me by my conscientious judgment that if I do what the authority directs (or believe what the authority tells me), the result will survive my own conscientious self-reflection better than if I try to figure out what to do/believe myself.

I conclude that what Raz calls the "normal" way to justify authority is not a necessary condition for justified authority in certain communities. This is not an objection to Raz because he makes it clear that his NJ thesis does not give necessary conditions for authority. But the ways in which the NJ thesis is inadequate are rectified by GJA. This point is important since there is both epistemic and nonepistemic authority in small communities, and in these communities it is precisely the ways in which authority is disanalogous to political authority that explains the way authority is justified.

2.2. Justifying Epistemic Authority in Small Communities

The examples of authority in small communities not only reveal the limitations of Raz's NJ thesis and the advantages of GJA, they also reveal some limitations of JAT 1 and advantages of JAT 2 for justifying epistemic authority. First, notice that there is epistemic authority in communities whose end is not epistemic, and the epistemic authority is justified by reference to the nonepistemic end. The end can be the transmission of a *techne* or skill or art, or a whole way of life, such as the life of Benedictine monasticism. The *techne* might be epistemic, as in the example of philosophy, but it often is not. Medicine, painting, baseball, playing the cello, and even deck-building are practices transmitted from an expert in the practice to others who gradually acquire expertise through training in the practice. The subject has reason to believe what the authority tells her because that is a necessary condition for learning the practice.⁶ Epistemic authority is justified by a reasonable expectation that the authority will serve an end that is not necessarily epistemic or is only partly so. That is not explained by JAT 1. I would not deny that a student of the cello reasonably believes that what her teacher tells her about the proper way to hold the bow is probably true, and similarly for most of the other beliefs novices in a practice learn from their teachers, but finding out the truth about how to hold the bow is only a means to an end—the end of learning how to play the cello well. What really justifies the student in taking beliefs on the teacher's authority is that that is what she conscientiously judges she should do if she wants to become a master of the practice. She judges that her future reflection upon her beliefs in conjunction with her desires and other states of the self will be more harmonious if she follows the teacher's authority than if she does not. JAT 2 therefore fares better than JAT 1 as a justification for taking a belief on the authority of a teacher of a practice.

Second, the example of the way authority operates in teaching philosophy shows how the subject sometimes needs to be taught the end, not just the means to the end. It can be reasonable for the student to modify her first-order reasons for the sake of which she consulted authority, and to do so on the word of the authority. I am not denying that there is philosophical truth, but the end of studying philosophy is not limited to truth, and

6. This is a point made by Simon (1991: chap. 3). It is also argued by Teichmann (2004).

this is the case even though the end is intellectual, not practical. Since JAT 2 is broad enough to permit epistemic ends other than truth, this case indicates another limitation of JAT 1, but not JAT 2.

Third, admiration for an epistemic exemplar can ground authority directly, and that can also explain why the subject might trust the authority more than her own judgment of the epistemic goals. If trust in oneself is more basic than the NJ thesis, that explains why it is sometimes reasonable to revise the reasons for accepting authority to which the NJ thesis refers on the basis of trust in that same authority, and it also explains why it can be reasonable to accept authority because of trusted features in the self other than beliefs about the authority. Epistemic admiration is one of those.

It is difficult to identify epistemic admiration in a way that does not beg any questions about the realm of the epistemic, but I think we distinguish admiration for persons in the intellectual realm from admiration in the moral or practical realms. The range of the epistemically admirable includes those persons who exemplify understanding or insight as well as knowledge or true belief, but I am excluding persons who exemplify such virtues as justice, benevolence, and efficiency. So I am leaving open the possibility that the epistemically admirable person might be unjust, or that the benevolent person is not epistemically admirable. A harder virtue to classify is wisdom, which combines intellectual and moral admirability. Could my admiration for a person's wisdom give me reason to accept that person as an epistemic authority? I think the answer is yes, but I will not pursue this question here. I want merely to extend my point in chapter 4 that trust in my emotion of epistemic admiration gives me reason to epistemically trust exemplars. Now I want to say that trust in that emotion can give me reason to take certain exemplars as epistemically authoritative. The justification of authority in these cases is given by JAT 2, not JAT 1, provided that trust in my emotion of admiration survives my reflection on my total set of beliefs and emotions.

Fourth, in addition to communal practical ends like building decks and playing music, there are communal epistemic ends. As a member of a community, I have the goal to obtain truth for the community, to add to the community's stock of truths, as well as to increase the community's understanding of the truths we have. We members of a community do our conscientious best in getting the truth in some domain. We pool our resources—experiences, historical memory, data, interpretations, and have a system of authority as well as a division of labor. According to GJA, I can be justified in taking a belief

from my community on the authority accepted by my community provided that I conscientiously judge that the result will survive my conscientious reflection better than a belief I would get on my own.

The process by which I do this needs to be explained and justified. Authority can expand to include groups of persons, sometimes united within a community and sometimes not. This raises two questions: (1) What justifies my community in taking a belief on authority? (2) What justifies *me* in taking a belief that *my community* is justified in believing on authority?

3. COMMUNAL EPISTEMIC AUTHORITY

To answer these questions, let us go back to the way trust expands. In chapter 4 I defended the principle that I have reason to trust those who are conscientiously trusted by those I conscientiously trust. This principle justifies trust in persons beyond those whom I directly judge to be trustworthy. An analogous transfer of authority principle would be the following:

General Expansion of Authority Principle

I have reason to accept as an authority someone who is conscientiously accepted as an authority by persons I conscientiously trust.

I think this principle is false, but the reason it is false is revealing.

Suppose I trust Sam because I conscientiously believe he is conscientious in his beliefs, and suppose he conscientiously takes Sarah as an authority in some domain because he conscientiously judges that Sarah is more likely to get the truth than he is in that domain. I will not always judge that if Sarah is better at getting the truth than Sam is, she is also better at getting the truth than I am. It depends upon how I judge my own ability relative to Sam.

The falsity of the General Expansion of Authority principle is even easier to see if we look at a case in which I trust someone who accepts an authority by JAT 2. Suppose that I conscientiously judge that Sam is conscientious, and Sam takes Sarah as an authority because he conscientiously judges that if he takes a belief from Sarah in some domain, the result will satisfy his own conscientious self-reflection better than a belief he would get on his own. But Sam's conscientious self-reflection might not yield the same result as my conscientious self-reflection, no matter how conscientious I judge that he is. We are two different persons, and what survives my reflection upon my total

set of beliefs, emotions, and experiences is not the same as what survives Sam's reflection upon his total set of psychic states. Authority does not expand by the general form of this principle, and it cannot justify authority over a community of people.

However, there are at least two versions of the Expansion of Authority Principle that justify taking a belief on authority because someone else judges that that person is an authority. One is the following:

Expansion of Authority Principle 1 (EAP 1)

The authority of B's testimony for me is justified by my conscientious judgment that A is more likely to get the truth in some domain than I am, and A judges that B is more likely to get the truth than she is.

Using the mechanism of this principle, we get what Keith Lehrer calls a hierarchy of experts.⁷ The principle works because truth for you is the same as truth for me, and if B is more likely to get the truth than A, and A is more likely to get the truth than I, then B is more likely to get the truth than I. This principle shows how authority for an individual can expand beyond those of whom she has direct knowledge. It results in higher and lower tiers of authority, so there is a built-in structure in the resulting community. But it is unlikely to justify authority in any community in which the authorities do anything more than pass around information. It works as long as the testimony has little impact on the personal experiences, emotions, and prior beliefs of the individual members of the community.

Another version of the principle has the potential to apply to beliefs that are deeper in one's psychic structure than ordinary information, and it avoids the problem that what survives someone else's conscientious reflection may not be the same as what survives my own.

Expansion of Authority Principle 2 (EAP 2)

The authority of B's testimony for me is justified by my conscientious judgment that A's judgment about who the authority is in some domain is more likely to satisfy my conscientious reflection than my own judgment, and A judges that B is an authority in that domain.

7. Lehrer (1977). A hierarchy of experts based on the judgment of the experts on the experts is also mentioned in Hardwig (1985: 341).

Like EAP 1, this principle justifies an expansion of authority for me beyond those persons whom I can judge satisfy the conditions for being an authority. EAP 2 also has the potential to justify beliefs on authority that impact my emotions and personal experiences. If I conscientiously judge that a particular person is better than I am at identifying an authority whose testimony in some domain will survive *my* reflection upon my total set of beliefs and emotions, given the experiences that are unique to me, I should accept her judgment of the authority, and treat the testimony of the authority as authoritative for me. Authority in communities of persons with similar life experiences, background beliefs, and emotion dispositions is justified by this principle. This mechanism permits epistemic authority to spread by chains of interpersonal trust. But I expect that it would be difficult to get a single authority over a large group of people by this mechanism since it is unlikely that there is a large group of people relevantly similar in their psychic structure, but we do get epistemic authority in a strong sense.

There has to be another mechanism, and I think the idea that a community is like an extended self is the key to the way authority operates in communities. The communities I have in mind can be based on shared religious beliefs, shared political beliefs, geographical location, heredity, and many other features. Some of the communities I mentioned earlier are not communities in the sense I mean here. An orchestra or homicide team is not an extended self, although I am not denying that it could be in some circumstances. A rough test of whether a community is an extended self is the way people refer to the community. If they always refer to it in the third person—for example “the United States,” they are not a member of the community, and if everyone refers to it in the third person, it is not a community at all. If they refer to it in the first-person plural, as “we” and “us,” that is an indication that they identify with the community in a way that makes conscientious reflection upon the community’s beliefs very different from reflection upon the beliefs of other persons, and much closer to reflection on their own beliefs.

A community in the sense I mean is an extended self because it has many of the features of a person, and the persons who are its members relate to it in the same *sort* of way they relate to themselves, although in healthy individuals there is never any doubt about the difference between self and community. A community has a communal consciousness with the same components upon which its members can reflect as individuals have when

they engage in self-reflection. A community has a history of experiences; it has communal beliefs; it may have communal emotions expressed and fostered in the community's stories. It often has hopes and plans for the future. It has values. It often acts as an agent. A member of the community will refer to these components of the community consciousness as "our" experiences, beliefs, values, and so on, and its acts as "our" acts.

The beliefs of a community cannot be identified with the beliefs of any one individual member, and the goal of truth for the community is not the same as the goal each member of the community has to get truth for herself. I think this is a point that is not sufficiently recognized.⁸ The community sometimes expresses these beliefs as assertions. We see this in scientific communities ("We teach evolution"), communities of historians ("We know very little about the early life of Charlemagne"), religious bodies ("We believe in one God"), some clubs and organizations ("A Scout is trustworthy, loyal, and helpful"), and whole nations ("We, the people of the United States, believe in the Bill of Rights"). We testify to others, both to members of the community, and to the outside world. Communal testimony within the community informs or reminds the members of what We believe and is analogous to the outcome of self-reflection. Communal testimony to those outside the community is analogous to the testimony of one individual to another.

What We believe may differ from what an individual member of the community believes on some issue, and to the extent that a member of the community rejects a belief We have, she will feel disaffected from her community. If many people within a community differ on an issue of importance, the community may split along lines of a given belief or set of beliefs. This has happened many times in religious communities.

8. John Hardwig (1985) briefly mentions the possibility that what we believe as a community differs from what the individual members of the community believe at the end of that paper. Hardwig's point is about knowledge. He is thinking of a scientific team in which the justification of the belief is divided among the members. Of this possibility Hardwig says, "This community is not reducible to a class of individuals, for no one individual and no one individually knows that *p*. If we take this tack, we could retain the idea that the knower must understand and have evidence for the truth of what he knows, but in doing so we deny that the knower is always an individual or even a class of individuals. This alternative may well point to part of what Peirce may have had in mind when he claimed that the *community* of inquirers is the primary knower and that individual knowledge is derivative" (349).

The justification of the system of acquiring beliefs within the community is its conscientiousness. It is a conscientiously developed system. We think that we will do a better job of getting the truth by following this system with its authority structure than if we use an alternative. Obviously we are not guaranteed truth, but we think it is a more trustworthy system in the long run than alternatives, and we can be conscientious in that belief.

A conscientious person may reasonably see the beliefs of her community as authoritative, justified by analogues of both JAT theses:

Justification of Communal Epistemic Authority 1 (JCEA 1)

The authority of my community is justified for me by my conscientious judgment that I am more likely to believe the truth and avoid falsehood if I believe what We believe than if I try to figure out what to believe in a way that is independent of Us.

Justification of Communal Epistemic Authority 2 (JCEA 2)

The authority of my community is justified for me by my conscientious judgment that if I believe what We believe, the result will survive my conscientious self-reflection better than if I try to figure out what to believe in a way that is independent of Us.

The fact that We believe p can give me a preemptive reason to believe p , but since I am part of the community and accept it as an extended self, the authority that We have is not something alien to me, as the authority of the political state often is. The ultimate authority over me is still myself, and what I take to be the authority is an extension of myself. The attentive reader will see that the analogous point applies to authority over states of the self other than beliefs. A justification of communal authority thesis can justify practical as well as epistemic authority, but this book is primarily focused on authority over beliefs.

What "We" believe is determined in different ways by different communities. A crucial decision that a community needs to make is the structure of authority it will accept. My position is that epistemic authority in a community is justified by the community's conscientious judgment that the community is more likely to get the truth or get beliefs that survive communal reflection if it comes to a belief by the method it chooses than by alternatives. In cases in which a community lasts a very long time, the method itself can be gradually developed by communal reflection, which means it is

determined in part by the authority that has already been developed in the past.

In some communities the members trust the belief of a majority of its members more than a belief of any one person, and such communities will have an authority structure that is democratic. In these communities authority resides in the results of a procedure followed by the entire community rather than in a person or body of persons invested with the authority to protect, develop, and transmit what "We" believe within the current community and to its future members. But other communities have different structures. A community is less likely to adopt a democratic structure when the community lasts for many generations since a democratic structure of authority favors the present over the past. But whatever authority structure a community adopts, it is justified by the communal judgment that We are more likely to believe the truth if we use our authority structure than if we attempt some other method. This is a judgment we make when we are conscientious. It is justified by our communal conscientiousness in the same way my individual judgment justifying my acceptance of authority is justified by my personal conscientiousness.

When I accept the community as an extended self, I acquire reasons to believe what my extended self believes. The justification of the community's beliefs is the community's conscientiousness. The justification of my acceptance of the authority of the community is the fact that my acceptance of my community survives my own critical self-reflection. I can in this way acquire reasons to adopt beliefs on the authority of my community.

Like the second form of the other authority theses, JCEA 2 has a feature that I think is significant because of what it reveals about the way the self is extended temporally. This thesis justifies something I do now because of a judgment I make about my future self. My effort to have a harmonious self, one that resolves dissonance without creating further dissonance, requires a conscientious judgment about what I will be like in the future when I have somewhat different beliefs and emotions, and have had new experiences. JCEA 2 is future-oriented, whereas JCEA 1 is not. JCEA 2 reminds us that the executive self has a dual task in managing its psychic states. It aims to make its states fit their objects, and it aims to make its states fit each other. Ultimately, the test of whether a state fits its objects is that it fits one's other states, not just in the present, but in the future.

The point of authority is to help me believe and act in ways that will withstand self-reflective scrutiny. The test of critical self-reflection might include passing the test of what exemplars say, and since the exemplar can be an authority, the justification of authority in some cases can be partially circular. So the justification of the authority of a particular person might be something that the subject cannot fully grasp without accepting the person's authority. I do not think this is problematic. If an authority shows me how to go about reflection in a way that survives critical self-scrutiny, the authority has helped me attain a higher level of integrity of the self.

4. THE EPISTEMOLOGY OF IMPERFECTION

I think that it is an important part of epistemic humility to regard the search for truth as a communal project. We may have trouble trusting a community because its track record is not perfect, as far as we can tell. But I think we should try not to be like the humans who were outperformed by rats in the experiments mentioned in chapter 5. The animals did better than the humans because the humans were unwilling to go with a rule that they knew was less than perfect even though it was better than anything else they could figure out. We are very unforgiving of authorities when they make mistakes, and may begin to lose trust in the authority even when the authority remains better at getting the truth in some domain than we are, and we are aware of that fact. We have to remind ourselves that in some domains of our epistemic lives we need to accept that errors will be common even when we adopt the best strategy for getting the truth we can find.

When Francis Vincent was commissioner of Major League Baseball, he made an astute observation about the epistemological import of that sport:

Baseball teaches us, or has taught most of us, how to deal with failure. We learn at a very young age that failure is the norm in baseball and, precisely because we have failed, we hold in high regard those who fail less often—those who hit safely in one out of three chances and become star players. I also find it fascinating that baseball, alone in sport, considers errors to be part of the game, part of its rigorous truth.⁹

9. From a speech delivered at Fairfield University, quoted in Kurtz and Ketcham (1992: 245).

In many epistemic domains we can expect to get a better truth/falsehood ratio than .333, but in other domains that is the best the authority can do and I will do much worse than that on my own. There is also the option of refraining from believing anything in such domains, and that is sometimes the better option, but not always. It depends upon which alternative will better satisfy future conscientious self-reflection: nonbelief or belief on authority. Earlier in this chapter I said that one of the reasons we need epistemic exemplars as authorities is that they may know better than we do when a belief is important enough to take the risk of falsehood. What we believe is like what we do in more than one way. One way is that it can be justified by our later judgment that it was worth the risk we took at the time. A belief can be later judged to be worth the risk of falsehood because it turned out to be true and important, but it can also be worth the risk because even though false, it led to other important truths that we would not have gotten without it.

The fact that the authorities in a community are not perfect highlights the fact that getting truth is a corporate project, sometimes involving hundreds of years of effort. The role of members of a community is not limited to identifying the authorities and taking beliefs from them. Most communities grow in knowledge from the distinctive experiences and insights of their members, and the community's verdict on an issue is often the product of communal exchange and deliberation. I do not intend my defense of epistemic authority to be a defense of epistemic passivity. The epistemic demand of a conscientious person to take beliefs on authority needs to be understood in conjunction with an account of how the authority acquires and modifies the community's beliefs. The JCEA principles show that authority for a community resides in the community and it is only because of the community's conscientious reflection that a certain person or body of persons acquires the status of speaking authoritatively for the community. The individual members of the community will reasonably withdraw that judgment if the authoritative body ceases to be a voice that they judge does a better job of getting beliefs that satisfy conscientious reflection than beliefs they get on their own. If that happens, the individual members must find a way to build new communities, given that in most domains we are epistemically helpless on our own.

Moral Authority

1. THE *PRIMA FACIE* CASE FOR MORAL EPISTEMIC AUTHORITY

No domain of epistemic authority is more sensitive than the domain of the moral. People who accept scientific authority without question frequently balk at the idea that they ought to accept authority in any of their moral beliefs.¹ The purpose of this chapter is to examine moral epistemic authority and its limitations. Moral authority can be practical as well as epistemic, but my topic in this chapter is believing on the basis of another person's authority or the authority of a community, not acting at their direction.

Since we can rely upon another person in many different ways, this topic is broader than the topic of moral testimony, although testimony is the simplest way to get a belief from another person. The question, then, is whether it can be reasonable to take a moral belief from another person in the preemptive way that is distinctive of authority, either by testimony, or by relying upon that person's powers in some other way.

My argument up to this point does not distinguish one domain of belief from another.

1. Patrick Hurley (2008) explicitly discusses the use of authority in moral or religious realms as an example of the Appeal to Unqualified Authority (*Ad Verecundiam* fallacy). He writes, "In deciding whether a person is a qualified authority, one should keep two important points in mind. First, the person might be an authority in more than one field. For example, a chemist might also be an authority in biology, or an economist might also be an authority in law. The second point is that there are some areas in which practically no one can be considered an authority. Such areas include politics, morals, and religion. For example, if someone were to argue that abortion is immoral because a certain philosopher or religious leader has said so, the argument would be weak regardless of the authority's qualifications. Many questions in these areas are so hotly contested that there is no conventional wisdom an authority can depend upon" (133).

I argued that I must have a basic trust in my epistemic faculties, and a particular trust in the beliefs I form when I am conscientious. I then argued that consistency requires me to have the same basic trust in the epistemic faculties of all other persons whose general similarity to me I come to believe when I am conscientious. Furthermore, I owe the same particular trust in the beliefs of others whose conscientiousness I conscientiously discover. There is a *prima facie* case for the same conclusion when the issue is the formation of moral beliefs. Perhaps the powers I use when I form beliefs in the moral domain are not the same powers I use in domains outside of moral judgments. For instance, they might include emotion dispositions. That is quite possible, even likely. But whatever they are, they are powers that other persons have as well, and if I am conscientious, I will come to believe that.

So if I believe that I am basically trustworthy in my moral beliefs, then since I am rationally committed to the principle that I should treat like cases alike, I am rationally committed to thinking that other persons are basically trustworthy in their moral beliefs. And if I place particular trust in certain moral beliefs because of my trust in the conscientious exercise of the powers that lead to these beliefs, then if I am consistent, I must trust the moral beliefs of others whose beliefs are produced in the same conscientious way I trust in myself.

The fact that there appears to be more disagreement in moral beliefs than in many other domains will no doubt bother me, and that may lead me to reduce my confidence in the trustworthiness of the faculties and dispositions that lead to moral beliefs. But that gives me no reason to trust myself more than others in general. As long as I see no relevant difference between others and myself in forming moral beliefs, then given that I trust myself, I should trust them.

It follows that we should reject epistemic egoism in the domain of morality. But this conclusion raises a puzzle. Many philosophers regard moral epistemic egoism as permissible, even mandatory, and some philosophers support the extreme version of epistemic egoism about morals. They say that the fact that another person has a certain moral belief never gives me a reason to believe it myself. I should always form my moral beliefs by the direct use of my faculties, never on the authority of another. This position is supportable only if moral beliefs have distinctive features that block the application of my argument against epistemic egoism to the moral domain.

I argued in chapter 4 that there are epistemic exemplars. There is a *prima facie* case for the existence of exemplars in the domain of moral belief. The same point applies to the arguments of chapters 5 and 6 in which I proposed theses on the justification of epistemic authority, each of which can be applied to moral belief and moral testimony. Since my arguments in those chapters do not depend upon the content of the belief, we have a *prima facie* case for the application of the argument to moral beliefs.

The justification of authority theses applied to moral beliefs give us the following:

Justification Thesis 1 for the Authority of Moral Testimony (JAMT 1)

The authority of another person's moral testimony for me is justified by my conscientious judgment that I am more likely to form a true belief and avoid a false belief if I believe what the authority tells me than if I try to figure out what to believe myself.

Justification Thesis 2 for the Authority of Moral Testimony (JAMT 2)

The authority of another person's moral testimony for me is justified by my conscientious judgment that if I believe what the authority tells me, the result will survive my conscientious self-reflection better than if I try to figure out what to believe myself.

The theses on the justification of the authority of moral belief (JAMB) from chapter 5 are parallel. For the most part I will not distinguish the JAMT theses from the JAMB theses since the differences are not relevant for most of my arguments in this chapter. The same point applies to the justification of communal epistemic authority (JCEA) theses of chapter 7, which I will mention briefly at the end of the chapter.

My previous arguments should lead us to accept both JAMT theses unless there are features of moral beliefs that prevent the application of these theses to the moral domain. As long as we aim to make our moral beliefs true and we aim for them to withstand conscientious self-reflection, then these theses will justify us in treating persons who satisfy the conditions given by the theses as authorities and taking moral beliefs from them preemptively. The purpose of the next section is to investigate whether there is something about moral beliefs that makes them exempt from the justification of authority theses.

2. SKEPTICISM ABOUT MORAL AUTHORITY

2.1. Skepticism about Moral Truth

There is a way to block JAMT 1 and my argument against epistemic egoism in the moral domain. JAMT 1 assumes that we aim to make our moral beliefs true, and epistemic conscientiousness is the quality of exercising my faculties in the best way I can in order to satisfy my desire for truth. But that is irrelevant to moral beliefs if moral beliefs do not have truth value, or if truth in the moral domain is constructed rather than discovered.

The issue of moral truth is enormously important, and positions on the way in which our moral beliefs can be true or false vary widely. Notice first that JAMT 1 does not require us to think that moral beliefs have truth value in the same way purely descriptive beliefs have truth value. As long as I aim to make my moral beliefs true in whatever way I think they can be true, I am justified in taking a moral belief on authority when I conscientiously judge that the testimony of someone else will give me a moral belief true in that sense. Of course, some philosophers maintain that moral beliefs do not have truth value in any sense, and so JAMT 1 is inapplicable from their perspective. But fortunately we need not settle the matter because these considerations have no effect on JAMT 2, which does not rely upon the idea that there are moral truths that some people are better at discovering. The moral testimony of certain others preempts my own reasons for and against a belief if I reasonably believe that taking a belief on their testimony will survive my own conscientious self-reflection better than a belief I get on my own. All sides admit that self-reflection is a good thing whether or not it leads to the discovery of truth. A person who is conscientiously self-reflective reflects upon more than those items of her psyche that aim at truth. She reflects upon her moral beliefs along with all her other beliefs, and she reflects upon her emotions, attitudes, and choices. A mental state that survives conscientious self-reflection is one that remains after reflection without dissonance. This aim is not dependent upon the aim of getting truth. It follows that even if someone prefers to leave aside the issue of moral truth, the case for moral epistemic authority justified by JAMT 2 stands.²

2. Although the topic of this book is epistemic authority, this point can be extended to authority over acts since there is a form of conscientiousness in the practical domain. Elizabeth

2.2. Moral Egalitarianism

A second ground for rejecting moral authority and accepting moral epistemic egoism comes from one of the dominant features of modern thought identified in chapter 1: egalitarianism. There are a number of ways in which persons could be morally equal. In one sense of moral equality persons are equal in dignity, where dignity is a value that is either infinite or does not admit of degree.³ So anyone with a dignity is evaluatively equal to anyone else with a dignity. This value arguably grounds the right of persons to equal treatment under the law and a range of civil rights. It might be obvious that equality in this sense is quite a different matter than equality in either the rightness of their behavior or the truth of their moral beliefs, but I think that one of the sources of resistance to recognizing some persons as superior to others in their ability to get moral truth is the thought that that might lead to violating the dignity of some persons. Think of the widely prevailing attitude that moral differences in lifestyles, particularly sexual lifestyles, should not be publicly recognized because that leads to treating some persons as less valuable as persons than others. And if we should not recognize inequality in the morality of lifestyles, then presumably we should not recognize inequality in the ability of people to figure out the morality of lifestyles. There certainly is no *necessary* connection between dignity and the accuracy of one's moral judgment, but the connection is close enough in the consciousness of many people that there is a (probably reasonable) fear that the public

Anscombe (1981: 46) wrote, “[O]nly a foolish person thinks that his own conscience is the last word, so far as he is concerned, about what to do. For just as any reasonable man knows that his memory may sometimes deceive him, any reasonable man knows that what one has conscientiously decided on one may later conscientiously regret.” Someone else may do a better job of getting to a conclusion I would not later conscientiously regret, and this is the case whether the conclusion is a belief or a decision about what to do.

3. Kant seems to say both in the *Groundwork*. In a famous passage Kant says: “[E]verything has either a price or a dignity. What has a price can be replaced by something else as its *equivalent*; what on the other hand is raised above all price and therefore admits of no equivalent has a dignity” (42; emphasis in original). This passage implies two different things. One is that anything with dignity is more valuable than any number of things with a price, no matter how high the price. The other is that things with dignity cannot be compared in value to anything else, not even to other beings with dignity. The value of a person has no equivalent. Kant apparently thought that the two aspects of dignity go together, but in fact they are independent. I discuss this passage briefly and give some reflections on the sense in which persons have a distinct kind of value in Zagzebski (2001b).

recognition of lack of equality in the ability to make moral judgments will lead to the treatment of persons as unequal in dignity.

I will call the thesis that persons are equal in their ability to figure out the moral truth “moral-epistemic egalitarianism.” It should be distinguished from both the thesis that human beings are equal in dignity, and the thesis that human beings are equal in the moral rightness of their behavior or the moral admirability of their character. I am suggesting that moral-epistemic egalitarianism is motivated, at least in part, by the fear that the protection of the public recognition of equality in dignity requires a public stance of moral-epistemic egalitarianism.

I think this fear affects the way philosophers handle the issue of moral-epistemic egalitarianism. In fact, philosophers generally do not handle it at all, although we sometimes hear the related but different claim that there are no moral “experts.”⁴ There is little, if any, argument for this claim either, but the fact that it is thought important to deny the existence of moral experts indicates a fear that some deep values are at stake, and the values are not hard to find. The word “experts” suggests publicly recognized authorities to whom others defer in some domain of human knowledge, and who claim that others should defer to them in this way.⁵ But if an expert can be expected to claim expertise, then if there are experts in the moral domain, they will expect us to believe what they tell us about the moral, and if they can do that, it is but a short step to telling us what to do—or such is the fear. The conclusion is that there had better not be any moral experts.

My support of moral authority on the JAMT theses has a complex connection with moral-epistemic egalitarianism. A weak sense of intellectual egalitarianism follows from the line of argument I have given in this book. In chapter 3 I argued that since human beings all over the world have the same basic epistemic faculties, the fact that somebody somewhere believes *p* gives me a *prima facie* reason to believe *p* myself, and I think that the thesis is unchanged when *p* has moral content. So I am a weak egalitarian in a way

4. Bernard Williams makes this claim in Williams (1995: 205). Even some of the philosophers who are friendly to moral testimony seem hesitant to acknowledge the existence of moral “experts,” for instance, Robert Hopkins (2007).

5. Of course, not all experts claim that others should defer to them, but typically that is because their expertise is not challenged. When it is challenged, they generally do demand public recognition of their epistemic authority. Obvious examples of this are climate change experts and evolutionary biologists.

that has the widest scope—it applies to virtually every human being. However, some people will want to go farther and say that everybody has an equal ability to figure out the truth about all or most moral matters. But notice first that taking a belief on authority justified by one of the JAMT theses does not require rejecting moral-epistemic egalitarianism. I might judge that somebody else is in a better position to get the moral truth in some situation than I am, not because he has better epistemic powers in the moral domain than I have, but because he has more experience or is in a better position to judge. So if I know someone who has worked on a hog farm, I might reasonably think that his judgment on the morality of consuming the products of that farm is more trustworthy than my own judgment even though we are equal in moral faculties. And there are many other such cases.⁶

However, I think that moral-epistemic egalitarianism is false anyway, and a conscientious use of my faculties will lead me to see that it is false. I have already argued that there are moral and epistemic exemplars, and I see no reason to deny that there are epistemic exemplars in the moral domain. Whether that means I support the existence of moral experts depends upon what is meant by an expert. If an expert is somebody publicly recognized as a moral authority, then my position does not require the existence of moral experts. My focus has been on first-person reasons for accepting a moral belief on authority, and my theses do not require publicly recognized experts.

Nonetheless, I think that if I accept someone as a moral authority in some class of cases, given what I argued in chapter 7, it would be very surprising if my judgment was not shared by some other persons, and those persons might be members of a community defined in part by shared judgments on the moral authorities. The recognition is public in the sense that it is openly shared within a community, however small, but of course it need not be public in the sense of possessing widespread recognition within a large population. I do not think the existence of such communities should be worrisome to strong egalitarians, but I will not pursue that further here.

6. Karen Jones (1999: 59–60) gives an example of a man who should take the belief that another man is sexist on the authority of two women friends. His grounds are that the women are more sensitive to this feature than he is even though he is equal to his female friends in moral faculties.

2.3. Autonomy

Egalitarianism is not the only feature of modern thought identified in chapter 1 that poses a problem for moral authority. An even more influential reason for rejecting authority in morals is the idea of autonomy. There may be no incompatibility between intellectual autonomy and epistemic authority in general, but there might be a special problem with taking moral beliefs from another person since autonomy is sometimes thought to require direct control over the moral domain of one's life, including the epistemic part of that domain. For instance, Gerald Dworkin says the idea of autonomy is "the view... that in the moral sphere it is always, finally, up to us to decide what moral principles are valid and how they apply to particular situations" (1988: 49). Nobody would object that her autonomy has been violated if it is not up to her to decide what principles of biology or physics are valid and how they apply to particular situations, so Dworkin must think that moral beliefs have a special relationship to a person that makes the person lacking in something she ought to have if she does not determine what morality requires by the direct use of her faculties.

It is not clear whether there is anything defective about such a person without some details. I do not know of a description by Dworkin, but his intuition seems to be similar to that of Christine Korsgaard, who gives an example based on the character Harriet Smith in Jane Austen's novel *Emma*:

Imagine a person I'll call Harriet, who is, in almost any formal sense you like, an autonomous person. She has a human mind, she is self-conscious, with the normal allotment of the powers of reflection. She is not a slave or an indentured servant. In every formal legal and psychological sense we can think of, what Harriet does is *up to her*. Yet whenever she has to make any of the important decisions and choices of her life, the way that Harriet does that is to try to figure out what Emma thinks she should do, and then that is what she does.

This is autonomous action and yet it is *defective* as autonomous action. Harriet is self-governed and yet she is not, for she allows herself to be governed by Emma. Harriet is heteronomous, not in the sense that her actions are caused by Emma rather than chosen by herself, but in the sense that she allows herself to be governed in her choices by a law outside of herself—by Emma's will. (2009: 162)

I think that Korsgaard is probably right that Harriet is lacking in admirability, but we need more information before we should make that judgment. According to Korsgaard, what is defective about Harriet is that her will is governed by Emma's will, and Korsgaard's Harriet is a person who treats Emma as a practical authority. Since I am primarily interested in epistemic authority, what is relevant for my purpose in this chapter is a case like that described by Korsgaard, but where Harriet takes her moral *beliefs* from Emma. I do not know whether Korsgaard would say that the modified Harriet is defective also, but if she is defective, it cannot be because her will is determined by Emma's will. Rather, her beliefs are determined by Emma's beliefs. Suppose, however, that the fact that they are determined by Emma is due to Harriet's reasoned judgment that Emma is a moral-epistemic exemplar whose authority is justified by one of the JAMT theses. Would there be anything wrong with Harriet in such a case?

Thomas Scanlon links autonomy with moral beliefs in a way that is pertinent to the Harriet case. Scanlon says that a person is morally autonomous if he "cannot accept without independent consideration the judgment of others as to what he should do. If he relies on the judgment of others, he must be prepared to advance independent reasons for thinking their judgment likely to be correct" (1972: 216). Notice that belief justified by JAMT 1 would count as autonomous on Scanlon's definition. In fact, such a belief should be acceptable to the epistemic egoist because Scanlon's condition amounts to a *statement* of standard epistemic egoism in the moral domain. So according to Scanlon, Harriet would not lack autonomy if she defers to Emma based on a reasonable judgment that Emma's belief about what she should do is likely to be correct.

But someone might still insist that Harriet is defective even if she reasonably believes that Emma's judgment is superior. Robert Hopkins argues that a person does not have a *right* to a moral belief unless she herself grasps the moral reasons for the belief.⁷ If so, Harriet is defective because she does not grasp the reasons for her moral beliefs. Of course, Emma could give Harriet the reasons for the belief as well as the belief, and in general, it is common

7. Hopkins (2007) investigates every reason for pessimism about moral testimony he can think of and rejects them all except for this one, which he calls "The Requirement." He does not connect the point with autonomy.

for people who tell others what they think is moral or immoral to also give them the reasons. This is generally the case when the testifier is acting as a teacher. Papal encyclicals always include arguments, usually very extensive. It might not be clear whether a person who reads an encyclical, understands that the reasons support the conclusion, and adopts the conclusion does so on papal authority or because of the reasons. But either way, he grasps the moral reasons for the belief, and so he satisfies Hopkins's restriction.

But what if he does not grasp the reasons? What if someone accepts a conclusion of an argument in an encyclical without even considering the argument? What if Harriet does not grasp the reasons for Emma's belief, but she adopts Emma's beliefs by applying one of the justification principles, say JAMT 1? Does Harriet have a right to her belief, as Scanlon allows by his principle? By hypothesis, Harriet has made a considered judgment that the process of getting her belief from Emma is more likely to get her the moral truth than the process of trying to figure it out on her own. If she lacks a right to the belief, she lacks a right to undertake a process that she reasonably believes gives her the best chance of getting the truth. Possibly she does lack such a right. If so, her right to get the moral truth is trumped by the demand to have moral beliefs that are based on reasons she grasps. If this is the case, it needs to be clear that truth is not the most important aim of moral belief. In fact, one does not have the right to make it the most important aim.

As I have said, some people want to deny that there is moral truth anyway, so they will be unmoved by this argument. But we still have JAMT 2. If I conscientiously judge that I will do a better job of forming a moral belief that will survive my own conscientious self-reflection if I believe what someone else believes than if I try to figure out what to believe myself, I am justified in taking the belief from her on JAMT 2. What can we say about the way Harriet fares on that principle?

I find it difficult to answer this question without considering the fact that different models of the executive self result from different strategies. Contemporary philosophers often prefer the model of the micromanaged self, in which an agent exercises direct executive control over each belief in his web of moral beliefs. Compare that with a model of a self that delegates some of its beliefs to others because of a good management decision that that is the better strategy. A good business executive exercises a wider range of control when he delegates some of the jobs in his sphere of responsibility to others. When he can find someone who is better than he is at some task (or will save

him time), he is reasonable in delegating the task to that person, and he is responsible for the result. A good executive self can operate in the same way. As I see Harriet, what makes her defective is not the lack of a micromanaged self, but the lack of a properly functioning executive self. Rather than treating Emma's beliefs as an executive treats the acts of a delegate, Harriet treats herself as a subordinate. She cannot make an executive decision to take Emma's beliefs on any of the JAMT theses because she has trouble exercising executive control over her self. Her dependency on Emma prevents her from having a strong executive self.

I think this case shows something interesting about epistemic authority. I have stressed the justification for taking beliefs on authority justified by the theses I proposed. Clearly, it takes a high degree of trust in another person in order to apply one of those theses to a particular case. But the case of Harriet shows something else. Taking a belief on authority justified by one of these theses also requires a high degree of trust in one's own executive ability to handle the results of believing what someone else believes on his authority. To justifiably believe on authority, I must make a conscientious judgment that if I believe what someone else believes, suitably qualified, the result will give me a more conscientiously governed self. When I adopt a belief from another person, I am still the one who has to add that belief to my total set of beliefs and other psychic states, and I am the one who must adjust for any dissonance that occurs. Harriet's problem is not her trust in Emma, which may be justified. Her problem is that she cannot apply any of the justification theses to belief on Emma's authority because she does not conscientiously reflect. She is unable to make the judgment that the result of taking Emma's belief on authority will survive conscientious reflection, and she would not know what to do if the result of taking a belief from Emma was psychic dissonance.⁸ The moral of the Harriet case is that one needs a strong

8. Actually, Korsgaard's Harriet is underdescribed. It is possible that she does trust Harriet because she judges that doing so is justified by one of the theses I have endorsed. But it is much more likely that she has no idea whether Emma's belief will be more trustworthy than a belief she can get on her own. An interesting possibility is that she thinks it will never happen that Emma's belief conflicts with some part of her self because she will change any part of herself that conflicts with any belief Emma has on moral matters or matters of personal decision. (Compare the mother who thinks, "Whatever my son does will be okay," not because she thinks that her son will always do what she judges to be okay, but because she will always judge to be okay whatever her son does.) In that case, the standard for her resolution of dissonance within her self is Emma, not the aim of making her psychic states fit their objects.

executive self. I submit that such a self is compatible with taking many beliefs on the authority of others, including moral beliefs.

3. MORAL AUTHORITY AND THE LIMITS OF TESTIMONY

3.1. Emotion and Moral Belief

So far I have focused on acquiring moral beliefs on the testimony of others, and my conclusions have been optimistic. But testimony is not an adequate model for most moral learning, and it has at least two important limitations. A belief acquired by testimony (1) lacks motivational force, and (2) does not give us understanding. The first is a practical defect; the second is epistemic. My position is that we can acquire understanding and motivational force from others, but the process is complex and often involves immersion in a community. Rarely does it occur by testimony.

In contemporary epistemology testimony is typically treated on a simple model of a propositional belief acquired on the say-so of another. Once the believer accepts the proposition, she can then use it the same way she uses any other belief. But the ability to act on a moral belief acquired by testimony is limited by the fact that a moral judgment needs to engage the motivations of the believer if it is going to have any effect on the way she lives her life, and a testifier cannot convey the motive in the straightforward way she can convey the cognitive content of a proposition.

This problem is illustrated by the common situation of people who sit in the pews impassively while the homilist is delivering a fervent sermon on a moral issue. Perhaps he exhorts them to treat the immigrants in their community as they would want to be treated, and in some sense they believe it. But it does not connect to their motivational structure in a way that leads them to act. Anticipating that reaction, the homilist might include an appeal to biblical teaching that a moral belief that does not result in action is empty. The listeners might believe that also, but that is another belief that is not motivating. There is something missing, and if they are aware of that fact, they might also be missing what it takes to do something about their recognition that something is missing.

This is a limitation of judgments taken on testimony. Even though we can believe what another person tells us, testimony cannot give us motivational power. This point is compatible with a number of views on the relation

between moral judgments and affective states. Perhaps a person who makes a moral judgment without the motive to act on it does not *really* believe what she judges. In that case, testimony does not give us what it takes to really believe a moral proposition. Alternatively, a moral belief is detachable from the motive to act on it, in which case testimony can give us the former but not the latter.⁹ Either way, moral beliefs acquired by testimony have limited usefulness to the moral life.

Testimony is an important topic in epistemology because it is the simplest process whereby a valuable epistemic state such as knowledge or justified belief is transmitted from one person to another. If Rob tells Nina that the Albuquerque Balloon Fiesta is the first weekend in October, that may be sufficient information for Nina to make plans. With that model in mind, we imagine Rob saying, "Eating meat is wrong," or "Late term abortion should be illegal," and we imagine that Nina accepts the belief on Rob's authority, justified in the appropriate way. But what happens next? Is Nina going to stop eating meat? Will she engage the political process to stop late term abortions? We need a different model if we want to understand how people can act on the moral beliefs they take on authority.

Testimony is sufficient to engage the motivating emotions of the subject if she already has the appropriate emotion in situations of a certain kind, and then learns by testimony of a particular instance of the kind. So if Nina already has a horror of partial-birth abortion and learns that it is legal in her state, she will be motivated to try to stop it there. Most of us can generalize our emotions to some extent. Our ability to feel sympathy for the plight of a friend can extend to sympathy for the plight of potential friends. Our ability to feel gratitude towards the giver of gifts can extend to gratitude to the country that makes blessings easier to obtain. Our ability to feel joy at the increasing prosperity of our own city can extend to joy at signs of worldwide economic improvement. But emotional generalization has its limits. For most of us, emotions are dramatically weakened when directed towards persons with whom we have no contact except in imagination. In such cases we

9. In other work I have argued that thick affective concepts are components of moral judgments in their ground-level uses, which explains why the judgments have intrinsic motivational power. I then describe a process of thinning the moral judgment, in which the motivating aspect of moral judgments gets detached from the cognitive content. This process explains why we find moral weakness puzzling even though we also recognize that it is common. See Zagzebski (2003a); see also chapters 2 and 3 of Zagzebski (2004a), especially 59–82.

are able to perceive the relevant similarity to situations that elicit an emotion such as sympathy, and we may be able to judge that acting as sympathy would dictate is morally right, but we lack sufficient sympathy to motivate action. I take this familiar phenomenon as evidence that the human conceptual ability to generalize exceeds the human emotional ability to generalize. Testimony can convey the conceptual judgment, and it can convey the relevant similarity to persons or situations that elicit a given emotional response in us, but it is not sufficient to produce the emotional response.¹⁰

One diagnosis of this problem comes from John Henry Newman, who argued that it takes particular images to move a person from assent to an abstract proposition to what he called “real assent” by which we live our lives, and particular images are not shareable in the way we can share abstract generalizations by testimony (1891: chap. 4, sec. 2). This is not to say that we cannot share images. In fact, images are often used to bolster moral judgments, as when photos of earthquake victims accompany pleas for help, or photos of the coffins of soldiers killed in Iraq or Afghanistan were used to support antiwar positions. But in these cases the images function as a reason for the belief, and the source of the images is not treated as an authority. Someone else by showing me the photo might cause me to form the belief that I ought to help the earthquake victims, but the fact that she has the belief is not my reason for the belief. The image in the photo is the reason—or part of it.

Similarly, we can acquire an emotion from another person, and the emotion can be for us a justifying reason for a moral judgment, but the other person has no place in justifying the judgment. The emotion’s role in justifying the judgment is like the role of the image of the earthquake victim in justifying my judgment that I ought to help. In chapter 4 I mentioned a well-known series of studies by Jonathan Haidt (Haidt, Koller, and Dias 1993) that shows that people use their emotion of disgust as a justification of moral judgment, for instance, the moral condemnation of a man who has sex with a chicken carcass and then cooks it and eats it for dinner. A person who tells me a story about the man might cause me to feel disgust, and my disgust can be my reason for moral disapproval of the man. My reason for the judgment does not refer to anything about the person who told me the story even

10. The human difficulty in generalizing the scope of an emotion is a good thing when the emotion is negative. For instance, it is fortunate if people can envy only a few people of their acquaintance.

though he caused my emotion of disgust and might have caused it intentionally. In this way I can get a moral belief from another person, but the fact that he has a certain belief is not a reason for my belief, and I am not believing on his authority.

There is room for authority, however, once I subject my emotion to critical self-reflection. I argued in chapter 4 that emotions may or may not fit their objects, and a conscientious person reflects upon her emotions and makes judgments about their reasonableness. In this way emotions are like beliefs and unlike images.¹¹ Sometimes we can justify an emotion by the direct use of our faculties, just as we often do with beliefs. I have ways to tell whether something that disgusts me is disgusting, whether something I fear is fearsome, whether something that makes me angry deserves that emotion, and so on. But we also justify our emotions indirectly, by the emotions of others whose emotion dispositions we trust. I can ask myself whether an emotion I acquired from another person satisfies conscientious self-reflection better than an emotion acquired independently. Sometimes the answer to that question is yes.

An example of an emotion that is frequently spread from person to person is the emotion that goes with seeing something as tasteless, or what is colloquially called "tacky." I do not know the name of the emotion, but it seems to me to be a distinctive type of disdain. Suppose I acquire that emotion from someone else on some occasion, and suppose that it survives future critical self-reflection. Let us also suppose that the experience of getting an emotion of that type from the other person is repeated in other circumstances. At some point I come to trust her emotion disposition in this domain, and it may be reasonable to do so. I can reasonably judge that her emotion in this domain is more likely to satisfy my own conscientious self-reflection than an emotion I would acquire independently. In my argument against emotional egoism in chapter 4, I said that the fact that someone whose emotion dispositions I trust has a given emotion gives me a *prima facie* reason to have the same emotion. I think that the characteristic feeling that goes with seeing something as tasteless is an example of an emotion that can be justified in this way. The justification is parallel to the justification of taking a belief on authority.

11. I am not denying that photographs can be faked, but we do not criticize the appropriateness of the relation between the image and reality the way we criticize the appropriateness of an emotion and the situation construed a certain way in an emotional state.

The emotion that accompanies seeing something as tacky is not a moral emotion, but I think the same pattern of justification applies to many emotions that are distinctively moral, such as scorn, indignation, compassion, and certain forms of disgust, among others. If I reasonably trust someone's disposition to feel compassion, his compassion for a terrorist gives me a *prima facie* reason to feel compassion even if I find the emotion surprising or puzzling. If I acquire compassion from the person I trust in this instance, my emotion can be justified by my conscientious judgment that an emotion acquired from him will satisfy conscientious self-reflection better than an emotion I would have had independently. If the emotion is the ground for moral judgment—say, the judgment that the terrorist should not be abused in prison—then I am not taking a moral judgment on authority, but the ground of my judgment is an emotion justified in a way that parallels the justification of the authority of belief. The function of authority in cases of this kind is subtle and indirect, but it is significant. I do not take a belief on authority; I take an emotion on authority, and the emotion is the ground for my moral belief.

This process does not presuppose that emotions can be acquired voluntarily. I am assuming only that we do get emotions from others, and when we do, they can be justified. The transmission of emotions is one of the most important functions of communities.¹² Families share patterns of emotion dispositions, some of which are grounds for moral judgments. Religious communities and political groups transmit moral beliefs to their members in part through the transmission of emotions. These communities also transmit moral beliefs by testimony, but since testimony does not explain how the beliefs succeed in guiding our moral lives, the transmission of emotions is a critical part of effective moral authority. Emotions are crucial for acting morally, and to the extent that we acquire those emotions from others, those other persons are crucial for our moral acts.

3.2. Moral Authority and Understanding

There is a second limitation of getting a moral belief on testimony. If a person tells me that *p* and leads me to believe *p*, I might be justified in believing *p*

12. For a discussion of this point, see Hills et al. (2010).

and can even know that p , but that process cannot give me understanding of p . I assume that understanding includes grasping relations between one fact and another. I not only grasp *that* the car will not start, but I see *why* it will not start. Someone else can tell me that the car will not start. She can even tell me that it will not start because the battery is dead. But she cannot make me see the relation between those two facts by testimony.

The relation between one fact and another need not be causal. It also takes understanding to see logical relations, and here too only the recipient can make the link. If you are teaching *modus tollens*, you can explain what the rule is, and you can give numerous examples, but if the student does not get it, there is nothing you can do. This does not mean that a good teacher cannot aid a student's understanding. On the contrary, giving students understanding is an important pedagogical accomplishment, and the teacher should often be given the major part of the credit for producing understanding in the student. But there is still a step that the students have to do themselves, a step that goes beyond grasping what is said by the teacher and accepting it.

The same point applies to understanding in the moral domain. Moral understanding includes seeing the connection between moral reasons and moral judgments, and perhaps also the connection between certain emotions and moral judgments. Understanding permits us to see how to extend a moral judgment to different situations, and to see how distinct moral judgments relate to each other, perhaps because they are grounded in the same general value or principle. Testimony cannot give us this ability.

The idea that we cannot get moral understanding from testimony goes back to Plato, where an important concern for Socrates is the issue whether virtue can be taught. It appears that his answer is yes, but not by testimony. Socrates says in the *Apology* that when his *daimonion* tells him not to do something, he does not do it, but he denies that he thereby has *epistēmē*. What is interesting about this case is that Socrates' *daimonion* is a voice in his head, but he treats it like the authority of another person.¹³ Even though

13. Socrates says his *daimonion* tells him to stop doing what he is about to do, but it does not tell him what to do (*Apology* 31d2–4). However, this does not seem to be the reason the *daimonion* does not give him *epistēmē*. The issue is whether he has *epistēmē* of the fact that he should not do what the *daimonion* says he should not do, and Socrates says he does not, even though the *daimonion* is of divine origin and is presumably reliable.

Socrates thinks the *daimonion* is divine in origin, believing on its authority is not good enough for *epistēmē*, the kind of knowledge that gives understanding.¹⁴ But that is not to say that one cannot get *epistēmē* from other persons. On the contrary, as I mentioned in chapter 1, Plato claims that one should find a person with *epistēmē* and learn from him. Similarly, one could get moral *epistēmē* from another person if one could find a person who has it. If one could surmount the difficulty of finding such people, there are ways one could get moral understanding from them, but we should not expect to get it by testimony.

A version of Plato's objection to moral testimony has appeared in the recent literature in Alison Hills's (2009) argument that virtue requires understanding, and the failure to produce understanding is a significant limitation of moral testimony. Hills says that we do not act virtuously unless we act for the right reasons, and acting for the right reasons requires understanding, not just having the correct moral beliefs. But we cannot get understanding on testimony. So according to Hills, testimony can give us the correct moral beliefs, but that is insufficient to give us what is valuable in virtuous action.

I agree with Hills that we cannot get understanding from simple testimony, but I think that our investigation in this book shows that we need not be pessimists about moral authority, even for the goal of acting virtuously. Other persons help us get moral understanding by teaching, presenting us with detailed reasons, drawing analogies, telling stories, and showing us how to make connections between one moral situation and another, or between one moral judgment and another. More is demanded of the recipient when the goal is understanding than when it is true belief, but more is demanded of the authority also. Moral understanding is often imbedded in traditions, and it takes a community to transmit a tradition. Moral teaching within a community can be formalized, as it is in the teaching office of the Catholic Church, or it can be informally transmitted by immersion in a way of life, as it is for the Old Order Amish. More commonly, moral understanding—or the illusion of understanding, arises gradually through networks of people who trust each other in a certain domain because they are already like-minded in that domain. That is, they already share a number of moral beliefs,

14. I believe that Greek scholars agree that "*epistēmē*" in Plato is not knowledge in the modern sense. It is closer to understanding.

and they get understanding through extended exchanges, which in the old days took place face-to-face, but nowadays often occur in online forums and blogs.

This leads to the issue of whether understanding, like beliefs and emotions, needs to be justified. At the beginning of chapter 4 I said that I think that the problem of fit arises for any psychic state that aims at an object and can fail to reach it when the subject is unaware of the failure. A form of the skeptical problem arises for each such state. So there is a form of emotion skepticism as well as belief skepticism. What about understanding? Can we be in a state that seems to us to be understanding, but which is not, and which is such that we cannot detect that it is not? If so, we need some means to justify thinking we have understanding when we think we have it, and if there is a gap between such justification and a guarantee of understanding, we need basic trust in the powers that produce understanding, parallel to the need for basic trust in our belief-forming faculties and emotion dispositions. The possibility of illusory understanding also means that we would need to be justified in getting what seems to be understanding from another person just as we need to be justified in getting beliefs from another person. On the other hand, if understanding, like pain, is incorrigible, we would not need a means to justify us in thinking that what we take to be understanding really is understanding, whether we got it from someone else or by the direct use of our faculties. We would not need basic trust in the powers that produce understanding, and we would not need principles to justify us in getting understanding from another person.

Fortunately, we have the resources to go either way on this issue. It is interesting that skepticism about understanding gets little attention. Perhaps the reason we do not worry about whether our understanding is illusory is that understanding, unlike true belief, is *partially* transparent to the mind. The "Aha" experience is phenomenologically as trustworthy as the sense of certainty. It is transparent in that when we have it, we know we have it. It is tempting to add that it is incorrigible: When we believe we have it, we do have it. But the incorrigibility of understanding is much less plausible than its transparency. It is not hard to think of cases in which we later judge that we were mistaken in thinking we had understanding. For instance, at one time I thought I understood Husserl, but later decided I did not. There are no doubt many things I think I understand now, but do not. I may discover some of them in the future, but I have no guarantee that I will discover all of them. So illusory understanding does seem possible. If so, when I acquire a

state that appears to be understanding from another person, I will need a way to justify it, and I think it can be justified the same way I have proposed we justify any of our states: It is justified by my conscientious judgment that the state will survive conscientious self-reflection.

Other people can give us both conscientiously acquired moral beliefs and moral understanding. Although the process by which we get understanding differs from the process by which we get beliefs, both are forms of relying on someone we trust. If I believe what someone else believes because she believes it, I am imitating her. If I get understanding through a process of adopting her way of relating one belief to another, I am imitating her in a different way. In both cases the process is justified by what I conscientiously predict will survive future conscientious reflection, and when it does survive, the authority has helped me in acquiring a conscientiously governed self.

4. COMMUNAL MORAL AUTHORITY AND CONSCIENCE

The transmission of moral beliefs is an important function of many communities. I am justified in taking a moral belief from my community by the JCEA theses I defended in chapter 7. Communities are equally important as vehicles of moral discovery. Attaining moral truth is usually a corporate achievement. As Alasdair MacIntyre has reminded us, in practical deliberation we are liable to errors unless corrected by others: “[We] need the ruthless correction of our judgments by others who can see in us what we cannot see in ourselves, and that is why deliberation not conducted in the company of such others is deliberation on which we would be unwise to rely. We should always therefore treat solitary deliberation as peculiarly liable to error” (2009: 16). MacIntyre goes on to say that other people can be the source of corruption as well as correction, and therefore the community needs norms of objectivity. But the salient point is that regardless of the limitations of moral testimony or communal moral inquiry, we are often more likely to get moral truth and beliefs that satisfy conscientious self-reflection if we collaborate with a community rather than attempt it alone.

Moral deliberation is more likely to be accurate if many members of a community participate in it. Different individuals make different mistakes, and when deliberation is conducted together, members of the community are checks on each other. I am not suggesting that a reasonable community

should conduct its moral inquiries by a democratic process. There are many mistakes in moral deliberation to which all but a few people are prone because most personal vices and mistakes in reasoning are common, and cultural forces that are detrimental to the discovery of moral truth are apt to be unnoticed by the majority. But it is also true that most individuals have insights that most others do not, and the authority structure of a well-functioning system should include sympathetic consultation with the members of the community for viewpoints that reveal those insights, regardless of whether the final verdict on an issue is determined by a few people or a great many.

When I reasonably judge that what my community comes to believe is more likely to satisfy my conscientious self-reflection than what I would come to believe by independent inquiry, it is reasonable for me to believe what my community comes to believe in this way preemptively. But this conclusion will be resisted from a different direction than the ones that have their roots in modern philosophy. It is a traditional teaching of Christianity that a person has a right to follow his own conscience. In fact, a person is obliged to follow his conscience and can never be obligated to act against it. The notion of conscience is interesting because it has often been treated as a distinct source of moral knowledge, like Socrates' *daimonion*, a voice within him that prevented him from acting wrongly. But unlike Socrates' *daimonion*, conscience is supposed to give a moral verdict on what one should do on a particular occasion. Conscience therefore does not deliver general moral judgments, but particular judgments that apply to the agent.

Conscience can be opposed to authority as the latter is judged by the agent only if conscience is a power by which I judge independently what is right and wrong in some case. In that sense of conscience, it can, of course, be opposed to the judgment of a community that I judge to be more likely than myself to get the moral truth. But that cannot be the sense of conscience that has final authority.¹⁵

15. In *Veritatis Splendor*, Pope John Paul II objects to the view of conscience that gives the individual absolute authority. He says of that view: "The individual conscience is accorded a status of a supreme tribunal of moral judgement, which hands down categorical and infallible decisions about good and evil. To the affirmation that one has a duty to follow one's conscience is unduly added the affirmation that one's moral judgment is true merely by the fact that it has its origin in the conscience. But in this way the inescapable claims of truth disappear, yielding in their place to a criterion of sincerity, authenticity and 'being at peace with oneself,' so much so that some have come to adopt a radically subjectivist conception of moral judgment" (sec. 32).

In Aquinas's treatment of conscience, it is not a power, but an act. It is the practical judgment by which we apply knowledge to some action (*ST I*, q. 79, a. 13, corpus). The dictate of conscience is our best moral judgment, all things considered. We ought to follow conscience in that sense, but conscience in that sense cannot conflict with our judgment that the authority is more likely to be right than we are independent of the authority since in that case the dictate of conscience would be to follow the authority.

However, if conscience is our best moral judgment in some case, conscience can conflict with the dictate of an authority whom we *previously* judged to satisfy the JCEA thesis. Our judgment that a putative authority is an authority justified by one of the justification theses (JAMT or JCEA) is revisable. It can happen that someone whom we take to be a moral authority makes a moral judgment that leads us to withdraw or modify our previous judgment that the authority is more likely to get the moral truth in some domain than we are, and hence, to withdraw or modify our previous judgment that the putative authority is an authority. Making such a judgment is significant and fraught with the dangers of self-deception, wishful thinking, and other vices. But it can sometimes be reasonable to make such a judgment. Notice that this is not a case in which conscience conflicts with the judgment of someone the agent takes to be authoritative. As long as conscience is identified with one's best moral judgment, one's conscience never conflicts with the verdict of someone whom one takes upon reflection to be an authority justified by one of the justification theses.

Moral authority exists because both epistemic and emotional egoism are incoherent. Moral communities aid their members in satisfying their own ends. These communities often are not limited to the moral domain. That is, they are communities whose ends are not limited to acquiring moral truth. Often moral communities are also something else—political communities or, especially, religious communities. Moral authority can overlap religious authority, the topic of the next chapter.

Religious Authority

1. RELIGIOUS EPISTEMIC EGOISM

The domain of religious belief is an especially interesting one for the arguments of this book since religious authority can have a large and complex structure with a vast reach in time and space, sometimes including authority over acts as well as beliefs. But religion also has features that lead some people to reject authority and to accept some form of egoism about religious belief. In chapter 8 we looked at skepticism about authority in the moral domain. Many philosophers have thought that the value of autonomy makes epistemic egoism permissible, even mandatory in the domain of moral beliefs. While religious beliefs do not have the connection with autonomy that moral beliefs allegedly have, they share other features with beliefs about morality that lead some philosophers to deny the existence of religious authority. The most obvious feature is widespread disagreement about religious matters. In the passage I quoted in chapter 8 (note 1), Patrick Hurley cites this as a reason why there are no religious authorities. Hurley says, “[T]here are some areas in which practically no one can be considered an authority. Such areas include politics, morals, and religion. For example, if someone were to argue that abortion is immoral because a certain philosopher or religious leader has said so, the argument would be weak regardless of the authority’s qualifications. Many questions in these areas are so hotly contested that there is no conventional wisdom an authority can depend upon.” Hurley rejects authority on grounds of disagreement, the lack of “conventional wisdom.” We will discuss the conscientious response to disagreement in chapter 10, but there are some implied assumptions in Hurley’s claim that bear on the arguments I will give in the next two sections.

Notice that if widespread disagreement between persons of unknown reliability and between such persons and oneself affects the reasonableness of a belief, epistemic egoism in that domain is false. Religious epistemic egoism is false on this line of reasoning. Hurley implies further that the numbers count. There would be no point in his mentioning that *many* people disagree with many other people, unless the credibility of a belief increases or decreases with the numbers. There is *prima facie* reason to count as credible both the many who believe one thing and the many who believe another because there are many of each. So to take widespread religious disagreement as a problem is to assume the falsity of the egoist position, and it assumes some credence in common consent arguments. However, religious epistemic egoism is so common, it requires special attention. Let us begin by looking at how the argument of chapter 3 against epistemic egoism applies to the religious domain.

The extreme epistemic egoist says that the fact that another person has a belief never gives me a reason to believe it. If somebody wants to convince me to believe *p*, he needs to demonstrate to me that *p* follows from my own beliefs and the rules of reasoning I accept. Extreme epistemic egoism applied to religious belief would be the following:

Extreme religious epistemic egoism

The fact that another person has a certain religious belief never gives me a reason to believe it. I should not believe that God exists unless I have a demonstration of the existence of God that uses premises I accept myself, and I should accept the beliefs of a particular religion only if the same conditions can be satisfied for each doctrine of that religion.

It is very doubtful that any religion can satisfy these conditions. Theism satisfies these conditions for some extreme religious egoists and atheism for others, whereas still others will be convinced neither by arguments for atheism nor by arguments for theism and will become agnostic. Examples of extreme religious egoists are legion, especially among contemporary philosophers.

The weaker *standard epistemic egoist* says that the fact that another person has a certain belief gives me a reason to believe it, but only when I have evidence that the other person is reliable. When we apply standard epistemic egoism to religious belief we get the following:

Standard religious epistemic egoism

The fact that someone else has a certain religious belief gives me a reason to believe it, but only if I have evidence of the reliability of the source.

I interpret John Locke as a standard epistemic egoist about belief in Christianity. Locke defines revelation as a communication from God, and faith as the acceptance of beliefs on the word of God. Locke says we have good evidence that the Gospels are a communication from God, given the miracles performed by Jesus, and he refers to the miracles of the prophets as confirmation that their message was from God.¹ According to Locke, belief in particular Christian teachings does not require demonstration of the content of the revelation, as it does for the extreme egoist. But reason judges whether something is a revelation, that is, that it is reliable. And Locke allows that it can be rational to believe a revelation even when the content is improbable. He says we would expect that revelations would be about matters we would independently judge to be improbable since revelation is about matters beyond the limit of our faculties to attain on our own, for instance, that the angels rebelled against God.²

It seems to me that the argument I gave against epistemic egoism in chapter 3 applies to religious epistemic egoism. I assume that the natural desire for truth extends to a desire for truth about religious matters. Is there such a desire? Of course, it depends upon what questions a person wants answered. Presumably truths about religious matters would be among the most important truths we could get. If there is a God, that would be an important thing to know, and if there is no God, that would be important to know also. The same point applies to many other questions that religions attempt to answer, such as questions about the origin of the universe and an afterlife. Many of us think it is desirable and possible to get answers to these questions. Of course, someone might think these answers are not desirable or not attainable, or they might think that even though they are *prima facie* attainable, reasonable disagreement in religion is a defeater for the belief in their

1. See Locke, "A Discourse of Miracles" (1709).

2. "Of Faith and Reason, and Their Distinct Provinces," in *Essay*, Bk. IV, chap. 18, sec. 7. In "A Discourse of Miracles," Locke similarly argues, "it cannot be expected that God should send any one into the world on purpose to inform men of things... that are knowable by the use of their natural faculties. This would be to lessen the dignity of his majesty in favour of our sloth, and in prejudice to our reason."

attainability. But it is also the case that if they conscientiously believe that there are many people who conscientiously believe that getting the answers to these questions is desirable and possible, they have a *prima facie* reason to think so also.

If we think that the desire for truth about religious matters is satisfiable, we must also trust that our faculties are adequate to the task. Granted, it is unlikely that our faculties are as well suited to figuring out the truth about religious matters as they are for figuring out the truth about the physical world, but greater desirability compensates for lesser attainability. Aristotle remarks that there is greater satisfaction in a lesser grasp of “celestial things” than a complete grasp of mundane matters, “just as a half glimpse of persons we love is more delightful than an accurate view of other things” (*Parts of Animals* 644b32–35). For Aristotle, the desirability of the grasp of celestial things is great enough that it makes even a lesser degree of knowledge or understanding of those things a good worth striving for. I see no reason to reject the assumption that religious truths are among the truths we desire, and that we believe we have some level of attainability of those truths. With these assumptions, the argument of chapter 3 against epistemic egoism applies to religious epistemic egoism. When I am conscientious I will come to believe that many other people have the same desire for religious truth and the same general faculties I have for attaining it. I have argued that the general trust I must have in my faculties in advance of evidence of their reliability should be extended to all others whose similarity to myself I discover when I am conscientious. There is no reason to think that people with beliefs about religious matters (either for or against) are using special faculties that I do not share. If there *are* people with a special religious faculty or ability, that is all the more reason to reject religious epistemic egoism. But absent a reason to think that *I* am the one with a special religious faculty denied to others, I should accord other persons the same trust in the domain of religious beliefs that I accord myself.

2. RELIGIOUS EPISTEMIC UNIVERSALISM

I have argued that the principle of epistemic trust in others commits me to two levels of trust in others: a general trust in their faculties because they have the same faculties I have, and a particular self-reflective trust in those

others whose conscientiousness I discover by being conscientious. Let us focus on the general trust. Given that I must have a presumption in favor of my faculties without evidence, then if I believe in a way that I trust that other persons have the same general faculties that I have, I should accord them the same general presumption in favor of their faculties. That means there is an epistemic presumption in favor of the veridicality of the deliverances of the faculties of other persons until shown otherwise by a further use of my faculties. That is what I have called epistemic universalism. When applied to religious belief, this commits me to the following principle:

Religious epistemic universalism

The fact that another person has a certain religious belief gives me a *prima facie* reason to believe it.

In chapter 3 I argued that epistemic universalism supports the credibility of belief based on common consent. The most historically important example of a common consent argument is the *consensus gentium* argument for belief in God. In classical philosophy we find this argument in Seneca's *Ad Lucilium Epistulae Morales* and in Cicero's dialogue *On the Nature of the Gods*. Both Seneca and Cicero thought belief in the gods was universal. Today it is not universal, and I do not know if it ever was, but it is certainly widespread. In any case, the argument from consent does not require universality of belief. Lack of universality is a problem only to the extent that there are not only large numbers of people who believe in God, but there are also many people who disbelieve. Clearly, if the number of people who believe in some proposition counts, so does the number of people who disbelieve that same proposition.

As I mentioned in chapter 3, the issue of independence is relevant to common consent arguments. If millions of people believe in God because they all acquired the belief by testimony from a small number of sources, the vast number of believers does not count as much as the same number of beliefs acquired independently. Most people who believe in God come to do so in the early part of their lives by testimony from their parents and other trusted adults. It is doubtful that the beliefs children get from adults have much more credibility than the beliefs of the adults alone, so we probably should discount the number of children's beliefs as relevant to the *consensus gentium* argument. However, I would not discount the number of theistic

beliefs of adults who originally acquired their belief on testimony since there is a defeasible reason to trust their trust in the lack of defeaters for their belief. Furthermore, there are a great many beliefs in God that *were* independently acquired in widely separated communities with little if any influence from each other. I think we should conclude that the epistemic presumption is in favor of the belief. The fact that another person believes in God gives each of us a defeasible reason to believe. The fact that many millions believe increases the reason, and the fact that many of those millions acquired their belief independently increases the reason further. It is an implication of self-trust that the fact that so many people all over the world at all times have believed in a deity gives each of us a *prima facie* reason to believe in a deity, a reason that exceeds the reason we would have for believing in God if we were aware of only one or a few believers. For the same reason, the number of disbelievers counts in favor of disbelief.

Given what I have said, the fact that consistent self-trust leads to trust in others supports a form of the traditional *consensus gentium* argument for theism that we can formulate as follows:

***Consensus gentium* argument from self-trust**

1. I must have a general attitude of self-trust in my epistemic faculties as a whole. This trust is both natural and shown to be rational by philosophical reflection.
2. The general attitude of epistemic self-trust commits me to a general attitude of epistemic trust in the faculties of all other human beings.
3. So the fact that someone else has a belief gives me a *prima facie* reason to believe it myself.
4. Other things equal, the fact that many people have a certain belief increases my *prima facie* reason to believe it, and the reason is stronger when the beliefs are acquired independently.
5. The fact that other people believe in God is a *prima facie* reason to believe that God exists, and the fact that many millions of people constituting a very high majority believe or have believed in prior ages that God exists increases my *prima facie* reason to believe in God myself. Even though many of those beliefs were acquired from others, I have reason to trust the beliefs of trusted others who acquired their beliefs from persons they trust. That is, I have reason to trust their trust. Furthermore, even if we completely discount for dependence, there are still many millions of people who independently believe or have believed in past ages in the existence of God.

A parallel argument can be given for atheism since the fact that there are many people who disbelieve in God gives me a *prima facie* reason to disbelieve in God. If there are many more people who believe than disbelieve under the conditions I have described, the *prima facie* reason for belief is stronger than for disbelief.

Since the argument I have given here is a special case of the common consent argument I defended in chapter 3, it differs from the traditional *consensus gentium* argument in that it is not an argument to the best explanation. In traditional forms of that argument, there is a certain datum that needs to be explained: the fact that so many people in so many parts of the world believe in God. The issue is whether the best explanation for that datum is the truth of the belief. Various explanations of the datum are then weighed against each other from the standpoint of a neutral arbiter. No connection is assumed between the arbiter's trust in the way she goes about evaluating alternative explanations for the datum and the arbiter's trust in the persons evaluated. The arbiter could just as well be a person from another planet. In contrast, the form of argument I support links trust in the beliefs of others with self-trust. It is because of trust in the self that we are committed to granting *prima facie* credibility to the belief of another, and the credibility is greater when the belief is widespread.

The reason for belief in God that we get from the beliefs of other people can be defeated. In particular, it can be defeated by the discovery of lack of conscientiousness on the part of a believer, or by the discovery that they have inferior faculties. Is there any reason to think that belief in God is over-represented among those who are the least conscientious or have inferior faculties? Hume notoriously claimed that belief in miracles is more common among "ignorant and barbarous" peoples (*Enquiry*, Sec. X, Pt. II, p. 94). Similarly, some intellectually sophisticated atheists claim that atheism is dominant among intellectually sophisticated persons in Western countries, and some have argued that religious belief is negatively correlated with intelligence.³ So they must think that when they are conscientious, they can identify a property of a believer that makes a belief formed out of that property

3. See, for example, Richard Dawkins (2008: 123–30). Richard Lynn, John Harvey, and Helmuth Nyborg (2009) recently argued that, in the United States and Europe, there is a "negative relationship between intelligence and religious belief" (11), noting, however, that "the United States is anomalous in having an unusually low percentage of its population disbelieving in God (10.5 percent) for a high IQ country" (14).

trustworthy—intelligence or intellectual sophistication, and that many theists in history lack that property. If so, they must trust the truth-conduciveness of that property more than the truth-conduciveness of the conscientiousness of others.

The objector who says that theism is correlated with lack of intelligence or sophistication is offering a theoretical reason for rejecting the *consensus gentium* argument, and he is no doubt interpreting that argument as proposing a theoretical case for theism, but the form of the argument I have given is based on deliberative reasons. A person who engages in conscientious self-reflection begins with a general trust in all other human beings, and trust in a vast amount of personal experience, trust in particular persons, and trust in many beliefs that have already satisfied past reflection in an attempt to make her states fit both their objects and each other. Part of what I am arguing in this book is that the project of self-reflection reveals that other persons can help us in that project, and that all persons form a community in the common project of attempting to make our states satisfy conscientious reflection. What we share with all other persons is significant and should be taken very seriously. But it is not surprising that given that other persons have had different experiences and have engaged in their own project of self-reflection, the common conceptual content of theistic belief is exceedingly vague. Common consent cannot support specific theological claims about God, not even the unicity of God, although that attribute has wider acceptance than such traditional attributes as personhood, omnipotence, and perfect goodness.⁴ What we get from widespread agreement is closer to the “half glimpse” mentioned by Aristotle than to a clear view of the object. The peoples of the world give different descriptions of the object glimpsed, but they agree that there is a glimpse of something important. Perhaps the vagueness of the idea of God makes the *consensus gentium* argument uninteresting to some people, but I agree with Aristotle that even a half glimpse of the profound is more valuable than a complete view of lesser things.

4. Notice, however, that if “God” is a proper name, it is not necessary that there be a common descriptive content in the beliefs of theists in order for them to share a common belief.

3. BELIEVING DIVINE TESTIMONY

3.1. Faith and Believing Persons

Epistemic egoists will not take a belief from another person without a theoretical case for the belief—either a direct case, or a case for the reliability of the source. The extreme egoist expects the former. The standard egoist will take the latter. Taking a belief from Scripture or from a religious tradition would not be justified at all according to the extreme egoist. That is to say, the belief would not be justified because it is from Scripture or tradition. If the belief is justified, it is because it is justified by the use of my faculties anyway. The fact that the same belief is included in Scripture or the tradition is irrelevant. The standard egoist may take a belief from Scripture, tradition, or an authority, but only if she can determine what was said by the source and weigh the evidence that the source is reliable in the relevant domain, and the evidence must be independent of the testimony of the source.

Locke thought that faith can be defined in a way that is compatible with his epistemic egoism and the evidence model of testimony. As I mentioned above, Locke thinks of faith as assent to something on the testimony of God, where the evidence of miracles is evidence that the source is divine and hence reliable. I think Locke is right that faith is tied to belief on testimony, and the rationality of faith is therefore tied to the rationality of belief on testimony, but I think that the evidence model of testimony Locke uses prevents him from explaining the key component of faith.

Suppose I believe that there is sufficient evidence for a proposition contained in Scripture because I judge that the source is reliable. I may be rational in believing the content of the proposition, but the identity of the source is an accidental feature of my reason for believing. The source need not be God. It could just as well be a reliable devil. If your aim is to get information, that is okay, but I assume that faith is not limited to receiving information. No set of theoretical reasons will ever be enough for religious faith since faith is more than a cognitive state. The relationship of trust between a person and God explains both the nonepistemic component of religious faith, and the difference between rational belief on faith and irrational belief. A believer can have good deliberative reasons to believe the teachings of a revealed religion whether or not she has sufficient theoretical reasons.

On the trust model of testimony I defended in chapter 6, testimony is a request for trust that the recipient may accept. The rationality of the acceptance of trust requires deliberative reasons since theoretical reasons are neither necessary nor sufficient for the rationality of trust. The trust itself provides the believer with a deliberative reason to believe the testimony.⁵ Since trust is partly a nonepistemic state, it follows that states that are partly nonepistemic can give a person epistemic reasons for a belief. The way trust operates as an epistemic reason is not just one of its interesting features; I believe it is crucial to understanding how faith can be rational.

If God tells me that p , God takes responsibility for the truth of p for me and for all other intended recipients of his revelation. God intends that I believe him, and he acknowledges that we who are the recipients place epistemic trust in him by believing him. On this model of faith, then, religious faith is believing God, as Elizabeth Anscombe (1979) argued long before the dispute between Moran's assurance model and the evidence model of testimony came into the literature. My position is that the ground of faith is trust in God, which gives me a deliberative, first-person reason to believe what God tells me. As in ordinary cases of testimony described by the trust model, the testifier is responsible for the justification of the content of the belief, and the recipient is responsible for having appropriate trust in accepting the testimony. Although faith is a gift, I suggest that it is justified the same way any other psychic state is justified—by survival of conscientious self-reflection.

There are a number of different ways trust can be an appropriate deliberative reason to believe testimony from God. For some people, trust is grounded in other deliberative reasons such as their religious experience or the admiration they have for the Scriptural message. For others, trust is grounded in the conscientious judgment that if they trust, they will get more important true beliefs in the relevant domain than figuring out what and

5. The Polish logician Joseph Bochenski analyzes different schemata for the justification of religious beliefs in *The Logic of Religion* (1965). One class of what he calls metalinguistic justifications uses authority. Belief on authority can be justified by trust, which Bochenski says is an insight into the truth of two sentences: (1) A knows the situation in the field in which A is an authority. (2) A speaks truthfully about the elements of that field to the subject. He says that the acceptance of (1) and (2) need not be based on any reasoning. The trust described may completely replace it. When applied to religious belief, the trust in question is trust in the revealing agency of God. See chap. 5, esp. 122–28.

whether to believe independently. Believing a person who is currently speaking to me or who has written a book or sent me an email is not very mysterious, but believing God requires a theory of revelation to explain how communication between God and me can succeed. There are a number of accounts of revelation compatible with the view of faith I am endorsing here. I will describe three models, all of which are intended to explain how something communicated by God at a certain moment of time can be transmitted to a great number of people over a very long period of time in widely varying circumstances in such a way that the communication succeeds in producing a state of faith that includes reasonable belief.

3.2. Models of Revelation

A religious community is extended in time as well as space. If it is to last longer than the lifetime of its original members, it must develop a structure to preserve and transmit the practices that define it to distant and future members of the community. Since a community is like a person, it has a memory as well as a future. Tradition is the memory of a community (from the Latin *traditio*, or “handing on”). In most of the major religions, a structure of authority is a component of the tradition that functions to guard the faithfulness of its development. Religious epistemic authority is interwoven with moral epistemic and practical authority, the training of emotions, and more generally, the cultivation of a particular worldview. In Judaism, Christianity, and Islam sacred scriptures are an important part of the tradition. All three religions also have extensive oral traditions with varying degrees of authority.

Joseph Dan says that the term “Kabbalah” was originally used to refer to a sacred tradition of divine origin, part of which is found in scriptures, and part of which is transmitted orally from generation to generation by Jewish leaders. This usage can be found in the Talmudic tractate *avot*, an important Hebrew text probably formulated in the second century:

The first section of this tractate describes the traditional chain of Jewish law and religious instruction, which was transmitted from generation to generation. The first stage of this transmission, as described in this tractate, is: “Moses received [kibel] the Torah on [Mount] Sinai and transmitted it to Joshua, who [transmitted it] to the Elders [of Israel] . . .”; the

text goes on to describe the oral transmission of this tradition to the judges, the prophets, and the early sages of the Talmud. This paragraph was used for nearly two thousand years to validate Jewish tradition as a whole, fixing the Mount Sinai revelation as the point of origin, deriving legitimacy from the sanctity of that event. The term “torah” in this sentence was understood to mean everything—scriptures, the law (halakhah), the rules of ethics, the expounding of scriptural verses (midrash)—everything related to truth of divine origin. Some even said that everything that a scholar might innovate was given by God to Moses: what may seem to be an innovative, brilliant religious observation was already known to Moses, informed by God in that all-encompassing revelation. (Dan 2006: 3–4)

Dan says that what is “received” in this context is kabbalah in its premedieval usage. It is essentially nonexperiential and nonindividual religious truth transmitted by tradition (4).

There are similar ways of understanding tradition and its transmission within Christianity and Islam. Christianity existed as a tradition decades before the Christian Scriptures were written down, and in the early Church, the Word of God did not primarily refer to a book. The Catholic Church consciously preserves a continuity of tradition that includes more than the written Scriptures. Similarly, Islamic scholars possess both the Qur’an and a treasure of divine wisdom passed orally from Muhammad to his disciples and to their disciples.

There are also traditions that are imbedded within the practice of a religion with their own system of transmission and authoritative structures. Christian monasticism is one such tradition. Another is the tradition of Kabbala, which in the thirteenth century came to refer to Jewish mystics who claimed to possess a secret tradition concerning the meaning of the scriptures and other sacred texts in addition to the ordinary kabbalah known to everyone. It was said that Moses received this tradition directly from God on Mt. Sinai and it was secretly transmitted orally from generation to generation (Dan 2006: 5–6).

All three religions have in common a revelation from God at a time in the past, and a crucial part of the practice of the faith is to transmit revelation. But there are different models of what is transmitted, how it is transmitted, and the type of authority structure necessary for faithful transmission.

On one model the transmission of a tradition is reducible to chains of testimony. What justifies belief in what the tradition transmits is a relation to something that happened at the origin—for example, the experience of Moses on Sinai, the Apostles' experience of Jesus Christ, or the revelation of Muhammad, and what happened at the origin is understood as immediate contact with the divine, the experience of which is transmitted by oral and written testimony to the present. The Talmudic *avot* can be interpreted as justifying Jewish legitimacy by the unbrokenness of a chain leading back to Moses' revelation from God, and similarly for other revelations from God at other points in Jewish history. On the chain model it is crucial that the chain is unbroken and that the transmission is accurate. This model assumes that *what* is transmitted remains the same as what it was at the point of origin. Revelation in this model is fragile because every time it passes from hand to hand, it runs the risk that some of it gets lost or distorted.⁶ On this model nobody can be as justified in a belief acquired through the mechanism of the tradition than the person who had divine contact at the beginning of the chain. The nearer one is to the source of revelation, the more complete and accurate the knowledge. Given that we are so far in time from the origin of the chain, the most we can do is to study old sources in greater depth, or perhaps discover ancient books that were lost at some point along the chain.

This model cannot explain how a religious tradition is transmitted without some additional elements. Why would it matter to us what a man called Abraham did, or that Moses had a religious experience in front of a burning bush if we are only the distant recipients of testimony about their contact with God? If what tradition passes on is a reconstruction of someone's experience a long time ago, it is hard to see it as anything more than a historical curiosity, and their written texts as anything more than artifacts of an ancient culture.⁷ Chains of testimony do not add up to a tradition in a sense that pertains to religious belief unless the content of the testimony bears on the future recipients of the testimony. In families there are sometimes stories passed from generation to generation about events in the past (e.g., what

6. Compare Locke's complaint that a chain of testimony gets less reliable the longer the chain (*Essay*, chap. 6, pp. 664–65).

7. Pope Benedict makes this point in several places. See Joseph Cardinal Ratzinger (1986: 28); and Pope Benedict XVI (2007: xvi). See also Scott W. Hahn (2009: esp. chapters 2 and 3).

Grandpa did in the war), and those stories bear on future family members because they are stories about *their* family. A future generation cares about the stories because they were already a family before the stories were told and after they are forgotten. In contrast, revelation defines the religious community to which it applies. No one would care about the secondhand (or millionth-hand) experience of Moses talking to God unless Moses is a part of *us*. What makes him part of *us* is that the intended recipients of the divine message received by Moses include people in the distant future.

There is a second model of tradition that focuses on the recipient's experience rather than the original experience. According to this model, what is handed on in tradition is Scripture, but Scripture speaks directly to the reader or hearer without any need for a tradition of interpretation or authority in its exposition and preservation. The purpose of the tradition is to make possible to present and future members of the community *firsthand* experience of the divine. In the Christian version of this model, the Holy Spirit acts upon the individual believer who reads or hears the Gospel proclaimed, and brings her to faith by grace.⁸ As Augustine says, faith is rational because of the operation of grace, which is God himself moving the person towards union with him, not because of arguments for the reliability of the Catholic Church's transmission of the original revelatory event.⁹ There may also be a Jewish version of this model in Martin Buber's influential proposal of the relation between the divine and human in *I and Thou* (2000).

This model minimizes the function of tradition in preserving the past intact, whereas the first model maximizes that function. It is doubtful that many actual religious communities adopt a pure version of either model, and the models are not mutually exclusive. But distinguishing the models is important because epistemic justification has a different explanation in the two models. For instance, a consequence of the second model is that those scholars who are engaged in scholarship aimed at seeking the "historical Jesus" or analyzing the original texts of the Bible have missed the point of what the tradition is transmitting. From the point of view of the adherent of the second model, the quest for the historical Jesus (or the historical Moses or the historical Muhammad) only makes sense on the chain model of tradition. But the continual action of the Holy Spirit makes the faith of a person in the twenty-

8. A detailed version of this model appears in Plantinga (2000b: chap. 8).

9. See Paul MacDonald (2010).

first century as justified as the faith of one of the Apostles. This model need not deny that the Apostles had a more direct and hence clearer and more detailed experience of Jesus than modern worshipers, but the point of the tradition is to produce faith, and the faith of a twenty-first-century Christian can be as accurate a relation to God as the faith of a first-century Apostle.

There are other models in which tradition is not just secondhand experience, but it is not a series of firsthand experiences either. In describing traditional Judaism, James Kugel argues that the idea of the Bible arose in the period of the ancient interpreters of the Jewish scriptures living at the end of the biblical period. What is transmitted from them is not what was originally written, so it is not the secondhand experience of Moses or the prophets; it is a tradition that transmits and deepens the understanding by each generation of the definitive "Oral Torah" given by the ancient interpreters. These interpreters, largely unknown to us by name, made certain assumptions about how the Bible was to be read: (1) They assumed that the Bible was fundamentally a cryptic text, with a meaning to be deciphered. (2) They assumed that the Bible was a book of lessons directed to readers in their own day even if it appears to be about the past. (3) They assumed that the Bible contains no contradictions or mistakes, and it is in accord with the religious beliefs and practices inherited from the oral tradition. (4) They believed the Bible is essentially a divinely given text in which God speaks through prophets.¹⁰

On Kugel's view, then, Jewish tradition is the transmission of an ancient oral tradition, so it has some features of the chain model, but it is not the transmission of an original experience or an original text, nor is it the transmission of something the point of which is to give the contemporary Jew firsthand experience of God. The ancient way of understanding God at the end of the biblical era was a high point in Jewish history, and the tradition transmits that way of understanding God to the present in a way that was intended by God to give lessons to contemporary Jews. It is a misconstrual of the intentions of the Bible's framers to read it without the four assumptions, or in a way that denies the authority of the ancient interpreters.

These four assumptions are criteria for distinguishing growth from deformation of the tradition, and the model described by Kugel also has implications for the methodology of modern biblical scholarship. Kugel writes:

10. James L. Kugel (2007: 14–15). Kugel gives the same four assumptions of the ancient interpreters in (1997: chap. 1).

The whole approach of modern biblical scholarship, which is predicated on disregarding the ancient interpretive traditions of Judaism (and, for that matter, Christianity) and rejecting the four fundamental assumptions that underlie them, must inevitably come into conflict with traditional Jewish belief and practice. The modern program rules out of bounds precisely that which is, for traditional Jews, the Torah's ultimate significance and its definitive interpretation. To insist on taking the Torah's words at face value, without regard to what the Oral Torah says about them, is thus for a traditional Jew somewhat comparable to telling a Christian that he or she must take the laws of the Old Testament at face value, without regard for all that Paul has to say about them in the New Testament, as well as about the new covenant of Christianity that has come to take their place. I do not know any Christians who would accept such a proposition.

My own view, therefore, . . . is that modern biblical scholarship and traditional Judaism are and must always remain completely irreconcilable.¹¹

Jewish tradition is not the only kind of tradition that can be understood on the model of a practice that reached a state of perfection by oral transmission a long time ago, and which has subsequently been transmitted to the present. Cases of this kind are not necessarily religious. A simple example is the practice of Italian glassmaking. By the sixteenth century Venetian glassmakers on the island of Murano reached a level that arguably has never been surpassed, and although their secrets were eventually discovered by glassmakers in other countries, the master glassmakers in Venice continue to make glass in the style and technique they perfected centuries ago.¹² It seems

11. Kugel (2007: 681). This passage comes at the end of a large and impressive work comparing the ancient and modern interpreters of the Bible.

12. An interesting contrast in ways of looking at the tradition of glassmaking was revealed when the American glass artist Dale Chihuly created a series of chandeliers made in glass factories in five different countries, working with Italian masters Lino Tagliapietra and Pino Signoretto, and then suspended them over the canals of Venice and in some of its most famous landmarks. Chihuly supervised the creation of the work in gratitude to Venice for what he learned from the Murano glass makers. In a video interview shown at the Oklahoma Museum of Art's permanent exhibit of Chihuly glass, one of Chihuly's assistants is quoted as saying that they learned an enormous amount from the Italian glassmasters with whom they worked, but the latter will change nothing in the way they make glass. My interpretation is that Chihuly's glass is intentionally original. In contrast, the Venetians intentionally preserve a particular tradition of glassmaking. Chihuly's glass expresses the personality of a particular man, whereas Italian glass expresses the refinement of centuries of work by many people. (The video was removed when the exhibit was redone at the end of 2011.)

to me that the performance of classical music and the practice of cooking in three-star Michelin restaurants in France transmit tradition in the same way. The musical performance or preparation of the culinary dish is not intended to be identical to those produced in the past since the cultural context gradually changes, but the changes are only those needed to keep the tradition alive at its best.

In Christianity there are many models of the transmission of tradition and its role in the faith of an individual believer. The first model of a chain originating in the writing of the text of Scripture is clearly one of them. On this model revelation is wholly contained in the original text and its faithful transmission is guided by the Holy Spirit. I have already mentioned the Christian version of the second model in which the transmission of the Gospels is the occasion by which the Holy Spirit produces faith as a first-hand relation to God with epistemic, affective, and conative components. There are also versions of the traditional Jewish model of tradition in Christianity. The Catholic Church has always attached great importance to oral tradition. In rabbinic Judaism there are two Torahs: the written Pentateuch and the traditions of its proper interpretation and application that were orally transmitted. Similarly, the Catholic Church teaches that Revelation is contained in both Scripture and tradition, some of which is orally transmitted or has been orally transmitted in the past, guided by the Holy Spirit. For the Jew, there is, in addition to biblical interpretation, the Oral Torah, which includes prayers and blessings for certain occasions, some areas of civil and criminal law, matters concerned with marriage and divorce, description of temple rites, purity statutes, and so on (Kugel 2007: 680). Similarly, Apostolic and post-Apostolic tradition recognized by Catholic Christians includes sacramental rites, liturgical practice, ecclesiastical discipline, practical conduct such as the Sunday obligation to attend Mass, and laws concerning marriage. As in the model of traditional Judaism described by Kugel, the tradition recognizes a high point in the past—in this case, the Apostolic age, and the Rule of Faith is the criterion for distinguishing authentic development from deformation of the tradition. The third model of revelation, then, is represented in both traditional Judaism and traditional Catholicism.

Tradition must develop and undergo gradual change because what is handed on (*traditum*) is not limited to a product like a text or the process for making a product like Murano glass, but a way of living in contact with

God.¹³ That includes having certain beliefs about God, human beings, and the relation between them, as well as characteristic emotions, attitudes, and sacraments that express religious truths in simple external ways in which ordinary people can participate. Tradition requires a structure that assigns responsibility for the authenticity of the tradition that is handed on, and the structure is part of the tradition. A hierarchy of one bishop with a college of presbyters and deacons with teaching authority as well as leadership was already established in Syria and parts of Asia Minor before the books of the Christian canon had been completed, and by the third quarter of the second century, every church of which we have information, with the exception of Alexandria, had a single bishop.¹⁴ That tradition continues in the Catholic Church, but it is doubtful that most Catholics give much thought to the precise date of the universal acceptance of episcopal hierarchy since what they trust is the tradition, and they expect the tradition to gradually evolve. But my purpose is not to endorse a particular interpretation of any particular tradition. The issue for this project is whether it is justified to accept authority within one's tradition, however that is understood within that tradition.

The way tradition is understood within Christianity was deeply affected by the Enlightenment and the subsequent privatization of religious activity. This was particularly important in English Puritanism, the Dutch Collegians, and in subsequent American Protestantism.¹⁵ The distrust of religious authority and its disappearance in some communities is often interpreted as the logical consequence of the rise of the value of autonomy, as I have observed before. But if I was right in my defense of the justification of communal authority theses in chapter 7, authority in a community is justified by the commitments of conscientious self-reflection. If I conscientiously judge that I am more likely to believe what God revealed if I take the authority of the Church than if I do not, I am justified in taking beliefs on authority by the

13. A standard way to define tradition is "the communication by the living Church of the Christian reality, and the expression, either oral or written, of that reality" (Fichtner 2003: 14:133).

14. Sullivan (1983: 43). This assumes that the traditional dating of the letters of Ignatius of Antioch is correct. Sullivan says that if the letters are authentic, they were written during the reign of Trajan (before 117). See p. 20, note 8 in Sullivan for a comment on a writer who disputes the date.

15. There are many discussions of the history of this period. See, for example, Adam B. Seligman (1997: chap. 5, esp. 140–41).

first justification thesis. If I conscientiously judge that I am more likely to get a belief that will satisfy my future conscientious self-reflection if I take the belief on the authority of the Church than if I do not do so, the belief is justified by the second justification thesis. Religious authority can be justified for a believer the same way any epistemic authority can be justified. Epistemic authority is justified by principles arising from the conscientious reflection of self-conscious beings.

4. CONSCIENTIOUS BELIEF AND RELIGIOUS AUTHORITY

Trust in a community is an expression of the virtue of intellectual humility. Rejecting epistemic egoism is the conscientious thing to do when we learn that many others are like ourselves. Accepting epistemic authority is the conscientious thing to do when we learn that others are epistemically superior to ourselves. Trust in a community that has existed for many hundreds of years is often more conscientious than trusting a community of my contemporaries, and much more conscientious than trusting myself alone. If tradition is the democracy of the dead, as G. K. Chesterton observed, ignoring it is a kind of egoism of the contemporary. Of course, there are some domains in which the beliefs of contemporary experts are more authoritative than the beliefs of the authorities of the past. Progress in these domains means not only accumulating knowledge, but replacing much of what was previously believed. This is reasonable in science because we know that if the past members of a scientific community were alive today, they would trust the contemporary experts more than themselves. But many domains are not like that. The moral and the religious are not, and probably the aesthetic domain is not either. I am not denying that there is progress in these domains, but the progress is like the progress in knowledge and understanding of a single person throughout a lifetime.

Pascal wrote, "The entire succession of men, throughout so many centuries, should be thought of as one and the same man, ever-present and learning continually."¹⁶ As Pascal sees human history, we contemporary humans constitute a state of that man, with a memory and a future. We can change,

16. *Opuscles* (ed. Brunschvig), p. 80. Quoted in Yves Congar (2004: 3).

but when we change we should be aware that we have a responsibility to future members of the community not to deprive them of their past. I am suggesting that the same thing applies to a community extended in time. Members of the community who are alive at the same time are like one stage of an individual person's life. They pass on their past to the community's future, and the members of the community have an extended past and an extended future through participation in the community.

If I am right that a community is an expanded self, then trusting it is an extension of self-trust. The difference is that self-trust is inescapable, whereas trusting Us is escapable. I can choose not to be a part of a community any longer, but doing so conscientiously occurs when a person has already changed in ways that lead her to believe that it is no longer the case that the beliefs of the community satisfy conscientious self-reflection better than beliefs she gets on her own. It is not likely to do so in the future because it has already failed to do so in the past. But membership in a community can sometimes be a very deep part of the self. The community's beliefs almost certainly *will* survive conscientious self-reflection for such a member because they already have done so for as long as the member has reflected. Furthermore, some of the community's beliefs may have the role of determining what *other* beliefs satisfy her conscientious self-reflection.

I proposed that we have two natural desires related to the self—to make our states fit their objects, and to make our states fit each other. Our ultimate test of the former is the latter. In some domains it is very difficult to tell whether one's community does a better job of getting the truth than oneself or some other community if the rules one uses for determining the truth are limited to the specifically epistemic. For instance, trust that a particular religious tradition puts one in the best position to get the truth depends in part on trust that it contains the highest attainment of the human spirit in relation to God. But to think that, one must have nonepistemic trust in the tradition and would need to determine that the tradition has that quality by the fact that its teachings satisfy conscientious reflection upon one's total set of psychic states, not just one's set of beliefs. The more fundamental thesis for the justification of religious epistemic authority, then, comes from JCEA 2:

Justification of Religious Epistemic Authority Thesis

The epistemic authority of my religious community is justified for me by my conscientious judgment that if I believe what We believe, the result

will survive my conscientious self-reflection on my total set of psychic states better than if I try to figure out what to believe in a way that is independent of Us.

This thesis can be extended to show how to justify a religious community's practical authority as well as its epistemic authority. The extended thesis would recognize the limitations on Raz's Normal Justification Thesis described in chapter 7. Given the conclusions of that chapter and my argument in this chapter that religious communities can satisfy the conditions of communal authority, I suggest the following general thesis:

Justification of Religious Authority Thesis

The authority of my religious community is justified for me by my conscientious judgment that if I engage in the community, following its practical directives and believing its teachings, the result will survive my conscientious self-reflection upon my total set of psychic states better than if I try to figure out what to do and believe in the relevant domain in a way that is independent of Us.

Other versions of this thesis would not refer to Us, but would refer to a community I might join. My general argument is easy to apply to such cases. I might join a community, not because I already identify with it, but because I conscientiously predict that I will in the future. The Church is more than a body with the authority to reveal truths of faith and morals. There are other natural desires the satisfaction of which can be better satisfied by participation in a wisdom community than on one's own. These desires include the desire to know and do the good, to acquire not just knowledge, but understanding, to learn patterns of living and principles of action that result in a more integrated self, to be surrounded by grace and beauty, and to experience the delights of living among persons whose own pursuit of these ends enhances one's own. The authority of a community can be justified by a conscientious judgment that these desires will be more likely satisfied by participation in the community.

John Henry Newman said that holiness is the true test of a church (1994: 144).¹⁷ In describing his gradual move towards Catholicism, he says that he

17. Newman (1994: 144). In this passage he says that that is what he believed at a time he was opposed to the Catholic Church, but he was attracted by its holiness.

came to believe that the church of Rome gives the freest scope to feelings of awe, mystery, tenderness, reverence, and devotedness (162). For him, the justification for joining the community was that it helped him have fitting emotions as well as beliefs. In this book I have argued that all persons have the same aim: to satisfy their natural desires, to have emotions that fit their objects, to have knowledge and understanding. Many others add another aim: to become holy. Authority is justified for a person by that person's own judgment that the authority of another person or community will help her achieve these aims. Nothing I have said in this book implies that authority is justified *only* in this way. My purpose is to show that the rejection of authority on grounds of its incompatibility with the autonomy of the self is badly mistaken. Authority can satisfy the demands of modern philosophers who require that authority be justified by its derivation from the authority of the self. Authority in religious bodies with an ancient authority structure can be justified the same way.

I think that the argument of this chapter indicates that there should be a reorientation in the epistemology of religion. In my opinion, firsthand experience gets too much attention in discussions of the rationality of religious belief. I can trust my tradition more than my own experience in many cases, and of course my experience is limited to the experience of one person. Given that we reasonably take beliefs from others or based on the experience of others, the structure of the process by which those beliefs are dispersed within a community and continued through the future life of the community needs epistemological models. The chain model is not sufficient to explain why I should trust the experience of Moses or the experience of the Apostles, reports of which have probably been distorted in multiple ways during the course of many centuries. I am suggesting a rule of justification that bypasses the chain model, although the belief that the chain is basically accurate can be one of the beliefs of the tradition, and I can be justified in believing that.

Religious faith is impossible to explain, much less justify, on the evidence view of testimony. That view forces us to either redefine faith as belief on a certain kind of evidence, as Locke did, or we must say that faith is nonrational, based on emotions that have nothing to do with epistemic justification. I have argued for the rationality of emotions such as trust, and have defended the justification of taking beliefs from the trusted other. Relationships of trust can be very deep, and they can form the basis for handing on

the insights of the wisest among us. From the viewpoint of the religious community, authority functions to foster those insights and to preserve faith within the community in the indefinite future. But I have argued that authority also serves to improve the self. That means there is always a way to reveal religious tyranny. Either religious authority satisfies conscientious self-reflection or it does not. When it does, no one should object. When it does not, the members of the community itself should object.

Trust and Disagreement

1. THE ANTINOMY OF REASONABLE DISAGREEMENT

At the beginning of this book we looked at some characteristic values of the modern period that affect the practice of contemporary epistemology: suspicion of authority, the ideal of self-reliance, and egalitarianism. Suspicion of authority arguably follows from either egalitarianism or self-reliance, but egalitarianism and the ideal of self-reliance pull us in opposing directions. It seems to me that the conflict between egalitarianism and self-reliance is at the root of the current debate about reasonable disagreement in belief. That problem arises when I discover that a person who seems to be as reasonable/conscientious as I am and who possesses all the same evidence relevant to some proposition p , disagrees with me about p . Later I will look at the conflict as it arises between communities, particularly religious communities.

Positions on the problem of epistemic disagreement vary widely, and there are well-known philosophers at both extremes and almost every point in between. I think that the existence of the two extremes can be explained by the fact that the value of self-reliance conflicts with egalitarianism, and either of these two values can drive a position at the expense of the other. Suppose first that I adopt the ideal of self-reliance. If I have a pure version of this position, I will be an extreme epistemic egoist and I will be unmoved by disagreement. I will think that as long as I have done my conscientious best and I believe p , why should it make any difference to me that somebody else believes not p ? It is *my* conscientiousness that counts for me, not hers. I trust *my* reasoning and other epistemic powers

rather than hers because my powers are mine and hers are hers. Her self-reflections are her business, not mine.¹

If I take this line I cannot think of others as my epistemic equals when I reflectively evaluate my beliefs. In fact, I must deny that *anybody* else is equal to me. I cannot even be a weak egalitarian, although I could be a methodological egalitarian. That is, I could think that other persons should take the same position in their own reflections about their own beliefs.² But even if we accept this way of looking at the primacy of the self, something has to be said about the general equality of others to myself. Egalitarianism is too important a value to be ignored.

On the other hand, suppose that I accept intellectual egalitarianism and ignore any sense in which the self is primary. If I take that line, I must say that my reasoning is just one bit of reasoning among many. When my belief conflicts with the belief of another person whom I judge has epistemic powers and virtues equal to mine, there is no reason to think that I am the one who is right, and so I have no reason to keep my belief.³ For many philosophers, this position is just as unpalatable as the first because it seems to lead to skepticism.⁴ I have another objection. As reasonable persons we aim to

1. A version of this position is defended by Alvin Plantinga (2000b). He argues that a person is internally rational if he has done his epistemic best with respect to a certain belief, and can be fully reasonable in continuing to believe in the face of disagreement. But Plantinga also says that awareness of religious disagreement is an undercutting defeater for one's own religious beliefs, and it might reduce the warrant of the belief by reducing the degree of belief. See the end of Plantinga (2000a).

2. Ralph Wedgewood (2007: 261) has given tentative support to an egocentric epistemic bias, arguing that I am justified in giving my belief extra weight in the face of peer disagreement because the belief in question is mine. Gideon Rosen (2001) and Peter van Inwagen (1996, 2010) both say that one can continue to believe a proposition in the face of peer disagreement despite lacking a symmetry breaker after all the evidence has been disclosed.

3. For a defense of the egalitarian view see Feldman (2006). In discussing a disagreement between Pro and Con, Feldman argues that if Pro recognizes both alternatives as reasonable, he lacks a reason to prefer his own alternative, so accepting that alternative rather than the other is arbitrary (2006: 226). Christensen (2007) says: "Given that my friend and I are generally reliable thinkers who have studied the same evidence, the fact that we disagree will be explained by the fact that at least one of us has made a mistake in this case. But intuitively, the explanation in terms of my friend's mistake is no more reasonable than the explanation in terms of my mistake, and I should acknowledge this by moving my belief toward hers" (198). See also Elga (2007).

4. For a different view, see Nathan King (2011), who argues that the epistemic principles most commonly discussed in the disagreement literature lack "skeptical bite" because we seldom know or have reason to believe that some other person is our epistemic peer.

resolve disagreement, and so we aim to share theoretical reasons, but if we assume that there is uniquely one epistemic attitude towards a proposition (belief, disbelief, suspension of belief) that is most reasonable on the theoretical reasons, and if the theoretical reasons are the only reasons there are, it follows that as reasonable persons we aim to be epistemically identical.⁵ The product of my powers is just one vote among many on what the epistemically identical stance should be. The problem with this position is the opposite of the first: it does not give the self sufficient importance. And since the self is partially constituted by beliefs, it threatens to dissolve a large part of the self.

The commonsense solution is to compromise. Why not say that we should not place too much more trust in our own epistemic faculties than in those of others, but we need not go so far as to think of others as our equals? So we can give others some degree of trust, but a degree less than we give ourselves.⁶ This approach would avoid both the unpalatable consequence that the beliefs of others do not count, and the equally unpalatable consequence that they count for so much that we have to give up our cherished beliefs when faced with unresolved disagreement.

This approach may give us the conclusion we want, but I think it is actually hopeless. Consider the following analogy. At the time of the Philadelphia convention of 1787, slaves were not allowed to vote. The slave states argued that each slave should be counted for purposes of determining the number of representatives a state would get in Congress, and those opposed to slavery argued that slaves should not be counted if they were not allowed to vote. The solution was the famous three-fifths compromise according to which each slave was to count as three-fifths of a free citizen for the purposes of determining representation.

Now it seems to me that there is a principled argument to count the slave as a citizen, and there is a principled argument not to, but there is no argument at all to count a slave as three-fifths of a free citizen. Of course, nobody

5. Roger White (2005) develops interesting arguments for the uniqueness thesis, and the principle is supported by Feldman (2006: 235). For resistance to uniqueness, see section 2 of Thomas Kelly (2010).

6. Adam Elga (2007) mentions the "split the difference" view wherein you assign your peer a value that either increases or decreases the level of your confidence in p , but Elga does not require any mandate that one must always follow the split the difference view. Christensen agrees (2007).

thought there was a legal reason to consider a slave three-fifths of a citizen; the rule was simply meant to break a practical impasse. But we have no such excuse when the issue is what it is reasonable for us to believe. There is a principled argument to count the belief-forming powers of apparent epistemic equals as equal in trustworthiness to my own, and a principled argument not to count them at all, but what is the argument to count them somewhat? There is no argument for a compromise position on trust other than the desire to avoid the two extremes.

Compare Kant's antinomy of free will and determinism. Kant argues that there is a strong and fully reasonable argument that we are free, and there is a strong and fully reasonable argument that we are determined, and the proposition that we are free conflicts with the proposition that we are determined. But can anything be said for a middle ground in which we are somewhat free and somewhat determined? Either universal determinism is true or it is not. If it is true, there is no free will. If it is not, there is no need to make a compromise with determinism in order to defend free will. The middle ground is supported by neither of the theses in conflict. The motive to adopt it is the practical desire to resolve the conflict; its support is not theoretical.⁷

I think this is a common problem with the reasonableness of theoretical compromise. A middle ground between two opposing principles is supported by neither principle, so it is theoretically worse off than either of the two extremes. The reason we like the middle ground is just that it is the middle ground, not that there is a theoretical justification for the position. I am not denying that it can be reasonable to weaken our principles and thus to block a conflict, but only if the weakened principles are plausible apart from the desire to resolve the conflict. Perhaps some compromise views on reasonable disagreement do arise from independently plausible principles, but I think that it is more likely that the conflict is a symptom of deeper problems or issues we have been ignoring. For Kant, the antinomy of free will and determinism arises because the two sides to the conflict apply to different domains. The self is part of the empirical world and as such, it is determined, but the distinction between such a world and the world of things-in-themselves leaves room for the possibility of the ego's freedom.

7. Here I am using the term "theoretical" in contrast to "practical." I am not referring to third-person reasons for belief as opposed to first-person, deliberative reasons.

I think that the problem of how a conscientious epistemic agent should handle disagreement uncovers the same type of problem about the self. If we treat the self as an empirical object, it is most reasonable to think of the self as equal to other selves in epistemic capacities. I am justified in trusting myself only because of my possession of qualities that many other persons possess. There is no reason to favor myself because I am myself. So egalitarianism is the reasonable stance to take when we think of the self as one of any number of selves. On the other hand, in this book I have focused on the self-reflective capacity of the self, a capacity that requires that there be an executive self that does the reflecting. The self that investigates, reasons, reflects, and sometimes changes its mind is *myself*. I have to assume that that self satisfies the conditions of an executive self in order to go through the reflective processes I perform. There is no need for me to treat other selves as satisfying those conditions because other selves are not doing my reflecting.

My point here is not that we have to embrace Kant's view of the transcendental ego. But I do think that the problem of reasonable disagreement arises from a deep conflict about the way a reasonable person ought to treat the self, a problem identified by Kant and for which his solution is the distinction between the empirical and the transcendental ego. But we do not need to accept a particular solution in order to recognize the problem in attempting to take a consistent point of view on the self.

Kant's discussion of antinomies may be applicable to the problem of reasonable disagreement in another way. Kant says that antinomies arise in the attempt to reason about the world as a whole. The attempt to reason about one's beliefs as a whole is the epistemic analogue of reasoning about the world as a whole. I argued in chapter 2 that our beliefs taken as a whole are not amenable to the kind of justification we can give for a particular belief in an ordinary context in which we assume the justification of certain other beliefs. Reflecting upon one's beliefs as a whole reveals the rational inescapability of self-trust. In chapter 3 I pointed out that a difference between self-trust and trust in others is that I cannot escape trusting my beliefs as a whole, whereas I can escape trusting the beliefs of another person as a whole. But I argued that when we look for a ground for self-trust—a *property* in virtue of which self-trust is justified, we find that any property that can serve as such a ground is shared by many other persons, and the most basic ground of self-trust is a property we share with all others: having human epistemic capacities.

That is why I cannot treat those who disagree with me as less trustworthy than myself just because I am I and they are they. The executive self determines the reasonableness of adopting or maintaining beliefs using whatever I trust, but my own self-reflections lead to the consequence that I should trust others. When I reflect upon my beliefs or potential beliefs, I do not simply ask myself, "What is the evidence for the proposition believed, and how should those reasons be weighed?" I also ask myself why I should trust any qualities I have that permit me to identify and handle evidence well. On what basis do I trust my ability to make the move from my reasons for a belief to successful attainment of the truth? Any answer I give to that question will be a property shared by others. One consequence is that I cannot consistently treat others as simply sources of evidence for me. Trust in the deliverance of their faculties is a rational consequence of consistent trust in the deliverance of my own faculties.

It follows that the egalitarian is right that any property I have that grounds my self-trust—in particular, epistemic conscientiousness—is a property shared by many others. But on the story I have told about self-trust, there is no *property* trust in which is as basic as trust in the executive self. I do not trust myself *because* I trust conscientiousness or intellectual virtue or the powers of reasoning or any other capacity. I do not reason as follows: I trust reasoning; therefore, I trust my own reasoning. I trust conscientiousness; therefore, I trust my own conscientiousness. I trust human perceptual powers; therefore, I trust my own perceptual powers. My reasoning is the reverse. Self-reflection reveals that natural self-trust cannot be avoided, and further reflection allows me to identify the particular aspects of myself that I trust. I trust myself in particular when I am a certain way, a way I have described as conscientious, and I see when I am conscientious that trust in my own possession of that property commits me to trust the instantiation of the same property in other persons. I trust the quality of conscientiousness because it is a rational commitment of self-trust. Self-trust is prior to trust in a quality such as conscientiousness or a faculty such as reason. It is prior to trust in any property at all. So the person who insists upon the primacy of the self is right that there is something about the self that is always more basic than any belief we have about properties shared by others, including belief in the trustworthiness of conscientiousness and the belief in egalitarianism, and it is more basic than any trust that derives from such beliefs.

2. DISAGREEMENT AND DELIBERATIVE VERSUS THEORETICAL REASONS

To trust myself means to trust whatever I do that I trust upon reflection when I exercise my powers as well as I can. The problem of disagreement between my beliefs and the beliefs of others arises because trust in myself includes both trusting my beliefs and trusting features of myself that commit me to trusting the beliefs of others. As I see it, then, the issue of disagreement is not a conflict between self-trust and trust in others; it is a conflict that arises within self-trust.

The fact that the problem of disagreement fundamentally arises within the set of things *I* take to be trustworthy upon reflection means that the executive self always has the last word. In the conflict between the modern values of the primacy of the self and egalitarianism, the primacy of the self wins, but it wins only in the sense that Kant's transcendental ego permits freedom to win over determinism. If we conclude that the conflict is resolved, we have failed to understand the antinomy. We cannot just declare one side the winner and ignore the other side. Kant does not think that determinism goes away, and his transcendental ego is not the same as the phenomenal ego that generated the conflict with determinism. In the disagreement problem, the primacy of the self wins in the sense that we have a way to pose the problem anew with the executive self as primary, but egalitarianism does not go away, and the aspects of the self that generated the conflict still conflict. The executive self has the power to change those aspects of the self by self-reflection. That means that the primacy of the executive self is *not* the primacy of my own prior beliefs. I have to face the fact that my own reflections may force me to change part of myself.

This line of reasoning shows that the problem of disagreement arises in its most interesting form out of the first-person, deliberative reasons for a belief, not the third-person, theoretical reasons. If the problem of disagreement was a problem of conflict among theoretical reasons, it would not be a problem worth the amount of attention it has recently received. To see why, suppose I think the only reasons there are are theoretical. When I examine those reasons, if I am an extreme epistemic egoist, all that I consider relevant is the evidence that bears directly on the truth or falsehood of some proposition *p*. When I do that, the epistemic attitude that I or somebody else has towards *p* is irrelevant. The extreme epistemic egoist therefore does not find

disagreement a problem. If instead, I am a standard egoist, I expand the relevant evidence to include second-order beliefs about the reliability of persons. When I do that, I may discover that two reliable persons have conflicting beliefs whether p . Sometimes one of those persons is myself. The fact that one is myself is irrelevant when the issue is theoretical reasons. My total evidence then includes the evidence I acquired by the direct use of my faculties for p , the higher-order evidence that I am a reliable person who believes p , and the higher-order evidence that some other person is reliable and believes not p . This type of conflict within my theoretical reasons is neither uncommon nor very interesting. We have conflicts between subsets of our theoretical evidence all the time. For the most part, we resolve the conflict by waiting for more evidence, but of course that sometimes cannot be done. The failure to do so at any time is due to the limits of our ability to get evidence. But there is nothing distinctive about this problem.

It follows that if the problem of disagreement is a problem of conflict among the third-personal, theoretical reasons, it is either no problem at all, or it is an ordinary problem within a body of evidence. In neither case is there a special problem of disagreement. What makes the problem of disagreement distinctive and deserving of special attention is that my reflection upon my trust in my power to properly handle the theoretical reasons and anything else relevant to getting the truth whether p leads me to trust many others, some of whom disagree with me. Trust in others is a commitment of self-trust. The problem therefore arises within my first-person, deliberative reasons, and that is where I need to look to resolve it.

3. SELF-TRUST AND RESOLVING DISAGREEMENT

How far can we go in resolving the conflict? Let us start with my version of how the dissonance arises. Suppose that after reflecting upon my need for self-trust and my grounds for trusting myself, I come to believe some form of the principle of trust introduced in chapter 3:

Principle of Epistemic Trust in Others

In any case in which, by believing in a way I trust in myself, I am led to believe that others have the same property I trust in myself (to the same degree as I have myself), I have a *prima facie* reason to trust them as much as I trust myself.

It is not important for the problem I am addressing that I trust the principle exactly as I have worded it here. It is sufficient that I trust some principle that accords epistemic trustworthiness to some other person equal to myself in some set of circumstances.⁸

Unfortunately, I believe many things that conflict with the beliefs of others whom I take to be trustworthy upon reflection, and some of them are key elements of my membership in a community. For instance, the Trinity is a central element of Christian belief, whereas denial of the Trinity is entailed by the Muslim version of monotheism. Membership in a Christian or Muslim community is in part defined by the beliefs that divide them.

The same point applies to moral and political disagreement. Communities are partly defined by beliefs that separate them from each other. One side believes that liberty trumps equality, and the other side believes the contrary. One group believes that society should foster a conception of the good that includes protection of all innocent human life, whereas another group believes that the foremost aim should be the protection of individual autonomy. The existence of distinct communities exists in part because people *do* disagree about matters, some of which are important. Of course, some of these disagreements are not fully reflective for individual members of the community, but the community itself often embodies a high degree of reflection upon the particular beliefs that divide it from other communities. This is especially true of religious beliefs. It is less true of political beliefs, given that they are usually associated with political parties whose leaders are primarily concerned with getting elected, and whose members spend little time thinking about the deeper beliefs that are supposed to distinguish one party from another.⁹ Nonetheless, I assume that there are communities divided by moral or political beliefs, and the community itself may arrive at their beliefs by conscientious communal reflection.

There are views of truth that permit the conclusion that many apparently intractable disagreements are only apparent. It might be hard to make this

8. There are similar principles in the literature on reasonable disagreement, for instance, the principle Christensen (2007: 211) calls "cognitive parity."

9. For instance, it is often assumed that Republicans are more fiscally conservative than Democrats, but if you point out to a fiscally conservative Republican instances in which that is not the case, the response is confusion. Which side should he be on? He has to decide whether he is a Republican first and a fiscal conservative second, or the other way around. Of course, you can have the parallel conversation with a Democrat.

position plausible when the disagreement is over values like liberty versus equality, but religious disagreement is sometimes thought to be in this category. Take the doctrine of the Trinity as an example. I believe

- (1) There is a Trinity of divine persons,

but I know that other people conscientiously believe

- (2) There is not a Trinity of divine persons.

As I mentioned in chapter 2, some philosophers are suspicious of the idea of truth, and it is possible to deny that (1) and (2) conflict. So some writers, often inspired by Wittgenstein's "Lectures on Religion" (1972), maintain that there is no truth about the matter of the Trinity to disagree about. There are also those whose view of truth permits a much looser correspondence between the cognitive content of a proposition such as (1) or (2) and the reality it attempts to represent, and both (1) and (2) turn out to be true.¹⁰ There is no conflict, then, if neither (1) nor (2) is true, and there is no conflict if both (1) and (2) are true.

This is not the way ordinary believers of (1) or (2) think about the matter, nor is it the way scholars thought about the matter until they thought they were forced to review their position on the nature of truth under pressure from the problem of disagreement. The original position of all of us is that there is disagreement, and the fact that some people believe (1) and others believe (2) is an example of it. I want to approach the problem as it arises for this initial position. The suggestion that the idea of truth must be given up or modified is something we can leave open as a solution to the problem.

How should I think about my disagreement with people whom I take to be as conscientious as myself? In my discussion of psychic dissonance in chapter 2, I mentioned that when there is a conflict among the things

10. The most influential version of this position is John Hick's pluralism according to which the major world religions are different manifestations of the same noumenal reality—God or the Real. He does not say they are equally accurate, but many apparent conflicts in belief are cases in which both sides are right in the only way a human being can get divine reality right—through the lens of his particular culture. Hick does not say both of two apparently conflicting propositions are true, but they can both correspond to different phenomenal realities. Hick has explained and defended this view in many places. See, for instance, Hick (2004).

I trust, often the degree to which I trust what I trust shifts unconsciously and I do not have to make a decision about the reasonable way to proceed. So I might briefly look on Amazon at the description of a newly published book and form a negative judgment, but when someone whose opinion I trust tells me he has read the book and it is excellent, my opinion changes without struggle or need to deliberate. I resolve the conflict automatically, sometimes even below the level of consciousness. In this book I have treated conscientiousness as the reflective exercise of faculties that in many circumstances operate without the need for conscious deliberation. I need to reflect about the structure of what I trust when the conflict is not resolved automatically, or when a moment comes that forces me to make a choice that requires me to favor one of the things I trust over another. The conflict between (1), my belief that many conscientious people believe (2), and the Principle of Epistemic Trust in Others can easily be of that kind.

Given the argument of this book, it is reasonable to resolve the conflict in favor of what I trust the most when I am thinking in the way I trust the most, that is, conscientiously. In fact, there is nothing else I can do. My own reflections show me that there is no deeper principle upon which to base a decision about which of a set of conflicting beliefs I should keep and which I should give up. In some cases I might decide that the evidence supports one belief more than a conflicting belief, but the evidence is helpful in deciding among competing beliefs only when I trust that the alleged evidence has been accurately identified and evaluated more than I trust the competing beliefs. In chapter 2 I argued that what I take to be the evidence and what I take to be the relation between the evidence and a given belief depends upon a more basic trust in myself. Trust in evidence is not the bottom line, and in some cases my trust in a particular belief can be greater than my trust in what appears to be the relevant evidence, or my trust in the relation between the relevant evidence and the belief it supports. In such cases it is reasonable to keep the belief I trust the most upon conscientious reflection even when it is not the best-supported belief in relation to what I otherwise take to be the relevant evidence.

If a belief is firmly rooted in a network of other beliefs and has already survived considerable self-reflection, it can be reasonable to treat putative evidence against it as either false or misleading. Many people did that when faced with evidence of differences in IQ between races in Herrnstein and

Murray's well-known book, *The Bell Curve*.¹¹ The problem was not that there was not enough evidence or that it was misleading. There was a substantial amount of evidence, or at least, there was enough evidence that if it had supported the opposite conclusion, the book's detractors would have been satisfied with the quantity of evidence. But the conclusion was rejected out of hand by many people regardless of the evidence. Although I would not take that to be a conscientious response myself, I think it might have been a reasonable thing to do for many conscientious persons. That is because it is possible to conscientiously trust a belief more than even a substantial amount of putative evidence against it.¹²

Disagreement with people we conscientiously judge to be conscientious should be handled the same way: in a way that we conscientiously judge will survive conscientious self-reflection. I encounter disagreements between conscientious persons and between such persons and myself about a multitude of issues—the best method for brewing coffee, the sun requirements of Japanese maple trees, the best health plan for the country, to take a few examples at random. In some cases the disagreement will eventually be settled by evidence both parties accept, but that is not the case for all disagreements, and in any case, what is relevant is not whether the disagreement will be settled by future or present evidence that somebody has somewhere. What is relevant for me is what I conscientiously believe, and what I predict will satisfy my future self-reflection, given what I conscientiously predict about myself. I face a problem now when my belief about the proper way to brew

11. Herrnstein and Murray (1996). What riled people the most was the public policy proposals at the end of the book, but the scientific evidence was never found faulty. It is ironic that the idea that intelligence is inherited is not necessarily a right-wing claim, and as Steven Pinker (2002: 301–2) has pointed out, people are often shocked to find out that the idea that there are innate differences in intelligence has been used to support left-wing political goals, e.g., Wooldridge (1995). Herrnstein and Murray themselves claimed that the inheritability of intelligence ought to lead to a greater commitment to Rawlsian social justice, but that is not the way *The Bell Curve* was perceived. The outcry against it is relevant to my claims about trust because it seems to me that on a charitable interpretation of the book's detractors, they might have been reasonably trusting a network of beliefs even if they were mistaken in their judgment about where the conclusion about racial differences in IQ fits in that network. However, it would have been preferable if they had acknowledged what they were doing.

12. Another example has been given by Alvin Plantinga (1991), who has argued that it is possible for a person rationally to be a young earth creationist in full knowledge of the evidence against it.

coffee conflicts with the belief of someone I trust, and I cannot wait to resolve the issue by studying all the evidence that exists now, much less in the future, because I will want to brew coffee again tomorrow. The same point applies to us collectively in making public policy decisions about medical insurance and many other issues. I think, then, that even though the problem of irresolvable disagreement gets the most attention in the literature, it is not more difficult from the point of view of the believing agent than many cases of disagreement that are in principle resolvable.

The general procedure I advocate is no different if the conflict is over a trivial matter. For instance, I believe that the shape of a wine glass affects the taste of the wine, but I also believe that my friend Jack believes that it does not. So I believe:

- (3) The shape of the wine glass affects the taste of the wine.

Jack believes:

- (4) The shape of the wine glass does not affect the taste of the wine.

After talking with Jack I believe:

- (5) Jack believes (4).

Once I reflect upon my belief and his, I may come to believe:

- (6) I am trustworthy in believing (3).
 (7) Jack's belief (4) arises from the same properties I trust in myself in believing (3).

I may also believe:

- (8) The Principle of Epistemic Trust in Others,
 and I may be conscientious in that belief as well.

Awareness of this set of beliefs produces psychic dissonance and the desire to resolve the conflict by giving up one of the beliefs. Like the case of disagreement over the newly published book, the dissonance might be resolved unconsciously, and that might happen gradually over a period of time. I said in chapter 2 that unconscious resolution of dissonance is our first model of

what rationality is, and we need basic trust in that process. Our second model of rationality is what we do when dissonance is not resolved and we reflect upon it. When I do that, I will attempt to target the belief that should be given up, and to that end I will ask myself which beliefs are more likely to satisfy my future self-reflection. If upon reflection I trust (5), (7), and (8) more than I trust (3) and (6), it is reasonable for me to adopt a skeptical attitude towards (3). I might even come to adopt Jack's belief (4)—if, for instance, I begin to believe that Jack has been more conscientious than I am in his belief, or if I come to believe that he has more relevant evidence or background knowledge than I do. But there are many other possibilities. I might trust (3) and (6) more than (5) and (7), in which case I would have no reason to give up (3) even if I have a high degree of trust in (8). Alternatively, I might trust (3) and (6) more than (8). I have given reasons for adopting some form of the principle of trust in earlier chapters, but there is no reason to assume that my trust in that principle will exceed my trust in every other belief I possess. I have reasons for accepting (8) and may continue to accept it upon conscientious reflection, but maybe I can say the same thing about my belief (3).

It is possible that we think that our trust in normative principles should exceed our trust in our first-order beliefs, but what reason do we have to think that? One answer is that a second-order normative principle such as the Principle of Epistemic Trust in Others is *a priori*, whereas a belief such as (3) is based on testimony or experience, and we may think that neither of the latter is as trustworthy as the former. This idea is arguable, and I will not discuss the general trustworthiness of what appears to be known *a priori* here. However, in the case we are considering, I can see why someone might find (8) more trustworthy than (3) because she trusts the *a priori* basis of (8) more than the particular testimony or experience upon which she bases (3). Nonetheless, I find it doubtful that what seems to the conscientious believer to be known *a priori* is always more trustworthy than what seems to the believer to be known by experience or by testimony. So I doubt that trust in (8) always does or should exceed trust in a first-order belief on the grounds that (8) is *a priori*.

A better reason for taking a second-order normative proposition like (8) to be more trustworthy than a proposition like (3) is that one's belief in a general proposition is typically more deeply imbedded in one's network of beliefs and is itself the product of extensive previous self-reflection, so giving

up (8) has more drastic consequences for one's other beliefs than does a belief like (3). Again, this does not mean that belief in a second-order normative proposition like (8) is always more trustworthy than belief in a first-order proposition like (3), but it does mean that when the conscientious person compares the trustworthiness of two beliefs, she needs to consider how much she trusts any beliefs she would have to give up if she gave up each of these beliefs. In my case, if I gave up (3), I would not have to change many other beliefs, whereas giving up (8) would put me in the position of having to rethink the reasoning I have given in this book that shows me that self-trust commits me to trusting others who possess the same qualities I trust in myself.

But there are also reasons against trusting a very general normative principle such as (8). The more general a principle is, the more likely it is that it is not trustworthy in its full generality. One of the reasons philosophers like to consider particular cases when discussing normative rules is that we sometimes trust intuitions about particular cases more than a general rule or principle, and the intuitions about a particular case can force a modification to a rule we generally trust.

There is another reason to think that a conflict between a first-order belief and the Principle of Epistemic Trust in Others should not always be resolved in favor of the principle of trust. As we have seen in earlier chapters, one of the things I should do as a conscientious person is to imitate epistemic exemplars, persons whose superiority to myself in some range of epistemic behavior is something I come to accept when I am conscientious. If I think an exemplar's belief (3) or (4) is relevant to whether I should believe (3) or (4), then presumably an exemplar's response to a conflict between her belief (3) or (4) and the belief of another person is also relevant to the reasonableness of my response to such a conflict. I cannot consistently treat an exemplar's belief as a *prima facie* reason to change my own belief but not also treat her response to epistemic conflict as a *prima facie* reason to respond in the same way. In my experience, persons I judge to be epistemically superior sometimes behave as if they trust something like the principle of trust more than a first-order belief, but sometimes they do not.

Now let us go back to the conflict between a religious belief like (1) and another person's belief (2). Like the case of a conflict between my belief (3) and another person's belief (4), the conflict arises from the combination of my belief (1), my belief that I am trustworthy in believing (1), my belief that

the other person is just as trustworthy as I am when she believes (2), and the principle of trust or a related principle. Like the case of the conflict with my belief about the relation between the shape of a glass and taste, I am reasonable when I resolve the conflict in favor of what I trust the most when I am being epistemically conscientious. I consider carefully which beliefs will survive future conscientious reflection, and those are the beliefs I trust the most. So the conflict emerges from the following situation.

I believe:

- (1) There is a Trinity of divine persons.

My friend Ali believes:

- (2) There is not a Trinity of divine persons.

So I believe:

- (9) Ali believes (2).

After reflecting about the properties of myself that ground my belief (1) and the similar properties of Ali in believing (2), I may conscientiously come to believe:

- (10) I am trustworthy in believing (1).
 (11) Ali's belief (2) arises from the same properties I trust in myself in believing (1).

And after thinking about the argument of chapter 3, I may also believe:

- (8) The Principle of Epistemic Trust in Others, and I may be conscientious in that belief as well.

If I trust (9), (11), and (8) more than I trust (1) and (10), it would be reasonable for me to change my belief (1). But I might trust (1) and (10) more than (9) and (11) when I am conscientious, in which case change would not be the reasonable thing for me to do as a conscientious person. Of course, I might not trust (11) very much because some of my background beliefs conflict with Ali's and I trust my background beliefs more than his. But reflection would then lead me to formulate the same problem about conflict in

our background beliefs as I have with the conflict between (1) and (9). And again, there are a number of ways I might resolve that conflict, but I still might end up having a lot of trust in (11).

The procedure for resolving the conflict over (1) should be the same as the procedure for resolving the conflict about (3). Even if I trust (9) and (11) as much as (10), I could trust (1) more than (8). I said that a general principle like (8) is typically more trustworthy than (3) because (8) is more deeply connected with other beliefs than (3) is, but is (8) more deeply connected to other beliefs than is (1)? It depends upon the person, of course, but in some cases the answer will be no. When asking myself how much I trust a belief like (1), I need to ask myself how much I trust any other beliefs I would have to give up if I gave up (1). The answer to that question can easily be different than for the parallel question about (3).

Religious disagreement differs from disagreement about the relation between glass shape and taste in another way. To imitate an exemplar includes imitating her attitude towards her beliefs, but as I mentioned, epistemic exemplars do not respond to disagreement in all areas of belief the same way. It would not be reasonable to base a change in my belief on the belief of a person whom I believe to be epistemically superior to myself when that person does not respond to similar cases of disagreement by changing her belief. It seems to me that epistemic exemplars do not change beliefs like (1) or (2) based on reflection about disagreement as often as they change many other kinds of beliefs for that reason, including a belief like (3).

Religious disagreement differs from disagreement in most other areas in a third way. Religious beliefs are commonly the beliefs held by a community, which I have described as an expanded self. People do not form communities around shared beliefs about the shape of wine glasses, although, of course, I am not denying that they often form clubs and associations around wine connoisseurship. The connection of religious beliefs with communities means that the resolution of religious disagreement for an individual depends upon her relation to her community and her reasonable acceptance of the beliefs of that community.

Suppose I acquire a belief from a community justified by my conscientious judgment that believing what We believe will satisfy my conscientious self-reflection better than a belief I get on my own, and suppose I then become aware of a conflicting belief of another person. The task of conscientiously determining what to trust the most requires evaluating the relative trustworthi-

ness of the other person and my community. To do that I must make a conscientious judgment about my future self—what I will believe, the emotions I will have, whom I will admire, and the exemplars I will want to imitate. If I continue to identify with my community, I will generally find my community's belief more trustworthy than the conflicting belief of someone outside the community, but my future self will change with changes in my community, and it is always possible for a conscientious person to cease identifying with a community. Conversion is possible and can be the conscientious thing to do because I might trust aspects of myself that require that I give up membership in my community along with its beliefs because I trust something else in myself more—perhaps admiration and the desire to imitate a person from a radically different culture. I think that that is often the way religious conversion works, and if so, it can be the conscientious way to resolve conflict within the self. But many religious beliefs cannot be changed without radical changes to states of the self—a multitude of beliefs, emotions, desires, and identification with a particular way of life. If so, we would expect that conscientiously resolving conflict within the self by making such radical changes in the self would be rare.¹³

4. COMMUNAL EPISTEMIC EGOISM AND DISAGREEMENT BETWEEN COMMUNITIES

Since I can conscientiously trust my community more than people on the outside, it might appear that the problem of reasonable disagreement is resolved, but it is not. I have argued that a community is like a self, and that means the problem of disagreement between myself and another individual reappears as the problem of disagreement between communities. Communities that embody long historical traditions with well-developed authority structures are no doubt more trustworthy in general than individuals, but we also know that there are religious and political communities whose views are more extreme than the beliefs most of the individual members would have had on their own.¹⁴ From the point of view of people outside the

13. I treat conversion as the decision to adopt an alternate self in Zagzebski (2006: 38–42).

14. James Davison Hunter (2010: 108) mentions this in his discussion of *ressentiment* in American political culture. Both the Left and the Right cultivate a sense of injury, which reinforces group solidarity and can lead to more extreme beliefs.

community, it looks like acceptance of beliefs on the authority of such a community is less conscientious than beliefs acquired independent of the community, and so it looks to us on the outside that the authority of their community is not justified by the justification of communal authority theses of chapter 7. But it is very hard to make that claim without begging the question against them since my justification theses have been given from a first-person perspective. What we call extremist they might consider conscientious. To counter this problem we need to look at the parallel to my argument in chapter 3 that a conscientious person will inevitably come to have beliefs that commit her to trust in others.

I have said that a community is like a self in that it has beliefs, historical memory, and reasoning faculties, and it has self-trust. Communities deliberate according to procedures that the community reasonably believes will satisfy the community's future conscientious reflection as a community. A community trusts itself in virtue of its possession of certain properties or powers. Some of those properties or powers are shared by other communities. Some properties are shared by all other communities—for example, basic human reason—and some properties are shared by all religious communities—for example, having members who have religious experiences. Many religious communities accept the authority of the Hebrew Scriptures. Recognizing the extent to which one's community grounds its beliefs in common human experiences or powers or in shared background beliefs is important because we commit ourselves to trusting other communities to the extent that they share what we trust in our own community.

It seems to me that this principle is often ignored in American politics. Republicans and Democrats feel justified in ignoring or even disparaging views of the other party even when the view is based on principles that both parties share, such as the Bill of Rights or other constitutional principles, rather than on principles that divide them. But to the extent that we trust our own conscientious conclusions from a political principle, we commit ourselves to trusting the conscientious conclusions of a different community on the same principle, and cannot rationally ignore their reflections just because they differ from us in *other* ways, much less because the results of their reflections differ from ours.

The same point applies to religious communities whose doctrines and methods of theological reflection agree more than they disagree. To the

extent that the communities overlap in the beliefs and powers they use in conscientious reflection, the communities are committed to trusting each other. Not to do so is inconsistent with what they trust in themselves.

In chapter 3 I defined extreme epistemic egoism as the position that I have reason to believe p only when the direct exercise of my faculties gives me reasons for p . The fact that another person has a belief p gives me no reason to believe it. The parallel position for a community is the position that We have reason to believe p only when the exercise of the belief-forming procedures within our community gives Us reasons for p . The fact that another community has a belief p gives us no reason to believe it. Let us call this position *extreme communal epistemic egoism*.

Standard epistemic egoism is the position that the fact that another person has a certain belief can give me grounds for believing it, but only if, by using my own faculties, I have grounds for believing the source is reliable. The parallel position for a community is the position that the fact that another community has a certain belief can give Us grounds for believing it only if, by using our communal belief-forming procedures, we can see that the other community is reliable. Let us call this position *standard communal epistemic egoism*. It follows from my argument in this book that both forms of communal epistemic egoism are rationally unsupportable.

When the belief of a community conflicts with the belief of another community, the most reasonable way to respond to the conflict is to engage in conscientious reflection upon the elements of the community that define it in order to determine which elements are more trustworthy than others. Since engaging in conscientious reflection as a community is a good for the community, it should not be considered merely defense against attack. It is hard for anyone to know which beliefs need revision and which do not, which emotions are appropriate and which are not, and which desires are appropriate and which are not without encountering challenges from those with different beliefs, desires, and emotions. What makes this critically important is that we are not able to reflect upon the vast majority of our beliefs, emotions, and desires. We reflect upon them only when faced with dissonance. That was the focus of my remarks at the beginning of chapter 2. The same point applies to self-reflection within communities. For instance, it has been argued that modern Western societies have developed principles of justice that would be reasonable if they were starting from scratch, but those principles perpetuate unjust circumstances when applied to societies

containing peoples who have been historically mistreated.¹⁵ It is unlikely that a community would think of that possibility without challenge from a community supporting different principles. Of course, the objection may be incorrect, but we almost always overlook many things with even the most careful reflection on our beliefs, emotions, and practices, and so it is important to attend to challenges from the outside. The same point applies to challenges to our emotions. A nation may know that a group of people is being slaughtered half a world away, and yet not feel enough outrage to act until they see outsiders reacting with outrage.

The way that challenges are handled within a community is the business of that community, but I think that the falsehood of communal epistemic egoism leads to some general principles of response to disagreement between communities, and I want to propose a few of them.

The first is what I call the *Need to Resolve Conflict Principle*:

It is a demand of rationality for a community to attempt to resolve putative conflicts between its beliefs and the beliefs of other communities.

This principle follows from general trust in the epistemic powers of other human beings and the dissonance that results when a community becomes aware of the conflicting beliefs of another community whose powers it trusts upon reflection. The dissonance is greater when the other community shares a substantial number of background beliefs. In the case of individuals, many disagreements are resolved subconsciously. I am not sure that a community has a collective subconscious, but there is probably something similar, and if so, some disagreements with other communities may be resolved in that way over time, without any effort or even reflective awareness of the change. But many conflicts would need to be resolved in the way I have already described—by communal reflection on the degree of trust that the community has in its conflicting elements.

How strong is the need to resolve conflict? Some conflicts are trivial, and a community can reasonably judge that a problem is not worth the effort to resolve it. Theological details can be in this category, although what is considered a detail at one time in history can be treated as monumentally important at another. Think of the way the disagreement over the “Filioque”

15. See Charles Mills (2009) and other works by the same author.

clause in the Christian creed divided Eastern and Western Christianity, but few theologians or Church authorities now consider it worth so much fuss.¹⁶ I suggest that the degree of the need to resolve a conflict depends upon the degree of the dissonance created within a community by the conflict. The degree to which a community cares about a belief is one dimension affecting degree of dissonance.

The need to resolve conflict is a need to eliminate dissonance within the communal self, so the solution need not be acceptable to the other community. The dissonance can be resolved by changing some of our beliefs, by adding a belief that explains why the dissenting community is mistaken, or by modifying the belief that conflicts with the belief of the other community in a way the members of the other community would not accept but which resolves our own dissonance. All of these responses are common in cases of religious conflict.

Notice that if we resolve the conflict in a way that is not acceptable to the other community, we have not eliminated dissonance entirely, but have replaced it with dissonance of a different sort. So suppose my community believes in the Trinity, Ali's community believes the doctrine is false, and my community resolves the dissonance that arises from our general trust in Ali's community by adopting a belief that explains why Ali's community's belief is sincere but false. Clearly, Ali's community does not adopt that belief, so the dissonance is resolved for our community but not his. This move creates dissonance in another way because Ali's community largely shares the powers and background beliefs trust in which ground our trust in the belief that explains for us why Ali's community is mistaken. But, of course, they do not share all of the powers and background beliefs; that is why they are a different community. That forces us to conscientiously reflect upon the relative trustworthiness of what we share with them and what we do not share. To the extent that we mutually trust shared powers and beliefs (human reason, experiences of the divine, basic theological beliefs, acceptance of certain sacred scriptures), we should trust the results of attempting to use what we mutually trust in an attempt to resolve disagreement to our mutual satisfaction. If this is the conscientious thing to do, a stronger principle than the Need to Resolve Conflict Principle could be supported: It is a demand

16. I thank Anthony Coady for this point.

of rationality for communities to attempt to reach agreement when their beliefs putatively conflict.¹⁷

Of course, it cannot be rational to attempt to do the impossible, but many agreements are possible. The success of the Catholic-Lutheran joint statement on justification was the result of a lengthy process of dialogue, and the two communities started with a vast area of shared beliefs.¹⁸ Obviously we cannot expect the same success with disagreements among the major religions. But I assume that agreement is epistemically preferable to disagreement, and the Need to Resolve Conflict Principle arises from that assumption.

There is a second principle that follows from our recognition of the trustworthiness of our shared human powers, the *Rational Recognition Principle*:

If a community's belief is rational, its rationality is recognizable, in principle, by rational persons in other communities.

The idea here is that if a belief is justified in virtue of the possession of some property shared by others, then those others can see that the belief is so justified. Much has to be packed into the term "in principle" since there are many reasons why a rational person cannot always recognize rationality in others, but I think that the principle is basically correct. There are a number of variations of the principle that I think are defensible. Here are two that naturally fit with the line of argument of this book and which are given in the first person:

Recognition of Trustworthiness Principle

If our community's belief is trustworthy, its trustworthiness is recognizable, in principle, by persons in other communities whose trustworthiness we recognize as a community.

Recognition of Conscientiousness Principle

If our community's belief is conscientious, its conscientiousness is recognizable, in principle, by persons in other communities whose conscientiousness we recognize as a community.

17. In Zagzebski (1999); see also Zagzebski (2001a), where I defend the stronger principle and give it the name of the principle I propose here.

18. "Joint Declaration on the Doctrine of Justification" by the Lutheran Federation and the Catholic Church (1999). It was accepted by members of the World Methodist Council in 2006.

Again, these principles need to be qualified since there are many reasons why our community's belief can be conscientious or trustworthy even though conscientious or trustworthy persons outside our community cannot recognize that fact. But I think that with qualifications, there is *prima facie* reason to expect that the conscientiousness of a belief can be recognized by conscientious persons in other communities. If conscientious persons in other communities cannot recognize our belief as conscientious, there has to be an explanation for their failure to recognize it. We have particular reason to expect epistemic exemplars in other communities to recognize the conscientiousness of our beliefs, and so dialogue between epistemic exemplars in different communities is an important way for both communities to critique their beliefs.

I argued in chapter 3 that we commit ourselves to trust others who have the same properties we trust in ourselves. We may not have a high degree of trust in some of their particular beliefs because of the way those beliefs depend upon other beliefs we do not share, and because the beliefs may conflict with a belief we trust more. But given that we owe them trust, we owe them the components of trust. We ought to treat them as trustworthy, believe they are trustworthy, and feel the characteristic emotion of trust towards them. We owe them that even though we do not owe them trust in certain of their beliefs. Trust goes beyond the basic respect we owe them arising from the moral principle that every human being has dignity. It is a requirement of consistency in trusting ourselves.

I assume that human beings have always been conscientiously reflective, but for millennia communal epistemic egoism was the normal stance for almost everyone. I surmise that that is because they did not typically come to believe that other communities had the same properties they trusted in themselves even though they were conscientious for the most part. The widespread appreciation of the degree to which human beings are alike depends upon contingent features of human history that did not exist until some time in the early modern period. I therefore disagree with those commentators on the phenomenon of disagreement who say that religious writers were always aware of disagreement, and that the only thing different is that some modern writers respond differently to the phenomenon.¹⁹ I am

19. See Plantinga (2000b: chaps. 2 and 3) for an example of this viewpoint.

suggesting that for the most part modern and premodern writers are not responding to the same perceived phenomenon.

Every belief must withstand critical self-reflection. When we take a belief on authority, it has to withstand critical self-reflection. The principle of trust has to withstand critical self-reflection. Judgments of the trustworthiness of others have to withstand self-reflection. The beliefs defining a community have to withstand critical communal self-reflection. One's identification with a community has to withstand critical self-reflection. The ultimate ground is trust in the processes by which the executive self reflects upon itself and changes aspects of the self in order to withstand further self-reflection. But anything can be changed. Disagreement poses a threat to many beliefs, but I am the one who has the beliefs, emotions, perceptions, and other psychic states, and I am the one who exercises executive control over these states. We all know that our control is limited, but I control what is controllable if I have an executive self.

Autonomy

1. THE AUTONOMOUS SELF

1.1. The Norm of Conscientious Self-Reflection

Humans have always been self-conscious. Only self-conscious beings could create art, invent syntax, and worry about the afterlife. At some point we must have become aware of the interesting features of self-consciousness, but we did not turn self-consciousness into a method of philosophy until a few hundred years ago. The story of how we did that is interesting. After Descartes, many philosophers decided that philosophical problems of all kinds could be illuminated by going inward rather than outward—by carefully reflecting on the subject rather than the object of consciousness. The idea was truly revolutionary: You look at the self, not because you are interested in the self, but because you think that if you look at the self, you will understand the universe. And some philosophers were so committed to the method that they decided that if we cannot understand the universe by that method, we cannot understand the universe at all.

The method was interesting because it did not start with the assumption that we are conscious of an object and, *in addition*, we are conscious of ourselves being conscious of an object. The method focused *primarily* on ourselves being conscious of an object, and inquiry into the nature of the world took the form: What kind of object can there be of which we can be conscious? The method made consciousness the basic object of philosophical investigation. Some critics complain that the method is an expression of the modern tendency to value the self over everything else, but I see no reason to think that that is the case. The method requires the obsession with the self, not the reverse.

Consciousness of self-consciousness is also consciousness of point of view. As soon as I become critically aware that I am a self-conscious self, I realize that I look at the world in a distinctive way, and there are many other selves whose conscious view of the world differs from mine. The awareness of this fact, of course, was not new, but when a focus on point of view becomes part of a philosophical method, it is very difficult to avoid the conclusion that no point of view is privileged, and this affects the way we look at a host of philosophical problems. Several times in this book I have remarked on the irony that the modern era combines self-obsession with egalitarianism, but that is what we would expect, given that the focus on self-consciousness cannot avoid focusing on the consciousness of a self *as such*. Egalitarianism, or at least methodological egalitarianism, is the inevitable consequence of such a focus. But that leads to the fear that my self is nothing much—just a very limited and probably distorted point of view on the world outside. Life was easier when people were self-conscious without spending much time thinking about the fact that they were self-conscious.

A self-conscious being reflects. It thinks about what to do, what to believe, whom to trust. A being conscious of being self-conscious also reflects about the process of reflecting about what to do, what to believe, whom to trust, and so on. How should I go about reflecting in general? What is the point of reflecting? Consciousness of self-consciousness adds higher-order norms to the norms of acting, believing, and desiring—norms of reflecting about the norms of acting, believing, and desiring. “Autonomy” was invented as the name for what we do when we are conscious of being self-conscious in this way. Autonomy is the executive self’s management of itself, or what is usually called self-governance. In this book I have proposed that the basic norm of self-reflection is what I call conscientiousness, the property of exercising my faculties in the best way I can to make the outputs of those faculties fit their objects—to make my beliefs true, my desires of the desirable, my emotions appropriate to their intentional objects. Conscientiousness is the higher-order norm of self-reflective beings. It is the norm of autonomy.

My aim as a self-reflective being is to have the right relation to a world outside my mind, but I have argued that critical self-reflection shows me that the ultimate test that my faculties have succeeded in fitting their objects is that they fit each other. It is not as if I can look forward to the day when I will get a final determination of the truth-value of my beliefs, the fittingness of my emotions, and the actual value of anything I value. At least, that will

not happen in this life. Instead, as time goes on, I have more experiences, engage in more reflection, get more testimony from other persons. Some of my psychic states change below the level of conscious reflection, but some of them change only when the executive self makes a higher-order judgment that they ought to change. As I have said, that judgment may not be sufficient to cause a state to change, and when that happens, I will experience dissonance. But if a state survives changes of experience, changes of belief, and further reflection without dissonance, that is the best I can do to tell that the state fits its object. I trust that it fits its object because after reflection it fits my other states.

A state that survives conscientious self-reflection is one that I will not later have to give up due to later dissonance. It is a state I will not regret. It does not lead me away from my goals. It survives criticism from others. If I make a sacrifice for it, I later judge that the sacrifice was worth it. I have proposed in this book that my best judgment of what will survive conscientious self-reflection justifies taking beliefs from others, and it justifies treating a person or a community as an authority. What justifies me now in taking a belief from an authority is that when I am conscientious now, I judge that my future self will retain the belief in spite of whatever changes my self will undergo. I argued that this is rational because rationality is doing a better job of what we do naturally, and we naturally attempt to resolve dissonance in the self. A rational being aims to resolve or prevent both current and future dissonance. An ideally rational person is one in whom dissonance is resolved. But all I can do as a conscientious person is to conscientiously predict the future. If I predict correctly, and a state of belief survives without dissonance, the future justifies my belief. If I predict incorrectly in spite of being conscientious, I will later judge that I made a mistake, perhaps a forgivable mistake, maybe even one I could not have avoided, but nonetheless a mistake. I cannot be ideally rational unless I judge correctly.

That means that being rational is not completely independent of being right. The way we treat the resolution of current dissonance shows that we assume that. When upon reflection we correctly resolve dissonance between, say, two beliefs, or between a memory and a perception, we say we are rational. Why do we say that? I think that the answer is partly that we reflectively attempted to resolve the dissonance and the reflective attempt deserves to be called "rational," but it is partly because we succeeded in doing so. If we resolved the dissonance without reflection, we

would probably judge that what we did was subrational. If we attempted to resolve the dissonance but did it incorrectly, for instance, by replacing one of a pair of conflicting beliefs with a belief that conflicts with a third belief, we would, upon reflection, judge ourselves as less than fully rational. Luckily, at any given time, rationality in the sense of trying to resolve dissonance lines up fairly closely with rationality in the sense of judging correctly what resolves dissonance. We are pretty good at identifying and resolving current and fully conscious dissonance, and so it does not bother us that rationality requires both trying and succeeding. Unhappily, there is a weaker connection between trying to prevent future dissonance and judging correctly what will do so. This is one of the ways in which there is luck in rationality.

Rationality involves luck, but it is not pure luck. There is a connection between what I judge will survive my conscientious reflection in the future and what will in fact survive because what will survive is partly up to me. When there is future dissonance, there is always more than one way to resolve it, and I can resolve it in a way that preserves what I previously judged would be preserved as long as I am willing to pay the possible price of more dissonance elsewhere that will then need to be resolved.

There is a more important reason why there must be a connection between what I judge will survive conscientious reflection and what will in fact survive such reflection. We assume there is a connection between a belief I judge will survive my future reflection and the truth. Since we think there is an even closer connection between a belief that will in fact survive my future reflection and the truth, there must be at least a loose connection between what I conscientiously predict will survive and what in fact will survive. Of course, I have argued in this book that we need trust in the connection between conscientious self-reflection and success in getting the truth, but as long as the connection exists, rationality in its aspect of trying and rationality in its aspect of succeeding are connected.

We can make similar points about desires. Many desires are transient, but others are deep and long-lasting. One would have to be quite a bit more neurotic than average to worry about whether a transient desire survives future conscientious self-reflection, but deeper desires ought to survive such reflection, and they are worth the reflection we devote to them. We want our desires to survive future reflection on them since that is our only way to tell that we desire what is desirable.

I have used the notion of conscientiousness in both a narrower and a broader sense to describe what the reflective self does. Much of the time in this book I have been talking about epistemic conscientiousness, the property of using my epistemic faculties as well as I can to reach my goal of getting truth. But I have also used the notion of conscientiousness in a broader sense, as the property of using all of my faculties as well as I can to make them fit their objects. I have argued that the only way I can tell that I am doing a good job of that is that my psychic states survive both current and future conscientious reflection upon them without dissonance. I think, then, that there is also a broad notion of rationality as well as the narrow notion. A rational person in the narrow sense has cognitive functions that do a good job of getting her to the truth. A rational person in the broad sense is one whose faculties of all kinds do a good job of connecting self and world. The experience of dissonance, reflection upon dissonance, and the attempt to resolve it is the fundamental task of rationality. It is the way human beings adjust the self to rectify mistakes and to bring the self closer to having a fitting connection with the world.

1.2. Autonomy from the Inside and the Outside

In chapter 1 I discussed the fact that although self-governance has been an important value in Western philosophy since Plato, it acquired a new level of significance in the modern period. For philosophers like Plato and Aquinas, reason is naturally and self-evidently authoritative. A person is self-governing because of her share of reason. It is reason *in* the person that is authoritative, not the parts of the person other than reason. The relationship between reason and self-governance changed when Kant identified the source of authority as one's own rational will. As Korsgaard interprets Kant, authority ultimately resides in *my* rational will, rather than *the* rational will. Reason is not the most fundamental bearer of authority. The structure of self-consciousness reveals rules a self-conscious being must accept, and reason just is the rules of reflective self-consciousness. The self-conscious self is the ultimate bearer of authority.

This is one of the ways in which making consciousness of self-consciousness a philosophical method has interesting results—in this case, consequences for the nature of authority. The method makes the authority of the

self more basic than the authority of anything else, including the authority of reason. Authority is intrinsically connected to self-consciousness, and autonomy is the state of exercising the natural right of a self-conscious being to govern itself. In my view the political term “govern” is unfortunate because it sounds as if the self gives itself commands (a consequence Kant accepted, of course). I would not deny that I can command myself to do various things (“Learn Italian,” “Buy a birthday card for Rita,” “Answer an email from Nina”), but self-governance in the sense of commanding myself to do certain things is only one of the many ways a self-reflective being manages itself. The structure of self-consciousness is such that the self has an executive function with regard to itself. What I have called the executive self in this book is the self in its role of manager of itself—the self that changes parts of itself, the self that judges what ought to be changed (whether or not it subsequently changes), the self that follows the basic norm of conscientiousness in attempting to resolve dissonance, thereby becoming a more harmonious self. The executive self has goals, some of which are unconscious. Making up my mind to act in certain ways is only one of the things my executive self does. The authority I have over my acts is just one component of the authority of the executive self over the self.

As I mentioned in chapter 1, autonomy can be understood as a right—a claim to be permitted to govern oneself, or it can refer to the successful exercise of that right—governing oneself correctly. An important reason autonomy became so salient in public discussion in many different fields is that people feared that the executive function of the self can be impaired. When we reflect about the executive self, we notice its fragility. The executive self has to struggle. There are things that inhibit it or interfere with it from both the inside and the outside. That is why Kant thought there are two ways a will can be heteronomous.

From the outside, autonomy is violated when another person imposes her will on me. That could just be ordinary harm, but it is a violation of my autonomy when it prevents me from judging or implementing a judgment of what I should do. We can see from my account of autonomy so far why Kant thought that coercion of my will by another will is not just harmful in the way doing me bodily injury is harmful. It is a violation of my selfhood because it prevents me from exercising the reflective control I have over myself. Coercion can prevent me from acting on what I will. A more intrusive form of coercion prevents me from willing what I would otherwise will.

This kind of attack on autonomy is particularly insidious because interference of the will is much harder for the agent to detect than interference with the execution of her will. For that reason it is much harder to resist and prevent. This observation is familiar from the feminist literature. A woman may not notice that she is unable to make certain choices because of forces in her social or personal situation, choices that she would make if unencumbered.

Heteronomy of the will because of hindrance from the outside leads to the common view that autonomy means noninterference. You respect a person's autonomy by leaving the person alone. Autonomy in this sense has to do with the way people treat *other* people. It is in this sense of autonomy that we think that within limits that do not pertain to the topic of this book, a person should be permitted to make wrong decisions. Autonomy entails that there is a *prima facie* case for others to give me the space to govern my beliefs, desires, and acts independently of what others think about it.

Autonomy looks very different from the inside, and we can see that by looking at the second way that Kant thought a will can be heteronomous. A will is heteronomous in this way when it is governed by something other than reason, when it is determined by inclination or "empirical" causes. Such a will is not managing itself as it would if it were properly reflective. Given the framework of this book, such a person lacks autonomy because she is not conscientiously self-reflective. Conscientious self-reflection takes some effort. If the executive self does not resolve dissonance in a reflective way, either the dissonance will continue, or it will be resolved unreflectively. I have stressed that there is nothing wrong with the unreflective resolution of dissonance. I would not call that heteronomous. We cannot be reflecting on ourselves all of the time, and luckily we often unconsciously do what we would do if we were reflecting. But the more important respect in which my way of looking at autonomy from the inside differs from that of Kant is that I say that we govern all parts of the self, not just our decisions to act. But I have no objections to the Kantian view that autonomy is self-governance by reason as long as autonomy does not require continuous reflection, and reason in the relevant sense includes all the norms of the executive self. It is not limited to the rules that apply to the governance of my acts.

I have argued that the rules of a self-reflective being make it incoherent to rely only upon the direct use of my faculties in forming my beliefs. It follows that epistemic self-reliance is incompatible with autonomy. This point extends to those beliefs upon which I rely in making practical and moral

decisions. But the comparison of autonomy as viewed from the outside and autonomy as viewed from the inside explains why autonomy does have something to do with independence. Autonomy requires that other people treat me a certain way. It requires noninterference from others because other persons encroach on my project of managing myself. From the outside, then, autonomy is noninterference. But from the inside, autonomy is often the choice of interference, done intelligently.

I do not call the executive self a will. The executive self is the self in its function of managing itself, and that includes managing beliefs, emotions, and desires, as well as decisions to act. It includes making higher-order judgments about the components of the self that ought to change. I have said repeatedly that I do not assume that the executive self can always make components of itself change. Often it cannot. But we have both nature and other people on our side. I have stressed the rationality of trust and the conditions for treating communities and other individuals as authorities, but our self-reflective dependence on others also means that other persons can often help us do what we cannot do on our own. They can help us resolve dissonance in many ways—by helping us form a higher-order judgment about what ought to change, by influencing us to acquire the proper motivating emotion, by showing us exemplars of harmonious selves.

For the same reason that autonomy is not epistemic self-reliance, it has nothing to do with individualism as opposed to communitarianism. Other people can either help or hurt us in the project of becoming a fully autonomous self. Communities exist in part to help individuals in the process of reflection leading to a more harmonious self, one that will survive future reflection. I have argued that the authority of communities is justified by my conscientious judgment that the community will do a better job of getting beliefs in some domain that will give me a more harmonious self than I would get on my own. Beliefs taken on authority are justified by the different forms of the principle I have proposed, and are required by the demands of conscientious self-reflection. So the authority of a community for me follows from my autonomy, and in the conditions I set out, autonomy requires authority.

The principles I have defended only apply to those persons who are in the circumstances given by the antecedent of the principles. Perhaps somebody never conscientiously judges that a community will give her beliefs that will survive her own reflection better than a belief she can get on her own. Such

a person will not be justified in taking a belief on authority from a community, but that is not because of the value of individualism or self-reliance. Either that person is not lucky enough to find a community that fits her own beliefs and emotions well enough that she will identify with the community in the future, or else there are such communities, but she distrusts them. A breakdown of trust within communities is another feature of the modern age. People often feel disaffected from those communities that would, in fact, help them in becoming a more harmonious self, one whose beliefs and emotions survive over time, but once an emotional break occurs, it is very difficult to mend it. Resolution of emotional dissonance is much harder to resolve than dissonance in beliefs, and I have argued that trust has an emotional component. There are forces in contemporary society that urge distrust and often succeed. Distrust serves a political purpose. When you can break communities apart, you destroy not only the attachment of their members to the community, but you also destroy the power of the community as an agent in the world. Those who aim to destroy a person's trust in others may have the illusion that they are enhancing the person's trust in herself, but an attack on a person's trust in others is also an attack on her trust in herself. If someone's trust in another is destroyed, she realizes that she cannot trust her own trust; she cannot trust *what* she trusts, and that includes herself. A being without self-trust is not autonomous because the basic norm of self-management is conscientious reflection, and a being without self-trust cannot trust conscientious reflection. The same point applies to a community's trust in itself. An autonomous community in my sense is one that is able to engage in conscientious reflection without internal restraints and without interference from outside communities. Loss of trust within the community prevents the community from acting autonomously since trust is a condition for conscientious reflection within the community.

2. ATTACKS ON THE POSSIBILITY OF AUTONOMY: DEBUNKING SELF-TRUST

I have argued that a community is like a self. It has a history, a memory, background beliefs, sometimes communal emotions, goals for the future, and norms and procedures for deliberating as a community to serve those goals.

We cannot engage in conscientious reflection as a community without trust in what “We” the community do/believe when we are being conscientious. For example, historians agree on their methodological assumptions, in addition to sharing beliefs about certain facts. That is, beliefs about the norms of inquiry as well as factual beliefs about the past and present are part of the shared background of the research conducted within the community. Justification for any new bit of historical interpretation proposed for the acceptance of the community depends upon the acceptance of norms and data that are part of the community’s beliefs. The community’s self-trust in its norms and shared beliefs is required to form new communal beliefs.

For communities, trust in theoretical evidence is derivative from trust in certain modes of belief formation, just as it is for individuals. Changes in what the community trusts lead it to evaluate evidence in new ways. This phenomenon in modern biblical exegesis explains why different communities of interpreters interpret the same evidence differently. For instance, a community that has lost trust in the possibility of miracles will give very different interpretations of the same evidence than communities that believe in miracles or leave open their possibility. In some cases new evidence can lead to a loss of trust in old beliefs, but sometimes the loss of trust in old beliefs is the reason for a certain interpretation of the evidence. The debate over the historicity of Acts of the Apostles is an example. The belief that Luke had apologetic motives in writing Acts has led some commentators to think that its historicity is compromised. And it is maintained by a contingent of interpreters that if Luke was not an eyewitness to Paul’s journeys, his accounts should not be trusted for historical accuracy.¹ This is a debate about trust, not about evidence. Each side accepts the evidence used by the others. Where they differ is on a variety of issues of trust: Can we trust a report of someone who was not an eyewitness? If we find out (or suspect) that a writer has apologetic motives, can we trust his truthfulness? If a historical event is reported that is not confirmed by other sources (e.g., Paul’s letters), can we believe it? What if some event is disconfirmed by other historical reports? Different answers to these questions can split a community apart because they no longer can engage in conscientious reflection as a communal self.

1. See Dillon (1990) and Fitzmyer (1998).

Disagreement over the background belief in the possibility of miracles is an example of one of the milder ways a community can be damaged. The community does not disappear; it just breaks apart. There are other attacks on communities that are aimed at destroying them. The ultimate goal of such debunking strategies is to destroy a community by destroying its trust in itself. The common strategy of casting suspicion on motives often aims not merely at undermining particular beliefs that a community holds in common; its goal is to undermine the trust that holds the community together. Freud's attack on religion is a classic example. If religious belief would satisfy our need for safety and the exorcism of our fears, it is claimed that the belief is not caused by faculties aimed at truth, and so we cannot trust the connection between the dispositions in us that produce the beliefs and the truth. This is not only an attack on a belief, it attacks those aspects of the individual or communal self used in self-reflection. These strategies aim at destroying self-trust, not simply trust in somebody like the author of the book of Acts. Cynicism becomes a methodological presupposition, and this is a problem for all of us because the strategy tends to spread from one community to another.

There are other examples of broad attacks on epistemic trust that allegedly reveal causes of belief independent of truth. Trust in epistemic communities has been attacked for some time on the grounds that communities serve the interests of those in power, even when the communities are epistemic.² Another example is the claim that evolutionary theory reveals that broad classes of our beliefs, perhaps all of them, arise from mechanisms that were selected for survival, not correspondence to reality. These attacks aim at destroying trust, and sometimes they succeed.³

I think it is important to see that these are attacks on autonomy. It is impossible to be self-governing without a substantial amount of trust in oneself. Evidence we trust that indicates that conscientious reflection is not apt to produce true belief in broad areas of belief formation undermines trust in the reflective capacities needed for self-governance. Even if the attack is only directed at one category of belief, for example, moral, political, religious, or

2. One example of such an attack is Jim Tiles and Mary Tiles (1993). See also Linda Martin Alcoff (1991–92).

3. For example, Pinker (1997) and (2005). Barrett (2000) and Boyer (2003) focus on religious beliefs.

philosophical beliefs, loss of trust in the faculties and dispositions that lead to beliefs in that category can easily lead to a loss of trust in the faculties that lead to beliefs in many other categories because they are often the same faculties. We do not, after all, have very many belief-forming faculties, and if we have reason to think these faculties are untrustworthy in their deliverances in one area, there is reason to suspect their deliverances in general. At the worst, the evidence casts doubt on the trustworthiness of the norm of conscientious self-reflection—on the connection between following the norm and getting the truth. We cannot be self-governing if we doubt the basic norm of autonomy.

Challenges of this kind highlight the importance of reflection on trust—what it is, what grounds it, and what sorts of things can reasonably undermine it. To properly interpret and conscientiously respond to attacks on self-trust, it is important to distinguish arguments using theoretical, third-person reasons from those using deliberative, first-person reasons. Debunking arguments typically offer theoretical reasons and they need to be treated as such. Attacks on religion from Freudian psychology and evolutionary biology, and some (but not all) arguments for the political motives of epistemic authorities are theoretical. But as I have argued, theoretical evidence is not sufficient to undermine my first-person, deliberative reasons for belief.

The gap between theoretical evidence and first-person reasons is easier to see if we start with an example in which the belief in question is about something over which I have the control of an agent. I have heard it said that people who believe their marriage will survive in full awareness of the fact that 50 percent of marriages end in divorce are irrational. They allegedly exhibit the triumph of hope over reason. The objector believes that people ought to have weakened trust in their own marital relationship on the basis of the statistics on failed marriages. The objector recognizes that most people do not respond to the statistics that way, but rather than conclude that these people must be taking into account reasons he has overlooked, the objector concludes that most people are irrational. But we do not trust our emotions or beliefs, much less our vows, on the grounds that a high percentage of people who fit a certain general description have fitting emotions or true beliefs or vows they keep, and we fit the description. To think that way is to treat oneself in the third person, and it shows a singular lack of awareness of what a conscientiously reflective person does with theoretical evidence that may apply to herself.

When the issue is my belief in the future of my relationship, what is much more relevant than statistical probability is the testimony of a friend about myself and my relationship. What a person who knows me well tells me gives me deliberative reasons to form beliefs that are relevant to my conscientious reflections about myself in a much more direct and trustworthy way than the theoretical reasons of statistical evidence about married couples in general. Furthermore, first-person knowledge of one's own future decision and the decision of a person one trusts is not merely predictive. We do not predict what we will do in the future; we form intentions and change them only because we form other intentions, intentions that arise out of further conscientious reflection. The best way to know our own future decisions and the decisions of those close to us is to be fully cognizant of our current conscientious reflections as well as theirs. Of course, when we do that, our belief in the future state of our relationship can go either way. We may become more confident or less confident than 50 percent.⁴

The situation is somewhat different when the theoretical evidence is about something less directly connected with our agency. A well-known body of evidence shows that people in various studies are inclined to commit a number of inductive fallacies, including errors of assessing probability and the fundamental attribution error (e.g., "Since my customer pointed out a mistake I made in her favor, she must be an honest person"). The theoretical evidence of these fallacies may or may not apply to me, but it does count for something. Theoretical evidence is evidence, and as a conscientious person I must deal with it. Once I am aware of the common tendency to commit such fallacies, I am less likely to commit them, but whether I committed the fallacies before I reflected depends upon many factors I use in my self-reflections. To think that if 80 percent of people in a study commit a certain inductive fallacy, there is an 80 percent chance that I do also is too facile. But it is reasonable for me to engage in the following first-personal line of reasoning: Persons who are relevantly like myself trusted their faculties of reasoning and nonetheless made errors identified by others whom I trust. When the mistake is revealed to those who err, they should (a) rectify the

4. Richard Foley (2001: chap. 3) devotes a chapter to empirical attacks on self-trust, arguing that evidence similar to some of the evidence I am discussing here raises first-person issues for belief. For the reasons I have given in this book, I think that the gap between theoretical and deliberative reasons is important, and it affects the way this problem should be handled.

error in the future, and (b) have less trust in their reasoning abilities since it is likely that they make other mistakes of which they are unaware. Since I am relevantly like them, I also may make either the mistake identified in the study or a similar one, and I also should (a) identify and rectify any such error, and (b) have less trust in my reasoning abilities than I did before.

The compatibility of (a) and (b) in both cases requires that the decrease in self-trust dictated by (b) is not enough to prevent the agent from acting on (a). A conscientious person ought not to be surprised that she makes mistakes, although she might be surprised that she makes a particular mistake. If she can identify and remedy her mistakes, her self-trust should increase. But it is not reasonable for her to let a small amount of empirical evidence undermine her confidence in her conscientious reflection. At all times she needs to be aware that her trust in the former depends upon her trust in the latter.

Unfortunately, there is a wide range of mistakes that are not as easily overcome as inductive fallacies. Everyone agrees that what is commonly called "wishful thinking" is irrational. A conscientious person accepts the norm that we should not believe something because we want it to be true, yet almost everybody does that. For instance, people are much more likely to believe a damaging report about an acquaintance or political figure whom they dislike than about one they like. Wishful thinking also explains the selective attention to theoretical evidence. Most of us read and listen only to news outlets that reinforce our beliefs, whether it is Fox News, the *New York Times*, or publications and electronic sites sponsored by special interests.⁵ I do not know if any studies have focused on professional philosophers, but I would not be surprised if philosophers were not much better than other people in these respects.

These mistakes are harder to overcome than inductive fallacies because the beliefs are partly motivated by emotions. But wishful thinking can usually be made conscious, so the reflective person can identify it. As long as she can identify it in herself, this form of irrationality is not very threatening to her self-trust. The relevance of theoretical reasons to self-trust is more

5. I suspect that the desire to avoid dissonance is part of the reason why people usually read or listen to people with whom they know in advance they are likely to agree. It is not irrational to avoid dissonance. The problem is that we tend to treat what we read and hear as theoretical evidence in support of our beliefs that we can then use in discussion with other persons.

interesting when there is empirical evidence that not only shows that persons in certain situations are apt to act in ways a person would want to reject when conscientious, but which also explains why such persons, upon reflection, are apt not to notice what they are doing. Evidence of subliminal racism is an example of this. In one study, white participants connected to an fMRI were shown pictures of faces for such a short amount of time that their conscious mind could not process it, although they were subliminally aware of it. When they were shown a picture of a black face, their amygdala lit up much more than when they were shown a picture of a white face. Since the amygdala is the part of the brain that processes fear and anxiety, the conclusion was that these people subconsciously fear black persons (Cunningham et al. 2004).

This one is harder to handle with self-reflection since the evidence pertains only to what people do below the level of conscious reflection. When they consciously reflect, they may experience dissonance at the thought that they might unconsciously feel the subliminal fear described in the study. The attempt to resolve the dissonance depends upon the belief that there are ways to consciously control unconscious reactions. If there is no such connection, there is nothing to be done about it from the first-person standpoint. All I can conclude from the study is that there might be something I do of which I am not consciously aware that would create dissonance with some of my beliefs and values if I were consciously aware of it. But the possibility itself may create dissonance in the same category as that created by the Freudian hypothesis that most people have many unconscious motives, fears, and desires, some of which are unconscious precisely because we would have trouble handling the dissonance if we became conscious of them. Apparently, the human psyche protects itself from dissonance it cannot resolve by making some of what would create the most dissonance unconscious.

By hypothesis we cannot conscientiously reflect directly on the subconscious mind, but we can conscientiously reflect about the evidence of a subconscious and its effects. It is also relevant to reflect about the way other conscientious persons respond to such evidence. In my experience, when conscientious persons come to believe that the above hypotheses apply to themselves, it is rarely because of theoretical evidence supporting the hypotheses, but because of their own reflections about themselves and what someone else tells them about themselves (e.g., a psychotherapist), which

gives them deliberative reasons to consider these hypotheses appropriate in their own case.

It seems to me that unconscious states that have no effect on conscious beliefs, feelings, or acts are psychologically inert, so a belief in their existence should not create dissonance, and the mere possibility of their existence should cause even less dissonance. Evidence of subliminal racism that has no conscious effects would be in that category. Evidence of subconscious motives that cause behavior or conscious feelings needs to be handled differently, but in that case there is something in the conscious self upon which the conscientious person can reflect, and upon which she can make higher-order judgments. If she has feelings that create dissonance with her values, she has reason to reflect upon them anyway, apart from any hypothesis about their cause. If there are no conscious feelings or behaviors that create dissonance with her values, there is nothing for her to reflect upon.

The epistemic form of this hypothesis is that subconscious motives cause one's conscious beliefs. If the hypothesis is true, the causes of the belief are independent of the truth of the belief. The Freudian hypothesis about religious belief is in this category. However, we know when we are in the process of conscientiously reflecting, and when we do so, we can tell whether our beliefs are harmonious with deliberative reasons and theoretical evidence, or whether instead there is dissonance between the evidence and the belief. If the belief survives conscientious self-reflection, it does not matter whether the belief was originally acquired through desire, instinct, testimony, personal experience, deduction from some other belief, or by some other mechanism. Like all beliefs, our only way to tell that religious beliefs are true is that they survive conscientious reflection. The same point applies to belief in the Freudian hypothesis. That also will not be adopted by the conscientiously reflective person unless the hypothesis survives conscientious reflection. It might or it might not. If it does survive, it ought not to reduce trust in those religious beliefs that also survive conscientious reflection. That would be a case in which the conscientious person would judge that her religious beliefs are true even though she judges that the Freudian hypothesis is also true.

We ought to treat the evolutionary hypothesis about religious beliefs the same way. Ultimately, we do not have reason to think that the hypothesis is true unless it survives conscientious self-reflection, including reflection upon both our deliberative and theoretical reasons for belief in the

hypothesis. Since the reasons in support of the hypothesis are theoretical, and theoretical reasons are derivative from self-trust, there is a limit to the credence we can put in the theoretical reasons that undermine self-trust without the reasons being self-defeating. In any case, belief in the evolutionary hypothesis about religious belief will not survive the conscientious reflection of many persons because the hypothesis creates dissonance with many categories of belief: obviously religious beliefs, but also metaphysical beliefs, and very general scientific theories, including, some would say, the general belief in evolution.⁶ Furthermore, as I have stressed, conscientious reflection includes reflection about the way persons we trust when conscientious respond to a belief. If such persons either reject the evolutionary hypothesis about the origin of religious beliefs, or accept it but do not respond by mistrusting religious beliefs, that gives us a deliberative reason to think that such a hypothesis does not undermine trust in religion for the conscientious person.

A harder case for the conscientious person to handle is one in which the reasons against self-trust are deliberative. Many people who live for a time in another country, or study their wisdom literature in depth, find that their trust in their own beliefs is undermined. It is common to think, "I would have had different beliefs if I had grown up in a different place, and it is an accident of history that I have the beliefs I have. I could have been Hindu, Muslim, Buddhist, Christian, atheist, or many other things." The same line of thought applies to philosophical positions and attitudes about political arrangements. I am a believer in libertarian free will but I could have been a determinist. I am a believer in Western democracy, but I could have believed in Islamic theocracy. This is a first-person attack, and so it is more serious than the third-person debunking arguments we have considered above. It is not just reason to believe that people in general form beliefs in certain ways. It is an argument that *I* could have developed a very different self than the one I have, one in which a different set of beliefs and perhaps values directs my life. Unlike attacks from Freudian psychology and evolutionary biology, this attack does affect people's conscientious reflection upon their own beliefs, and unlike the hypotheses discussed above, conscientious persons *do* respond to this line of thought by feeling a loss of trust in their beliefs,

6. A well-known version of this point has been defended by Alvin Plantinga (2002a).

although I do not know how many people actually give up beliefs because of these experiences.

This leads to a dilemma of self-trust familiar from our discussion in chapter 10. On the one hand, I realize that persons whom I trust because of their similarity to me have very different beliefs and values, and that creates a problem for me within the things I trust. But on the other hand, I realize that I could not be a different person. When I look back at my life and the way my self has developed through a long sequence of experiences, emotions, and the acquisition of beliefs, some of which change over time, I realize that I could have gone off on a different track at various points. But I am not what I could have been; I am what I am. I must admit that I have the ability to change, and it is possible to radically change my beliefs. Conversion can be the conscientious thing to do for some people, but only when there is something they *do* trust (not *might have* trusted) that they trust more than all the beliefs they must give up. I suggested that that is often admiration for an exemplar. But they always change because of something in the self, not because of something that might have been in the self.

The awareness that conscientious persons can radically differ from me and from each other in their beliefs is humbling because it shows that a multitude of selves with the same natural faculties and the same ends, and with the same conscientious use of those faculties, end up with different results. It is often said that the difference is due to a difference in experiences, but I think that is the wrong way to look at it. For one thing, experiences are not that different, at least not different enough to explain the vast difference in beliefs. But in addition, much of the difference in experiences can be explained by differences in beliefs, not conversely, because people seek experiences and interpret the experiences they have in ways that make them harmonious with their preexisting beliefs, emotions, and values. But as I have argued, a partial change in my beliefs will often result in a less harmonious self than one that either does not change or that changes radically. Given that I do not have reason to make a radical change in my beliefs except in the special conditions in which conversion is rational, the realization of the vast differences in beliefs and values among persons gives me no reason to change. A self that doubts is less harmonious than one that has beliefs. But humility does change the self. Humility is fundamentally the realization of what a self is. We all engage in the task of creating a harmonious self, and we want it to be in contact with reality.

Perhaps we actually like the differences between ourselves and others because we think we would not be the self that we are if we were like other selves. Yet the fact that we attempt to make our states fit the world pushes us towards attempting to make our states match the states of other people. The discovery of a mismatch between our states and the states of others is therefore threatening even though we would not want to be a perfect match with any other self. Herein lies a puzzle for the self: We think there is something wrong about a clash of beliefs, emotions, and values with those of others, and yet we are also sure there is something right about it too.⁷ The autonomous person not only has to govern her states in an attempt to serve a dual purpose—make them fit the world and make them fit each other, but the autonomous person has to figure out the place of selves in the universe. The goal of autonomy includes attempting to figure out an answer to that question; it is very hard to govern ourselves without understanding what we are governing. That question also is one that is best answered with help from others.

3. EPISTEMIC AUTHORITY FROM THE OUTSIDE

Resolving dissonance can be difficult to do. Other people usually do not know what I should do to resolve my own dissonance. When they do, if they try to impose it on me, that almost never resolves dissonance because coercion itself produces dissonance even when what is coerced would produce harmony. That is one of the reasons why autonomy from the outside generally means leaving people alone. But I argued at the end of chapter 6 that if we accept the first-person version of the justification of authority theses, we probably think that it is not the actual making of a conscientious judgment that justifies authority, but rather the truth of what we conscientiously judge. That is implied by our practices of justifying ourselves to another person when we take a belief on authority. When challenged, I typically do not point out that taking someone as an authority satisfies JAT 1 for me. Rather, I give him reasons to think that it is true that taking a

7. My own view is that each person is ontologically unique or incommunicable, and that a fitting relation between one self and reality is not identical to a fitting relation between another self and the same part of reality. I argue for the incommunicability of persons in Zagzebski (2001b) and (2004a: chaps. 5–6).

belief from the putative authority is more likely to get me to the truth than if I try to figure it out myself. In doing so, I imply that I am using the third-person version of JAT 1:

Third-person JAT 1 thesis

The authority of a person's testimony for me is justified by the fact that I am more likely to satisfy my desire to get true beliefs and avoid false beliefs if I believe what the authority tells me than if I try to figure out what to believe myself.

Similarly, if I take someone as an authority because I conscientiously judge that doing so makes it more likely that I will get beliefs that survive my own conscientious self-reflection, I will justify that to someone else by arguing that taking him as an authority makes it more likely that I will get beliefs that survive conscientious self-reflection. If I mention that I made a conscientious judgment that that is the case, that is only as a way to get him to agree that it is probably the case. But I do not expect him to think that the mere fact that I made a conscientious judgment of a certain kind is what justifies me in taking the authority as authoritative. It is the probable *truth* of the judgment that is doing the justifying work. That implies that I am assuming that the following version of the JAT 2 thesis is correct:

Third-person JAT 2 thesis

The authority of a person's testimony for me is justified by the fact that I am more likely to get a belief that will survive my future conscientious self-reflection if I believe what the authority tells me than if I try to figure out what to believe myself.

Since I can point to the third-person versions of JAT 1 or 2 to show that I am justified in taking someone to be an authority, another person can sometimes show *me* that I am justified in taking someone as an authority because *he* recognizes that the conditions of the third-person JAT 1 or 2 obtain. If I conscientiously accept what he says, that will lead me to see that the putative authority satisfies the first-person JAT 1 or 2 for me, but the most natural thing for me to judge in such a case is that the putative authority was already authoritative before I made the judgment. What my interlocutor did for me was to show me something that was true before I recognized it. And, of course, I can

discover myself that A satisfied the conditions for authority long before I recognized A as an authority. If so, why should my *actual* judgment matter?

But it does matter, and to see why it matters, let us look again at how a conscientious person manages the self. Suppose I find out that A was always more likely than I am to get the truth in some domain, or suppose I judge that A was all along better than I am at getting beliefs in some domain that survive my own conscientious reflection. What will I judge about my past self? Surely I will judge that I should have believed that A satisfied the conditions for authority and, for that reason, I should have taken various beliefs on A's authority. But I will not judge that I should have taken beliefs on A's authority in the absence of a conscientious judgment that A is better than I am in the relevant respects. If I had done that, there would have been dissonance in the self. I would have been taking a belief on authority without a conscientious judgment that the authority is authoritative, and that is unreasonable.

So under the imagined conditions it is true that I ought to have done the following: (a) recognized A's authority, and consequently (b) taken various beliefs on A's authority. But it is false that I ought to have taken various beliefs on A's authority. It does not follow from the fact that I ought to do x and y, that I ought to do y. The fact that I ought to do y can be dependent upon my doing x.

This shows that even if the third-person JAT theses are true, they do not threaten our autonomy. The autonomous person ought to accept authority when it satisfies the conditions given in one of the third-person JAT theses, and *when she does so*, she ought to accept particular beliefs on the authority's authority. It is still the case that the ultimate authority is the self even though an autonomous person is always governed by the norm of truth. The autonomous person must submit to the truth, but it does not follow from that that an autonomous person ought to accept certain truths in the absence of accepting certain others.

At the end of chapter 6 I said that it can be justified for an authority to do something very much like commanding belief and for the subject to do something very much like obeying the authority. The considerations above show why this does not interfere with the autonomy of the subject even though no authority may coerce belief. There is a fact of the matter regarding the persons or communities I ought to take as authoritative, and some people are better than I am at identifying who those persons or communi-

ties are. That is, some people are better than I am at making the second-order judgment that a certain person or community satisfies a third-person version of the JAT theses. Since I already know that some people are better than I am at making first-order judgments, it should not be surprising that some people are better at making the second-order judgment. Accepting their judgment does not reduce the authority of the self. On the contrary, it means an autonomous self has more work to do in managing itself. It must be alert in identifying epistemic authorities, and it must also be alert in identifying those persons who can identify epistemic authorities. It is humbling to realize that I can be mistaken in my judgment about myself. That is partly because my self is constrained by the world, not just the world as perceived by me. It is also because my future self is constrained by my past and present self, but the latter is not completely transparent to the executive self at any given time.

The ancient monastic dictum that the task of the self is to withstand the temptation to follow one's own "will" is correct. That is not because the self is worthless, but for the opposite reason. The self is important enough to be aided in its task by wise, knowledgeable, sensitive, and more experienced others. One's current point of view is not a dependable way to accomplish that task. Rather than to conclude that autonomy is not valuable, we should conclude that autonomy requires dependence on others. In the epistemic domain it requires believing certain things because an authority tells me to do so. In the practical domain it no doubt requires doing certain things because an authority tells me to do so.

At the end of chapter 7 I proposed the following *General Justification of Authority Thesis* (GJA):

The authority of another person is justified for me by my conscientious judgment that if I do what the authority directs (or believe what the authority tells me), the result will satisfy my own conscientious self-reflection better than if I try to figure out what to do/believe myself.

The argument of this section leads to the following third-person version of the thesis:

Third-person GJA thesis

The authority of another person for me is justified by the fact that if I do what the authority directs (or believe what the authority tells me), the

result will satisfy my own conscientious self-reflection better than if I try to figure out what to do (or believe) myself.

Although the topic of this book is not practical authority, I propose that the third-person GJA thesis is the most basic principle justifying authority in both the practical and epistemic domains. It also can be expanded to include the justification of taking an emotion on authority, discussed in chapter 4.

4. SELF-FULFILLMENT

Being conscientious is the norm for change and development of the self, but it is not the goal of the self. Satisfaction of conscientious reflection is not the end. Rationality is not the end. Satisfaction of one's desires is a better candidate for the end, but it is doubtful that all desires need or even ought to be satisfied. I have argued in this book that we manage the self in such a way that we aim to have a harmonious self properly connected to reality, but I am not suggesting that either the harmony of the self or its proper connection to reality is the ultimate end. In fact, it is very unlikely that it is. It may be self-evident that harmony is better than disharmony, and I have relied upon the intuition that it is good to desire the desirable, believe what is true, and have emotions appropriate to their objects, but these goals leave some of the most important questions unanswered, indeed unasked. What does it mean for the self to be fulfilled? For instance, when people say they find fulfillment in a romantic relationship, how does that differ from saying the relationship satisfies their desires and they find joy in it? People say very similar things about very different life commitments. A life devoted to art or philosophy or political engagement, or a consecrated religious life can be fulfilling. What is the self that is thereby fulfilled, and what does it mean to fulfill it? I do not have an answer to that question, but I think there is one. In this book I have given a lot of attention to negative self-management—resolving dissonance, and I have also given attention to positive self-management in the modest sense of having psychic states that fit their objects. But I have said nothing about the more glorified goal of self-fulfillment.

At the beginning of this chapter I remarked that when philosophers turned consciousness of self-consciousness into a philosophical method,

that affected the way we formulate and attempt to solve many philosophical problems. Traditionally, philosophers talked about persons, not selves. As I said at the beginning of chapter 2, I assume that a self is the inner life of a person. It is a person from the inside viewpoint. A philosophical inquiry into the self has an egoistic focus even if I am right that it cannot consistently remain egoistic. But it is plain that the concept of a self raises an egoistic problem, whereas nobody would say that the concept of a person raises such a problem. My purpose in this book has been to show that even if we use a modern methodology focused on the self, we get a traditional conclusion: We ought to take certain persons and communities as epistemic authorities. This traditional conclusion does not violate our autonomy, and in fact, is a rational requirement of autonomy. But there is another way in which the focus on the self leads to a different way of formulating philosophical issues. The ultimate end of human life looks very different if we start with the notion of a self rather than the notion of a human person. Whereas Aristotle talked about *eudaimonia*, or the fulfillment of human nature, modern philosophers talk about self-fulfillment, the fulfillment of a self. Of course, philosophers still talk about human well-being, but when they do so, it often looks just like self-fulfillment. Fulfillment of the self is parallel to the Aristotelian idea that we have by nature a set of potentialities, and human well-being is a matter of actualizing those potentialities. The difference is that for the modern philosopher, the potential does not come from our nature, but from the self. This is mysterious because there are no potentialities of a self *as a self* to be actualized or fulfilled, so the sense in which a self is fulfilled cannot be the same as the sense in which a nature is fulfilled.

However, there is one classical view of happiness that I think comes close to what a modern philosopher wants in the notion of self-fulfillment. Aquinas says that the perfect good for a person is whatever wholly satisfies his will. The natural desire for happiness is the desire that one's will be completely satisfied (*ST I-II*, q. 5, a. 8, corpus). Suppose we assume that there is such a thing as self-fulfillment, and that it is something like the perfect satisfaction of the will. What does that do to the arguments of this book? Is self-fulfillment a state that must be brought about directly by the self? If so, self-reliance in all domains, including the epistemic, might seem to follow. I think, however, that it leads us to another way of defending authority. In chapter 7 I mentioned St. Benedict's way of justifying the practice of his *Rule*. In the prologue he asks whether the reader desires days of real

fulfillment, and the rule that follows is intended to satisfy that desire.⁸ Whether it does so is for those who follow it to say, but it seems to me that the implied justification for authority is sound. If I conscientiously judge that acting or believing on the authority of another will lead me to self-fulfillment, surely I am justified in acting or believing on authority in such a case.

The argument for epistemic authority I have given arises from the conviction that in spite of the delight we take in individual differences, human selves are similar enough that what other persons have discovered in their task of directing themselves often applies to myself, and if I am honest, I will realize that. Some of these discoveries are embodied in communities whose task is to transmit them over time and space. The self-reliant person who rejects the help of others in the task of governing herself will have a much harder time than the person who conscientiously accepts that there have always been persons wiser than herself, and the accumulated wisdom of many persons extends to beliefs, desires, and values she would do well to adopt herself. The trust needed to make such a judgment is justified by what happens after the authority is accepted. Does accepting authority lead to fulfillment? If it does, it is surely justified. The ultimate test that anything we do or believe is right is our later judgment that it succeeded. I have stressed future success in the sense of surviving conscientious reflection, but for many things in our lives, we cannot tell that we are on the right path until we see the goal at hand. The justification of trust is that we later see that we got something with it that we could not have gotten without it. This is another way in which the future justifies the past.

The same point may apply to the development of human history. What justifies particular economic and political arrangements is that they are conditions for something else we deeply desire, but we cannot see whether that is true until those arrangements have existed long enough that their effect

8. In paragraph 4, Benedict quotes Psalm 33: "*Quis est homo qui cult vitam et cupit videre dies bonos?*" The most direct English translation would be something close to the following: "Is there anyone who yearns for life and desires to see good days?" But Patrick Barry (Benedict 2003) translates it this way: "Who is there with a love of true life and a longing for days of real fulfillment?" Barry makes the question more modern, and the justification of Benedictine authority that follows is therefore based on a reply to the person who is searching for "fulfilling" days rather than "good" days. Although we may doubt whether the sixth-century St. Benedict intended to show prospective monks how to be self-fulfilled, the implied rule of justification of authority seems to me to be sound.

on the human self is apparent. For centuries we have tried to deny that authority is one of the conditions for human fulfillment because so many times in the past authorities were oppressive. But we are now seeing some of the destructive effects of the lack of authority on human life. In my opinion, epistemic authority needs to be regained, but in a modern form. In this book I have tried to show what a modern defense of epistemic authority would look like.

- Adams, Robert M. 2002. *Finite and Infinite Goods*. Oxford: Oxford University Press.
- Adler, Jonathan. 1997. "Lying, Deceiving, or Falsely Implicating." *Journal of Philosophy* 94:9, pp. 435–52.
- Alcoff, L. M. 1991–92. "The Problem of Speaking for Others." *Cultural Critique* 20, pp. 5–32.
- Alcoff, L. M. 2003. *An Introduction to Historical Epistemology*. Cambridge: Basil Blackwell.
- Alston, William P. 1986. "Epistemic Circularity." *Philosophy and Phenomenological Research* 47, pp. 1–30.
- Alston, William P. 1991. *Perceiving God: The Epistemology of Religious Experience*. Ithaca, NY: Cornell University Press.
- Alston, William P. 1999. "Back to the Theory of Appearing." *Nous* 33, pp. 181–203.
- Alston, William P. 2005. *Beyond Justification: Dimensions of Epistemic Evaluation*. Ithaca, NY: Cornell University Press.
- Anscombe, Elizabeth. 1978. "On the Source of Authority and the State." *Ratio* 20, pp. 1–3.
- Anscombe, Elizabeth. 1979. "What Is It to Believe Someone?" In *Rationality and Religious Belief*, edited by C. F. Delaney. Notre Dame, IN: University of Notre Dame Press.
- Anscombe, Elizabeth. 1981. "Authority in Morals." In *Ethics, Religion, and Politics*. Vol. 3 of *Collected Philosophical Papers*. Oxford: Blackwell.
- Anscombe, Elizabeth, ed. 2008. *Faith in a Hard Ground: Essays on Religion, Philosophy, and Ethics*. Charlottesville, VA: Inprint Academic.
- Antony, Louise. 1995a. "Comment on Naomi Schleman." *Metaphilosophy* 26, pp. 191–98.
- Antony, Louise. 1995b. "Sisters, Please, I'd Rather Do it Myself: A Defense of Individualism in Feminist Epistemology." *Philosophical Topics* 23, pp. 59–94.
- Appiah, Kwame A. 2008. *Experiments in Ethics*. Cambridge: Harvard University Press.
- Aquinas, Thomas S. 1948. *The Summa Theologica of St. Thomas Aquinas*. New York: Benzinger Brothers.
- Ardelt, Monika. 2004. "Wisdom as Expert Knowledge System: A Critical Review of a Contemporary Operationalization of an Ancient Concept." *Human Development* 47, pp. 257–85.

- Arendt, Hannah. 1968. "What is Authority?" In *Between Past and Future*. New York: Penguin.
- Aristotle [384–322 B.C.E.]. 1984a. "Metaphysics." In *The Complete Works of Aristotle*, vol. 2, edited by J. Barnes. Princeton, NJ: Princeton University Press.
- Aristotle [384–322 B.C.E.]. 1984b. "Parts of Animals." In *The Complete Works of Aristotle*, vol. 1, edited by J. Barnes. Princeton, NJ: Princeton University Press.
- Audi, Robert. 2006. "Testimony, Credulity, and Veracity." In *The Epistemology of Testimony*, edited by J. Lackey and E. Sosa. Oxford: Oxford University Press, pp. 25–49.
- Austin, John L. 1946. "Other Minds." *Proceedings of the Aristotelian Society, Supplementary Volumes*, 20, Logic and Reality, pp. 148–87.
- Ayer, A. J. 1935. "The Criterion of Truth." *Analysis* 3, pp. 28–32.
- Baehr, Jason. 2009. "Is There a Value Problem?" In *Epistemic Value*, edited by A. Haddock, A. Millar, and D. Pritchard. New York: Oxford University Press.
- Baier, Annette. 1995a. "Trust and Anti-Trust." In *Moral Prejudices: Essays on Ethics*. Cambridge: Harvard University Press.
- Baier, Annette. 1995b. "Sustaining Trust." In *Moral Prejudices: Essays on Ethics*. Cambridge: Harvard University Press.
- Baltes, P. B., and Jacqui Smith. 2008. "The Fascination of Wisdom: Its Nature, Ontogeny, and Function." *Perspectives on Psychological Science* 3, pp. 56–64.
- Barrett, Justin. 2000. "Exploring the Natural Foundations of Religion." *Trends in Cognitive Sciences* 4:1, pp. 29–34.
- Barsade, Sigal. 2002. "The Ripple Effect: Emotional Contagion and Its Influence on Group Behavior." *Administrative Science Quarterly* 47, pp. 644–75.
- Benedict XVI, Pope. 2007. *Jesus of Nazareth*. Translated by Adrian Walker. New York: Doubleday.
- Benedict of Nursia. 2003. *The Rule of St. Benedict*. Translated by P. Barry, in *The Benedictine Handbook*. Collegeville, MN: Liturgical Press.
- Benson, Hugh. 2000. *Socratic Wisdom*. Oxford: Oxford University Press.
- Benson, Hugh. (In progress). *Clitophon's Challenge: Platonic Dialectic and the Philosophical Method in Plato's Classical Dialogues*.
- Benson, John. 1983. "Who Is the Autonomous Man?" *Philosophy* 58:223, pp. 5–17.
- Berlin, H. A., Udo Kischka, and E. T. Rolls. 2004. "Impulsivity, Time Perception, Emotion and Reinforcement Sensitivity in Patients with Orbitofrontal Cortex Lesions." *Brain* 127:5, pp. 1108–26.
- Blackburn, Simon. 1993. *Essays in Quasi-Realism*. Oxford: Oxford University Press.
- Bluck, R. S. 1963. "'Knowledge by Acquaintance' in Plato's *Theaetetus*." *Mind* 72:286, pp. 259–63.
- Bochenski, Joseph M. 1965. *The Logic of Religion*, New York: NYU Press.
- Boyer, Pascal. 2003. "Religious Thought and Behaviour as By-Products of Brain Function." *Trends in Cognitive Sciences* 7:3, pp. 119–24.
- Broncano, Fernando. 2008. "Trusting Others: The Epistemological Authority of Testimony." *Theoria* 61, pp. 11–22.
- Buber, Martin. 2000. *I and Thou*. New York: Scribner Press.
- Bulbulia, Joesph. 2004. "The Cognitive and Evolutionary Psychology of Religion." *Biology and Philosophy* 19:5, pp. 655–86.
- Burge, Tyler. 1993. "Content Preservation." *Philosophical Review* 102:4, pp. 457–88.

- Burnyeat, M. F. 1980. "Aristotle on Learning to Be Good." In *Essays on Aristotle's Ethics*, edited by A. Rorty. Berkeley: University of California Press.
- Burnyeat, M. F., and Jonathan Barnes. 1980. "Socrates and the Jury: Paradoxes in Plato's Distinction between Knowledge and True Belief." *Proceedings of the Aristotelian Society, Supplementary Volumes* 54, pp. 173–91, 193–206.
- Buss, Sarah. 2002. "Personal Autonomy." In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta. <http://plato.stanford.edu/archives/fall2008/entries/personal-autonomy/>.
- Christensen, David. 2007. "Epistemology of Disagreement: The Good News." *Philosophical Review* 116:2, pp. 187–217.
- Christman, John. 2002. "Autonomy in Moral and Political Philosophy." In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta. <http://plato.stanford.edu/archives/spr2011/entries/autonomy-moral/>.
- Cicero [106–43 B.C.E.]. 1972. *De Natura Deorum: Cicero: The Nature of the Gods*. Translated by H. C. P. McGregor. London: Penguin.
- Clarke, Desmond. 2006. *Descartes: A Biography*. New York: Cambridge University Press.
- Clifford, W. K. [1877] 1999. "The Ethics of Belief." In *The Ethics of Belief and Other Essays*. Amherst: Prometheus Books.
- Coady, C. A. J. 1992. *Testimony: A Philosophical Study*. Oxford: Clarendon Press.
- Code, Lorraine. 1991. *What Can She Know: Feminist Theory and the Construction of Knowledge*. Ithaca, NY: Cornell University Press.
- Coeckelbergh, Mark. 2004. *The Metaphysics of Autonomy*. New York: Palgrave Macmillan.
- Congar, Yves. 2004. *The Meaning of Tradition*. San Francisco: Ignatius Press.
- Cooper, John. 2003. "Stoic Autonomy." In *Autonomy*, edited by E. F. Paul, F. Miller, and J. Paul. New York: Cambridge University Press.
- Copi, Irving. 1986. *Introduction to Logic*. New York: Macmillan.
- Cosmides, Leda, and John Tooby. 1997. "Evolutionary Psychology: A Primer." Retrieved September 23, 2010 from <http://www.psych.ucsb.edu/research/cep/primer.html>.
- Crane, David. 2005. *Scott of the Antarctic: A Life of Courage and Tragedy*. London: HarperCollins.
- Cunningham, William A., et al. 2004. "Separable Neural Components in the Processing of Black and White Faces." *Psychological Science* 15:12, pp. 806–13.
- Damasio, A. R. 1994. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Quill.
- Dan, Joseph. 2006. *Kabbalah: A Very Short Introduction*. Oxford: Oxford University Press.
- Darwall, Stephen. 2006. "The Value of Autonomy and Autonomy of the Will." *Ethics* 116, pp. 263–84.
- Davidson, Donald. 1970. "How Is Weakness of the Will Possible?" In *Moral Concepts*, edited by J. Feinberg. Oxford: Oxford University Press.
- Dawkins, Richard. 2008. *The God Delusion*. New York: Houghton Mifflin.
- Dennett, Daniel. 2006. *Breaking the Spell*. London: Penguin.
- DePaul, Michael, and Linda Zagzebski, eds. 2007. *Intellectual Virtues: Perspectives from Ethics and Epistemology*. New York: Oxford University Press.
- Descartes, René. 1980. *Discourse on Method*. Translated by D. A. Cress. Indianapolis: Hackett.

- Descartes, René. [1637] 1985. *The Philosophical Writings of Descartes*. Vol. 1, edited and translated by J. Cottingham, R. Stoothoff, and D. Murdoch. New York: Cambridge University Press.
- Dillard, Annie. 1982. *Teaching a Stone to Talk: Expeditions and Encounters*. New York: Harper & Row.
- Dillon, Richard J. 1990. "Acts of the Apostles." In *The New Jerome Biblical Commentary*, edited by R. E. Brown, J. A. Fitzmyer, and R. E. Murphy. Englewood Cliffs, NJ: Prentice-Hall.
- Doherty, R. W. 1998. "Emotional Contagion and Social Judgment." *Motivation & Emotion* 22:3, pp. 187–209.
- Driscoll, D. M. 2009. *The Effects of Prefrontal Cortex Damage in the Regulation of Emotion*. Iowa City: University of Iowa Press.
- Dworkin, Gerald. 1988. "Autonomy, Science, and Morality." In *The Theory and Practice of Autonomy*. New York: Cambridge University Press.
- Edwards, Paul. 1967. "Common Consent Arguments for the Existence of God." In *The Encyclopedia of Philosophy*, vol. 2, edited by P. Edwards. New York: Macmillan and Free Press.
- Elga, Adam. 2007. "Reflection and Disagreement." *Nous* 41:3, pp. 478–502.
- Elshtain, Jean. 2008. *Sovereignty: God, State, and Self*. New York: Basic Books.
- Emerson, Ralph Waldo. [1841] 1903. "Self-Reliance." In *The Complete Works of Ralph Waldo Emerson*, vol. 2, introduction and notes by E. W. Emerson. New York: Houghton Mifflin.
- Fairweather, Abrol, and Linda Zagzebski, eds. 2001. *Virtue Epistemology*. New York: Oxford University Press.
- Feinberg, Joel. 1980. "The Idea of a Free Man." In *Rights, Justice, and the Bounds of Liberty: Essays in Social Philosophy*, edited by J. Feinberg. Princeton, NJ: Princeton University Press.
- Feinberg, Joel. 1989. "Autonomy." In *The Inner Citadel*, edited by J. Christman. New York: Oxford University Press.
- Feldman, Richard. 2004. "Authoritarian Epistemology." In *Evidentialism*, edited by E. Conee and R. Feldman. New York: Oxford University Press.
- Feldman, Richard. 2006. "Epistemological Puzzles about Disagreement." In *Epistemology Futures*, edited by S. Hetherington. New York: Oxford University Press.
- Festinger, Leon. 1954. "A Theory of Social Comparison Processes." *Human Relations* 7:2, pp. 117–40.
- Fichtner, Joseph. 2003. "Tradition (in Theology)." In *New Catholic Encyclopedia*, edited by Thomas Carson and Joann Cerrito. New York: McGraw Hill.
- Field, Hartry. 1994. "Deflationist Views of Meaning and Content." *Mind* 103:411, pp. 249–84.
- Fiennes, Ranulph. 2003. *Captain Scott*. London: Hodder and Stoughton.
- Fine, Gail. 2004. "Knowledge and True Belief in the *Meno*." *Oxford Studies in Ancient Philosophy* 27, pp. 41–81.
- Fitzmyer, J. A. 1998. *The Acts of the Apostles*. New York: Doubleday.
- Flannery, Austin, ed. 1975. *Vatican Council II: The Conciliar and Post Conciliar Documents*. New York: Costello.
- Flathman, R. E. 1972. *Political Obligation*. London: Lowe & Brydone.

- Flood, Anthony. 2003. "Self-Governance in Aquinas and Pre-modern Moral Philosophy." Ph.D. dissertation, University of Oklahoma.
- Foley, Richard. 2001. *Intellectual Trust in Oneself and Others*. New York: Cambridge University Press.
- Foot, Philippa. 1978. "Moral Arguments." In *Virtues and Vices and Other Essays in Moral Philosophy*, edited by P. Foot. Berkeley: University of California Press.
- Frankfurt, Harry G. 2005. *On Bullshit*. Princeton, NJ: Princeton University Press.
- Fricke, Elizabeth. 1998. "Self Knowledge: Special Access versus Artefact of Grammar—a Dichotomy Rejected." In *Knowing Our Own Minds*, edited by C. Wright and C. MacDonald. Oxford: Clarendon Press.
- Fricke, Elizabeth. 2006a. "Second-Hand Knowledge." *Philosophy and Phenomenological Research* 73:3, pp. 592–618.
- Fricke, Elizabeth. 2006b. "Testimony and Epistemic Autonomy." In *The Epistemology of Testimony*, edited by J. Lackey and E. Sosa. New York: Oxford University Press.
- Fricke, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. New York: Oxford University Press.
- Friedman, R. B. 1990. "The Concept of Authority in Political Philosophy." In *Authority*, edited by J. Raz. New York: New York University Press.
- Gallistel, C. R. 1990. *The Organization of Learning*. Cambridge: MIT Press.
- Gass, G. Z., and W. C. Nichols. 1988. "Gaslighting: A Marital Syndrome." *Contemporary Family Therapy* 10:1, pp. 3–16.
- Glaeser, Edward, Giacomo Ponzetto, and Jesse Shapiro. 2005. "Strategic Extremism: Why Republicans and Democrats Divide on Religious Issues." *Quarterly Journal of Economics* 120:4, pp. 1283–1330.
- Godfrey, Joseph, ed. 2006. "Conceiving Trust." In *Trust: Prospects for Science and Religion*. Stony Brook: Metanexus Institute, Templeton Research Lectures, SUNY Stony Brook.
- Godfrey, Joseph. 2012. *Trust of People, Words, and God: A Route for Philosophy of Religion*. Notre Dame, IN: University of Notre Dame Press.
- Goldman, Alvin I. 2001. "Experts: Which Ones Should You Trust?" *Philosophy and Phenomenological Research* 63:1, pp. 85–110.
- Greco, John, ed. 2004. *Ernest Sosa and His Critics*. Malden, MA: Blackwell.
- Greco, John. 2007. "Knowledge as Credit for True Belief." In *Intellectual Virtues: Perspectives from Ethics and Epistemology*, edited by L. Zagzebski and M. DePaul. New York: Oxford University Press.
- Greco, John. 2010. *Achieving Knowledge*. Cambridge: Cambridge University Press.
- Grice, Paul. 1989. *Studies in the Way of Words*. Cambridge: Harvard University Press.
- Hahn, Scott W. 2009. *Covenant and Communion: The Biblical Theology of Pope Benedict XVI*. Grand Rapids: Brazos Press.
- Haidt, Jonathan, Silvia Koller, and M. G. Dias. 1993. "Affect, Culture, and Morality, or Is It Wrong to Eat Your Dog?" *Journal of Personality and Social Psychology* 65:4, pp. 613–28.
- Haidt, Jonathan, Simone Schnall, Gerald L. Clore, and Alexander H. Jordan. 2008. "Disgust as Embodied Moral Judgment." *Personality and Social Psychology Bulletin* 34, pp. 1096–1109.
- Hansen, Charlotte. 2011. "Newman, Conscience and Authority." *New Blackfriars* 92:1038, pp. 209–23.

- Hardwig, John. 1985. "Epistemic Dependence." *Journal of Philosophy* 82:7, pp. 335–49.
- Hardwig, John. 1991. "The Role of Trust in Knowledge." *Journal of Philosophy* 88, pp. 693–708.
- Hare, R. M. 1963. *Freedom and Reason*. New York: Oxford University Press.
- Hart, H. L. A. 1958. "Legal and Moral Obligation." In *Essays in Moral Philosophy*, edited by A. I. Melden. Seattle: University of Washington Press.
- Hart, H. L. A. 1982. *Essays on Bentham*. Oxford: Clarendon Press.
- Hatfield, Elaine, John T. Cacioppo, and Richard L. Rapson. 1993. "Emotional Contagion." *Current Directions in Psychological Sciences* 2, pp. 96–99.
- Hatfield, Elaine, John T. Cacioppo, and Richard L. Rapson. 1994. *Emotional Contagion*. Cambridge: Cambridge University Press.
- Herrnstein, Richard, and Charles Murray. 1996. *The Bell Curve*. New York: Free Press Paperbacks.
- Hick, John. 2004. *An Interpretation of Religion*. 2nd ed. New Haven: Yale University Press.
- Hieronymi, Pamela. 2006. "Controlling Attitudes." *Pacific Philosophical Quarterly* 87:1, pp. 45–74.
- Hills, Alison. 2009. "Moral Testimony and Moral Epistemology." *Ethics* 120, pp. 94–127.
- Hills, Alison, David G. Rand, Martin A. Nowak, and Nicholas A. Christakis. 2010. "Emotions as Infectious Diseases in a Large Social Network: The SISa Model." *Proceedings of the Royal Society B* 277, pp. 3827–35.
- Hinchman, Edward S. 2005. "Telling as Inviting to 'Trust.'" *Philosophy and Phenomenological Research* 70:3, pp. 562–87.
- Hobbes, Thomas. [1651] 1994. *Leviathan*. Edited by E. Curley. Cambridge: Hackett.
- Hobbes, Thomas. 2010. *Behemoth, or, The Long Parliament*, edited by P. Seaward. New York: Oxford University Press.
- Hookway, Christopher. 2001. "Epistemic Akrasia and Epistemic Virtue." In *Virtue Epistemology*, edited by A. Fairweather and L. Zagzebski. New York: Oxford University Press.
- Hopkins, Robert. 2007. "What Is Wrong with Moral Testimony?" *Philosophy and Phenomenological Research* 74:3, pp. 611–34.
- Horwich, Paul, ed. 1994. *Theories of Truth*. New York: Dartmouth.
- Hume, David. [1748] 1995. *An Enquiry Concerning Human Understanding*. Edited by C. W. Hendel. Upper Saddle River, NJ: Prentice Hall.
- Hunter, James D. 2010. *To Change the World*. New York: Oxford University Press.
- Huntford, Roland. 1985. *The Last Place on Earth*. London: Pan Books.
- Hurley, Patrick. 2008. *A Concise Introduction to Logic*. Belmont, CA: Wadsworth.
- Hurtado, Guillermo. 2006. "Notas sobre de utilitate credendi." *Tópicos* 31, pp. 135–46.
- Idziak, J. M. 1980. *Divine Command Morality: Historical and Contemporary Readings*. New York: Edwin Mellen.
- Idziak, J. M. 1989. "In Search of 'Good Positive Reasons' for an Ethics of Divine Commands: A Catalogue of Arguments." *Faith and Philosophy* 6, pp. 47–64.
- John Paul II, Pope. 1993. *Veritas Splendor: Encyclical Letter*. London: Catholic Truth Society.
- Jones, Karen. 1996. "Trust as an Affective Attitude." *Ethics* 107, pp. 4–25.
- Jones, Karen. 1999. "Second-Hand Moral Knowledge." *Journal of Philosophy* 96, pp. 55–78.

- Jones, Karen. 2002. "The Politics of Credibility." In *Mind of One's Own: Feminist Essays on Reason and Objectivity*, edited by L. Antony and C. Witt. Boulder, CO: Westview.
- Jones, Karen. Forthcoming. "The Politics of Intellectual Self-Trust." *Social Epistemology*.
- Jones, Russell. 2010. "Virtue and Happiness in Plato's *Euthydemus*." Ph.D. dissertation, University of Oklahoma.
- Kahneman, Daniel, Amos Tversky, and Paul Slovic, eds. 1982. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.
- Kagan, Shelly. 1989. *The Limits of Morality*. Oxford: Clarendon Press.
- Kant, Immanuel. [1785] 1997. *Groundwork of the Metaphysics of Morals*. Translated and edited by M. Gregor. New York: Cambridge University Press.
- Kant, Immanuel. [1798] 2006. *Anthropology from a Pragmatic Point of View*. Translated and edited by R. Louden. Cambridge: Cambridge University Press.
- Kekes, John. 1991. *Moral Tradition and Individuality*. Princeton, NJ: Princeton University Press.
- Kelly, Thomas. 2005. "The Epistemic Significance of Disagreement." In *Oxford Studies in Epistemology*, vol. 1, edited by J. Hawthorne and T. G. Szabo. Oxford: Oxford University Press.
- Kelly, Thomas. 2006. "Evidence." In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta. <http://plato.stanford.edu/archives/fall2008/entries/evidence/>.
- Kelly, Thomas. 2010. "Peer Disagreement and Higher Order Evidence." In *Disagreement*, edited by R. Feldman and T. Warfield. Oxford: Oxford University Press.
- Kelly, Thomas. 2011. "Consensus Gentium: Reflections on the 'Common Consent.'" In *Evidence and Religious Belief*, edited by K. J. Clark and R. J. Van Arragon. New York: Oxford University Press.
- King, Nathan L. 2011. "Disagreement: What's the Problem, or, A Good Peer Is Hard to Find." *Philosophy and Phenomenological Research*. DOI: 10.1111/j.1933-1592.2010.00441.x.
- Korsgaard, Christine M., ed. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.
- Korsgaard, Christine M. 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.
- Kramer, D. A. 1990. "Conceptualizing Wisdom: The Primacy of Affect-Cognition Relations." In *Wisdom: It's Nature, Origins, and Development*, edited by R. J. Sternberg. Cambridge: Cambridge University Press.
- Kugel, James L. 1997. *The Bible as It Was*. Cambridge: Belknap Press of Harvard University Press.
- Kugel, James L. 2007. *How to Read the Bible*. New York: Free Press.
- Kurtz, Ernest, and Katherine Ketcham. 1992. *The Spirituality of Imperfection*. New York: Bantam.
- Lackey, Jennifer. 2008. *Learning from Words*. New York: Oxford University Press.
- Lackey, Jennifer. Forthcoming. "Disagreement and Belief Dependence." In *The Epistemology of Disagreement: New Essays*, edited by D. Christensen and J. Lackey. Oxford: Oxford University Press.
- Lackey, Jennifer, and Ernest Sosa, eds. 2006. *The Epistemology of Testimony*. Oxford: Clarendon Press.
- Lazarus, R. S. 1991. "Cognition and Motivation." *American Psychologist* 46, pp. 352-67.

- Lehrer, Keith. 1977. "Social Information." *Monist* 60:4, pp. 473–87.
- Lehrer, Keith. 1997. *Self-Trust: A Study of Reason, Knowledge, and Autonomy*. Oxford: Clarendon Press.
- Lewis, C. S. 1974. *The Abolition of Man*. New York: HarperCollins.
- Locke, John. [1691] 1824. *The Works of John Locke*. Vol. 4: *Economic Writings and Two Treatises of Government*. London: Rivington.
- Locke, John. [1709] 1958. "A Discourse of Miracles." In *The Reasonableness of Christianity, and A Discourse of Miracles*. Stanford: Stanford University Press.
- Locke, John. [1689] 1975. *An Essay Concerning Human Understanding*. Edited by P. Niddich. Oxford: Clarendon Press.
- Locke, John. [1689] 2009. *A Letter Concerning Toleration: Humbly Submitted*. New York: Classic Book America.
- Lucas, J. R. 1969. "True." *Philosophy* 44, pp. 175–86.
- Lutheran World Federation and Roman Catholic Church. 1999. *Joint Declaration on the Doctrine of Justification*. Grand Rapids: Wm. B. Eerdmans.
- Lynch, M. P. 2004. *True to Life: Why Truth Matters*. Cambridge: MIT Press.
- Lynn, Richard, John Harvey, and Helmuth Nyborg. 2009. "Average Intelligence Predicts Atheism Rates across 137 Nations." *Intelligence* 37, pp. 11–15.
- Lyons, Jack. 1997. "Testimony, Induction and Folk Psychology." *Australasian Journal of Philosophy* 75:2, pp. 163–78.
- Lyons, William. 1980. *Emotion*. Cambridge: Cambridge University Press.
- MacDonald, Paul. 2010. "The Epistemology of Faith in Augustine and Aquinas." In *Augustine and Philosophy*, edited by P. Cary, J. Doody, and K. Paffenroth. Lanham, MD: Lexington Books.
- MacIntyre, Alasdair. 1986. "Which God Ought We to Obey and Why?" *Faith and Philosophy* 3, pp. 359–71.
- MacIntyre, Alasdair. 1990. *Three Rival Versions of Moral Enquiry: Encyclopaedia, Genealogy, and Tradition*. Notre Dame, IN: University of Notre Dame Press.
- MacIntyre, Alasdair. 1999. *Dependent Rational Animals: Why Human Beings Need the Virtues*. Chicago: Open Court.
- MacIntyre, Alasdair. 2007. *After Virtue*. 3rd ed. Notre Dame, IN: University of Notre Dame Press.
- MacIntyre, Alasdair. 2009. "Intractable Moral Disagreements." In *Intractable Disputes about Natural Law*, edited by L. Cunningham. Notre Dame, IN: University of Notre Dame Press.
- Mackie, J. L. 1969–1970. "The Possibility of Innate Knowledge." *Proceedings of the Aristotelian Society* 70, pp. 245–57.
- Mavrodes, George. 1988. *Revelation in Religious Belief*. Philadelphia: Temple University Press.
- May, Thomas. 1998. "Authority and Obligation." In *Autonomy, Authority, and Moral Responsibility*. London: Kluwer Academic Publishers.
- McMyler, Benjamin. 2007. "Knowing at Second Hand." *Inquiry* 50:5, pp. 511–40.
- McMyler, Benjamin. 2011a. *Testimony, Trust, and Authority*. New York: Oxford University Press.
- McMyler, Benjamin. 2011b. "Doxastic Coercion." *Philosophical Quarterly* 61:244, pp. 537–57.

- Mill, J. S. [1859] 1947. *On Liberty*. Edited by A. Castell. New York: Appleton-Century-Crofts.
- Mill, J. S. [1874] 1963. "Theism." In *The Collected Works of John Stuart Mill*, vol. 10, *Essays on Ethics, Religion, and Society*, edited by J. M. Robson. Toronto: University of Toronto Press.
- Mill, J. S. [1869] 1986. *The Subjection of Women*. New York: Prometheus Books.
- Mills, Charles. 2009. "Rawls on Race / Race on Rawls." *Southern Journal of Philosophy*, 47, annual supplement of the *Proceedings of the University of Memphis Spindel Conference*, edited by B. E. Lawson, pp. 161–84.
- Mlodinow, Leonard. 2008. *The Drunkard's Walk: How Randomness Rules Our Lives*. New York: Random House.
- Montmarquet, John. 1987. "Epistemic Virtue." *Mind* 96, pp. 482–97.
- Montmarquet, John. 1993. *Epistemic Virtue and Doxastic Responsibility*. Lanham, MD: Rowman and Littlefield.
- Moran, Richard, ed. 2001. "The Authority of Self-Consciousness." In *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton, NJ: Princeton University Press.
- Moran, Richard. 2005. "Getting Told and Being Believed." *Philosophers' Imprint* 5:5, pp. 1–29; reprinted in *The Epistemology of Testimony*, edited by J. Lackey and E. Sosa. New York: Oxford University Press.
- Moravcsik, J. M. 1979. "Understanding and Knowledge in Plato's Philosophy." *Neue Hefte Für Philosophie* 15–16, pp. 53–69.
- Moser, P. K. 2004. "Skepticism Undone?" In *Ernest Sosa and His Critics*, edited by J. Greco. Malden, MA: Blackwell.
- Nagel, Thomas. 2010. *Secular Philosophy and the Religious Temperament*. Oxford: Oxford University Press.
- Nakamura, Jeanne, and Csikszentmihalyi Mihaly. 2005. "The Role of Emotions in the Development of Wisdom." In *A Handbook of Wisdom: Psychological Perspectives*, edited by R. J. Sternberg. Cambridge: Cambridge University Press.
- Newman, John H. [1891] 1992. *An Essay in Aid of A Grammar of Assent*. Notre Dame, IN: University of Notre Dame Press.
- Newman, John H. 1994. *Apologia pro vita sua*. New York: Penguin Putnam.
- Nowak, Martin, and Karl Sigmund. 2005. "Evolution of Indirect Reciprocity." In *Nature* 473:27, pp. 1291–98.
- Nozick, Robert. 1981. *Philosophical Explanations*. Cambridge: Belknap Press of Harvard University Press.
- Nussbaum, M. C. 2001. *Upheavals of Thought: The Intelligence of Emotions*. Cambridge: Cambridge University Press.
- Nussbaum, Martha. 2006. *Hiding from Humanity: Disgust, Shame, and the Law*. Princeton, NJ: Princeton University Press.
- Olberding, Amy. 2008. "Dreaming of the Duke of Zhou: Exemplarism and the Analects." *Journal of Chinese Philosophy* 35, pp. 625–39.
- Orwall, Lucinda, and Marion Perlmutter. 1990. "The Study of Wise Persons: Integrating a Personality Perspective." In *Wisdom: Its Nature, Origins, and Development*, edited by R. J. Sternberg. New York: Cambridge University Press.
- Owens, David. 2006. "Testimony and Assertion." *Philosophical Studies* 130:1, pp. 105–29.

- Pappas, George. 2000. "Epistemic Deference." *Acta Analytica* 15:24, pp. 113–16.
- Pinker, Steven. 1997. *How the Mind Works*. New York: Norton.
- Pinker, Steven. 2002. *The Blank Slate*. New York: Penguin.
- Pinker, Steven. 2005. "So How Does the Mind Work?" *Mind and Language* 20:1, pp. 1–24.
- Pinker, Steven. 2008. "The Moral Instinct." *New York Times*, January 13.
- Pitkin, Hanna. 1966. "Obligation and Consent—II." *American Political Science Review* 60, pp. 39–40.
- Plantinga, Alvin. 1991. "When Faith and Reason Clash: Evolution and the Bible." *Christian Scholars Review* 21:1, pp. 8–33.
- Plantinga, Alvin. 1993. *Warrant and Proper Function*. New York: Oxford University Press.
- Plantinga, Alvin. 2000a. "Pluralism: A Defense of Religious Exclusivism." In *The Philosophical Challenge of Religious Diversity*, edited by P. Quinn and K. Meeker. Oxford: Oxford University Press.
- Plantinga, Alvin. 2000b. *Warranted Christian Belief*. New York: Oxford University Press.
- Plantinga, Alvin. 2002a. "Introduction: The Evolutionary Argument against Naturalism." In *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument against Naturalism*, edited by J. K. Beilby. Ithaca, NY: Cornell University Press.
- Plantinga, Alvin. 2002b. "Reply to Beilby's Cohorts." In *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument against Naturalism*, edited by J. K. Beilby. Ithaca, NY: Cornell University Press.
- Plantinga, Alvin, and Michael Tooley. 2008. *Knowledge of God*. Oxford: Blackwell.
- Plato [427–347 B.C.]. 1997. *Plato: Complete Works*. Edited by J. Cooper. Indianapolis: Hackett.
- Quine, W. V. O. 1986. *Philosophy of Logic*. Cambridge: Harvard University Press.
- Ramsey, F. P. 1927. "Facts and Propositions." *Proceedings of the Aristotelian Society, Supplementary Volumes* 7, pp. 153–70.
- Ratzinger, Joseph, Cardinal. 1986. *Feast of Faith: Approaches to a Theology of Liturgy*. Translated by G. Harrison. San Francisco: Ignatius Press.
- Ravenshear, A. F. 1899. "Testimony and Authority." *Mind* 8:29, pp. 63–83.
- Raz, Joseph. 1978. "Authority and Consent." *Virginia Law Review* 67, pp. 103–17.
- Raz, Joseph. 1988. *The Morality of Freedom*. Oxford: Clarendon Press.
- Raz, Joseph. 1999. *Engaging Reason: On the Theory of Value and Action*. Oxford: Oxford University Press.
- Raz, Joseph. 2009a. *The Authority of Law: Essays on Law and Morality*. New York: Oxford University Press.
- Raz, Joseph. 2009b. *Between Authority and Interpretation*. New York: Oxford University Press.
- Reid, Thomas. [1785] 1983. "Essays On the Intellectual Powers of Man." In *Thomas Reid's Inquiry and Essays*, edited by R. E. Beanblossom and K. Lehrer. Indianapolis: Hackett.
- Reid, Thomas. [1785] 1993. "Essays on the Intellectual Powers of Man." In *The Works of Thomas Reid*, edited by Sir W. Hamilton. Charlottesville, VA: Lincoln-Rembrandt Publishing.

- Reid, Thomas. [1764] 1997. *An Inquiry into the Human Mind on the Principles of Common Sense*. Edited by D. R. Brookes. Edinburgh: Edinburgh University Press.
- Rhonheimer, Martin. 2000. *Natural Law and Practical Reason*. Translated by G. Marsbury. New York: Fordham University Press.
- Riggs, Wayne. 2002. "Reliability and the Value of Knowledge." In *Philosophy and Phenomenological Research* 64, pp. 79–96.
- Riggs, Wayne. 2009. "Two Problems of Easy Credit." *Synthese* 169:1, pp. 201–16.
- Riley, Patrick. 1999. *Will and Political Legitimacy: A Critical Exposition of Social Contract Theory in Hobbes, Locke, Rousseau, Kant, and Hegel*. New York: Excel Press.
- Ripstein, Arthur. 2004. "Authority and Coercion." *Philosophy and Public Affairs* 32:1, pp. 2–35.
- Roberts, Robert C. 2003. *Emotions: An Essay in Aid of Moral Philosophy*. New York: Cambridge University Press.
- Roberts, Robert C., and W. Jay Wood. 2007. *Intellectual Virtues: An Essay in Regulative Epistemology*. Oxford: Oxford University Press.
- Rosen, Gideon. 2001. "Nominalism, Naturalism, Epistemic Relativism." *Philosophical Perspectives* 15, *Metaphysics*, pp. 69–91.
- Rousseau, Jean-Jacques. [1762] 1968. *The Social Contract*. Translated by M. Cranston. New York: Penguin.
- Rousseau, Jean-Jacques. 2009. *Emile*. Teddington: Echo Library.
- Rowley, Jennifer, and Frances Slack. 2009. "Conceptions of Wisdom." *Journal of Information Science* 35, pp. 110–19.
- Scanlon, Thomas. 1972. "A Theory of Freedom of Expression." *Philosophy and Public Affairs* 1, pp. 204–26.
- Schmitt, F. F. 2006. "Testimonial Justification and Transindividual Reasons." In *The Epistemology of Testimony*, edited by J. Lackey and E. Sosa. New York: Oxford University Press.
- Schneewind, J. B. 1998. *The Invention of Autonomy: A History of Modern Moral Philosophy*. Cambridge: Cambridge University Press.
- Seligman, Adam B. 1997. *The Problem of Trust*. Princeton, NJ: Princeton University Press.
- Seneca [4 B.C.–A.D. 65]. 1925. *Ad Lucilium Epistulae Morales*. Vol. 3. Translated by R. M. Gummere. Cambridge: Harvard University Press.
- Sextus Empiricus [ca. 2nd century C.E.]. 1994. *Outlines of Scepticism*. Translated by J. Annas and J. Barnes. New York: Cambridge University Press.
- Shieber, Joseph. 2010. "Between Autonomy and Authority: Kant on the Epistemic Status of Testimony." *Philosophy and Phenomenological Research* 80:2, pp. 327–48.
- Simon, Yves. 1991. *A General Theory of Authority*. Notre Dame, IN: University of Notre Dame Press.
- Solomon, Robert. 1980. "Emotions and Choice." In *Explaining Emotions*, edited by A. Rorty. Berkeley: University of California Press.
- Solomon, Robert. 1984. *The Passions: The Myth and Nature of Human Emotions*. New York: Doubleday.
- Solomon, Robert. 1993. *The Passions: Emotions and the Meaning of Life*. Indianapolis: Hackett.
- Solomon, Susan. 2001. *The Coldest March: Scott's Fatal Antarctic Expedition*. New Haven: Yale University Press.

- Sosa, Ernest. 1991. *Knowledge in Perspective*. Cambridge: Cambridge University Press.
- Sosa, Ernest. 1997. "Reflective Knowledge in the Best Circles." *Journal of Philosophy* 94:8, pp. 410–30.
- Sosa, Ernest. 2003. "The Place of Truth in Epistemology." In *Intellectual Virtue: Perspectives from Ethics and Epistemology*, edited by M. DePaul and L. Zabzebski. New York: Oxford University Press.
- Sosa, Ernest. 2008. "The Epistemology of Disagreement." Paper presented to the Princeton Workshop on Disagreement.
- Sosa, Ernest. 2011. *Knowing Full Well*. Princeton, NJ: Princeton University Press.
- Sosa, Ernest, and Barry Stroud. 1994. "Philosophical Skepticism." *Proceedings of the Aristotelean Society, Supplementary Volumes* 68, pp. 263–307.
- Squires, Euan. 1994. *The Mystery of the Quantum World*. New York: Taylor and Francis.
- Stern, Robin. 2007. *The Gaslight Effect: How to Spot and Survive the Hidden Manipulation Others Use to Control Your Life*. New York: Morgan Road.
- Sternberg, R. J. 1998. "A Balance Theory of Wisdom." *Review of General Psychology* 2:4, pp. 347–65.
- Sternberg, R. J. 2000. "Wisdom as a Form of Giftedness." *Gifted Child Quarterly* 44:4, pp. 252–59.
- Sternberg, R. J. 2003. *Wisdom, Intelligence, and Creativity Synthesized*. New York: Cambridge University Press.
- Stocker, Michael, and Elizabeth Hegeman. 1996. *Valuing Emotions*. Cambridge: Cambridge University Press.
- Strawson, P. F. 2008. "Freedom and Resentment." In *Freedom and Resentment and Other Essays*. New York: Routledge.
- Strawson, P. F. 1992. *Libertad y resentimiento*. Mexico City: Universidad Nacional Autónoma de México.
- Sullivan, Francis. 1983. *Magisterium: Teaching Authority in the Catholic Church*. Eugene, OR: Wipf & Stock.
- Taylor, Charles. 1976. "Responsibility for Self." In *The Identities of Persons*, edited by A. Rorty. Berkeley: University of California Press.
- Teichmann, Robert. 2004. "Authority." In *Modern Moral Philosophy: Royal Institute of Philosophy Supplement*, edited by A. O'Hear. Cambridge: Cambridge University Press.
- Thrash, T. M., and A. J. Elliot. 2004. "Inspiration: Core Characteristics, Component Processes, Antecedents, and Function." *Journal of Personality and Social Psychology* 87, pp. 957–73.
- Tiberius, Valerie. 2005. "Wisdom and Perspective." *Journal of Philosophy* 102:4, pp. 163–82.
- Tiberius, Valerie. 2008. *The Reflective Life: Living Wisely with Our Limits*. Oxford: Oxford University Press.
- Tiles, Mary, and J. E. Tiles. 1993. *An Introduction to Historical Epistemology: The Authority of Knowledge*. Oxford: Blackwell.
- van Inwagen, Peter. 1996. "It Is Wrong, Everywhere, Always, and for Anyone, to Believe Anything upon Insufficient Evidence." In *Faith, Freedom, and Rationality: Philosophy of Religion Today*, edited by J. Jordan and D. Howard-Snyder. Lanham, MD: Rowman and Littlefield.

- van Inwagen, Peter. 2010. "We're Right, They're Wrong." In *Disagreement*, edited by R. Feldman and T. Warfield. Oxford: Oxford University Press.
- Vatican II. 1964. *Dogmatic Constitution on the Church: Lumen Gentium*.
- Walsh, W. H. 1964. "Moral Authority and Moral Choice." *Aristotelian Society Proceedings* 65, pp. 1–23.
- Walton, Douglas. 1999. *Appeal to Popular Opinion*. University Park: Pennsylvania State University Press.
- Wedgwood, Ralph. 2007. *The Nature of Normativity*. New York: Oxford University Press.
- Weiner, Matthew. 2003. "Accepting Testimony." *Philosophical Quarterly* 53, pp. 256–64.
- Whitcomb, Dennis. 2010. "Curiosity Was Framed." *Philosophy and Phenomenological Research* 81:3, pp. 664–87.
- Whitcomb, Dennis. 2011. "Wisdom." In *Routledge Companion to Epistemology*, edited by S. Bernecker and D. Pritchard. New York: Routledge.
- White, Roger. 2005. "Epistemic Permissiveness." *Philosophical Perspectives* 19:1, pp. 445–59.
- White, Roger. 2007. "Epistemic Subjectivism." *Episteme* 4, pp. 115–29.
- White, Roger. 2008. "You Just Believe That Because..." Paper presented to the Princeton Workshop on Disagreement.
- Wintrobe, Ronald. 2006. "Extremism, Suicide Terror, and Authoritarianism." *Public Choice* 128, pp. 169–95.
- Williams, Bernard. 1985. *Ethics and the Limits of Philosophy*. Cambridge: Harvard University Press.
- Williams, Bernard. 1995. *Making Sense of Humanity and Other Philosophical Papers*. Cambridge: Cambridge University Press.
- Williams, Bernard. 2002. *Truth & Truthfulness: An Essay in Genealogy*. Princeton, NJ: Princeton University Press.
- Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press.
- Wittgenstein, Ludwig. 1972. *Lectures and Conversations on Aesthetics, Psychology, and Religious Belief*. Edited by C. Barrett. Berkeley: University of California Press.
- Wolff, Robert P. 1998. In *Defense of Anarchism*. Berkeley: University of California Press.
- Wolford, George, Michael Miller, and Michael Gazzaniga. 2000. "The Left Hemisphere's Role in Hypothesis Formation." *Journal of Neuroscience* 20:RC64, pp. 1–4.
- Wolterstorff, Nicholas. 2001. *Thomas Reid and the Story of Epistemology*. Cambridge: Cambridge University Press.
- Wooldridge, Adrian. 1995. "Bell Curve Liberals" *New Republic* 212:9, pp. 22–24.
- Zagzebski, Linda. 1996. *Virtues of the Mind*. New York: Cambridge University Press.
- Zagzebski, Linda. 1999. "Phronesis and Christian Belief." In *The Rationality of Theism*, edited by G. Brüntrup and R. K. Tacelli. Boston: Kluwer Academic Publishers.
- Zagzebski, Linda. 2001a. "Religious Diversity and Social Responsibility." *Logos* 4:1, pp. 135–55.
- Zagzebski, Linda. 2001b. "The Uniqueness of Persons." *Journal of Religious Ethics* 29:3, pp. 401–23.

- Zagzebski, Linda. 2003a. "Emotion and Moral Judgment." *Philosophy and Phenomenological Research* 66:1, pp. 104–24.
- Zagzebski, Linda. 2003b. "The Search for the Source of Epistemic Good." *Metaphilosophy* 34:1–2, pp. 12–28.
- Zagzebski, Linda. 2004a. *Divine Motivation Theory*. Cambridge: Cambridge University Press.
- Zagzebski, Linda. 2004b. "The Epistemology of Religion: The Need for Engagement." *Proceedings of the Twenty-Sixth Annual Wittgenstein Symposium: Knowledge and Belief*, edited by W. Löffler and P. Weingartner, Vienna: Holder-Pichler-Temsky.
- Zagzebski, Linda. 2005. "Epistemic Value and the Primacy of What We Care About." In "Immoral Belief," special issue of *Philosophical Papers*, edited by Ward Jones, 33:3, pp. 353–77.
- Zagzebski, Linda. 2006. "Self-Trust and the Diversity of Religions." In *Proceedings of the Center for Philosophic Exchange*, edited by G. Dicker. Brockport: SUNY Brockport Press; reprinted in *Liberal Faith: Essays in Honor of Philip Quinn*, edited by P. Weithman. Notre Dame, IN: University of Notre Dame Press, 2008.
- Zagzebski, Linda. 2007. "Intellectual Motivation and the Good of Truth." In *Intellectual Virtues: Perspectives from Ethics and Epistemology*, edited by M. DePaul and L. Zagzebski. New York: Oxford University Press.
- Zagzebski, Linda. 2008. "Ethical and Epistemic Egoism and the Ideal of Autonomy." *Episteme* 4, pp. 252–63.
- Zagzebski, Linda. 2009. *On Epistemology*. Belmont, CA: Wadsworth / Broadview Press.
- Zagzebski, Linda. 2010a. "Exemplarist Virtue Theory." *Metaphilosophy* 41, pp. 41–57.
- Zagzebski, Linda. 2010b. "The Rule of St. Benedict and Modern Liberal Authority." *European Journal of Philosophy of Religion* 2:1, pp. 65–84.
- Zagzebski, Linda. 2011. "Epistemic Self-Trust and the *Consensus Gentium* Argument." In *Evidence and Religious Belief*, edited by R. van Aragon. Oxford: Clarendon Press.
- Zagzebski, Linda. Forthcoming. "First Person and Third Person Reasons and Religious Epistemology." *European Journal for Philosophy of Religion*.

- Adams, Robert, 103n6
 Adler, Jonathan, 124n6
 Alcott, Linda Martin, 239n2
 Alston, William, 39–42, 45, 77n2, 80n6
 Anscombe, Elizabeth, 123, 124n5, 162n2, 190
 Antony, Louise, 6n6
 Appiah, Kwame Anthony, 81n7
 Aquinas. *See* Thomas Aquinas
 Ardel, Monika, 92n21
 Arendt, Hannah, 4
 Aristotle, 34n6, 71, 184, 188, 252
 Augustine, 194
 Austen, Jane, 166–169
 Austin, J. L., 123, 124n5
 Ayer, A. J., 34n9
- Baehr, Jason, 92n23
 Baier, Annette, 36n13, 37, 124
 Baltes, P.B., 92n21
 Barnes, J., 13n13
 Barrett, Justin, 239n3
 Barry, Patrick, 253n8
 Barsade, Sigal, 95n26
 Benedict of Nursia, 146–147, 252, 253n8
 Benedict XVI (Pope), 193n7
 Benson, Hugh, 13n14
 Berlin, H. A., 79n5
 Bluck, R.S., 13n13
 Bochensky, Joseph, 190n5
 Borges, Rodrigo, 33n5, 36n15
- Boyer, Pascal, 239n3
 Buber, Martin, 194
 Bulbulia, Joseph, 68n11
 Burge, Tyler, 56n3
 Burnyeat, M. F., 13n13
 Buss, Sarah, 19n20
- Cacioppo, John T., 95n26
 Chesterton, G. K., 199
 Chihuly, Dale, 196n12
 Christensen, David, 205n3, 206n6, 212n8
 Christman, John, 19n20
 Cicero, 185
 Clarke, Desmond, 17n17
 Clifford, W. K., 100n2
 Coady, C.A.J., 121n3, 225n16
 Code, Lorraine, 10n9
 Coeckelbergh, Mark, 19n21
 Cook, Monte, 17n17
 Cooper, John, 21n23
 Copi, Irving, 72
 Crane, David, 88n14
 Cunningham, William A., 243
- Damasio, A. R., 79n5
 Dan, Joseph, 191–192
 Darwall, Stephen, 19n19, 130
 David, Marian, 115n10
 Davidson, Donald, 31
 Dawkins, Richard, 187n3

Descartes, René, 7n7, 9, 16–17, 143
 Dillard, Annie, 88n14
 Dillon, Richard J., 238n1
 Doherty, R.W., 95n26
 Driscoll, D. M., 79n5
 Duns Scotus. *See* Scotus, John Duns
 Dworkin, Gerald, 166

Edwards, Paul, 72–73
 Eflin, Juli, 62n8
 Elga, Adam, 116–117, 205n3, 206n6
 Elliot, A. J., 89n17
 Elshtain, Jean, 140n1
 Elugardo, Ray, 11n11
 Emerson, Ralph Waldo, 27, 28, 55, 60n7

Feinberg, Joel, 19, 19n19
 Feldman, Richard, 205n3, 206n5
 Festinger, Leon, 94n24
 Fichtner, Joseph, 198n13
 Fiennes, Ranulph, 88n15
 Fine, Gail, 8n8, 13n13, 13n14
 Fitzmyer, J.A., 88n15, 238n1
 Flood, Anthony, 20n22
 Foley, Richard, 6n5, 39–40, 42, 55n2,
 58n6, 241n4

Foot, Philippa, 78
 Frankfurt, Harry G., 126n9
 Freud, Sigmund, 239, 240, 244
 Fricker, Elizabeth, 9–12, 15–16, 18, 27,
 53, 120n1, 123n4, 136
 Fricker, Miranda, 142n4
 Friedman, R. B., 5n4

Gallistel, C. R., 115n10
 Gass, G. Z., 62n8
 Gazzaniga, Michael, 115n10
 Godfrey, Joseph, 36n14
 Goldman, Alvin, 5n3
 Greco, John, 118n13
 Grice, Paul, 122

Hahn, Scott W., 193n7
 Haidt, Jonathan, 81, 82n8, 172
 Hardwig, John, 152n7, 154n8

Hart, H.L.A., 106
 Harvey, John, 187n3
 Hatfield, Elaine, 95n26
 Hegeman, Elizabeth, 78n4
 Hernstein, Richard, 214–215
 Hesiod, 4
 Hick, John, 213n10
 Hills, Alison, 174n12
 Hinchman, Edward, 124n5
 Hobbes, Thomas, 23, 23n26, 101,
 141–142
 Hookway, Christopher, 31n2
 Hopkins, Robert, 164n4, 167–168
 Horwich, Paul, 34n9
 Hume, David, 44n22, 71, 85, 128, 187
 Hunter, James Davidson, 221n14
 Huntford, Roland, 88n15
 Hurley, Patrick, 159n1, 181–182

Idziak, J. M., 21n24

Jesus Christ, 193, 194–195
 John Paul II (Pope), 179n15
 Jones, Karen, 36n13, 44, 165n6
 Jones, Russell, 13n14

Kagan, Shelly, 58n5
 Kahneman, Daniel, 68n10
 Kant, Immanuel, 18–26, 163n3, 207–208,
 210, 233–235
 Kelly, Thomas, 47n26, 70n13, 206n5
 King, Nathan, 205n4
 Kischka, Udo, 79n5
 Korsgaard, Christine, 22, 28, 166–167,
 169n8, 233
 Kugel, James, 195, 196n11

Lehrer, Keith, 38, 152
 Lewis, C. S., 72n15
 Locke, John, 5, 9, 11–12, 14–16,
 23n26, 100, 128, 183, 189, 193n6,
 202

Luke, 238
 Lynn, Richard, 187n3
 Lyons, Jack, 127n10

- MacDonald, Paul, 194n9
 MacIntyre, Alasdair, 81n7, 178
 May, Thomas, 116n12
 McMyler, Benjamin, 16n15, 122, 123n4,
 128–130
 Mill, John Stuart, 19n18, 71–72, 142n4
 Miller, Michael, 115n10
 Mills, Charles, 224n15
 Mlodinow, Leonard, 115n10
 Montmarquet, James, 90n20
 Montminy, Martin, 43n21
 Moran, Richard, 122–123, 126n9, 128,
 190
 Moses, 193
 Muhammad, 193
 Murray, Charles, 215
- Nagel, Thomas, 38
 Newman, John Henry, 172, 201
 Nichols, W.C., 62n8
 Nietzsche, Friedrich, 22
 Nowak, Martin, 100n2
 Nozick, Robert, 24
 Nussbaum, Martha, 81–82
 Nyborg, Helmuth, 187n3
- Oates, Lawrence, 87–88
 Orwall, Lucinda, 92n22
 Owens, David, 123n4
- Pascal, Blaise, 199
 Paul, 238
 Perlmutter, Marion, 92n22
 Pinker, Steven, 68n11, 81n7, 215n11,
 239n3
 Plantinga, Alvin, 44n22, 46n24, 194n8,
 205n1, 215n12, 227n19, 245n6
 Plato, 12–14, 16, 20, 175–176, 233
- Quine, W. V. O., 34n9
- Ramsey, F. P., 34n9
 Rapson, Richard L., 95n26
 Ratzinger, Joseph Cardinal.
See Benedict XVI (Pope)
- Ravenshear, A. F., 126n9
 Raz, Joseph, 5n4, 37n16, 99, 106–110, 114,
 115n11, 132–133, 136–137, 139, 140, 143
 Reid, Thomas, 34n6, 71, 85–86, 127n10,
 128–129
 Riggs, Wayne, 32n3, 118n13
 Ripstein, Arthur, 139n13
 Roberts, Robert, 77n3
 Rolls, E.T., 79n5
 Rosen, Gideon, 205n2
 Rousseau, 10n9, 23n26, 141–142
 Rowley, Jennifer, 92n21
 Rutledge, Jonathan, 30n1
- Sartre, Jean-Paul, 30
 Scanlon, Thomas, 167–168
 Schneewind, J. B., 19, 21n24
 Scott, Sir Robert Falcon, 87–88
 Scotus, John Duns, 21
 Seligman, Adam B., 198n15
 Sextus Empiricus, 34n8
 Sigmund, Karl, 100n2
 Simon, Yves, 140n1, 149n6
 Slack, Frances, 92n21
 Slovic, Paul, 68n10
 Smith, Jacqui, 92n21
 Socrates, 13–14, 35, 175–176, 179
 Solomon, Susan, 88n15
 Sosa, Ernest, 92n23, 118n13
 Squires, Euan, 34n6
 Stern, Robin, 62n8
 Sternberg, Robert J., 92n21
 Stocker, Michael, 78n4
 Strawson, Peter, 94
 Sullivan, Francis, 198n14
- Taylor, Charles, 23n25
 Teichmann, Robert, 149n6
 Thomas Aquinas, 20n22, 21n23, 180,
 233, 252
 Thrash, T.M., 89n17
 Tiberius, Valerie, 92n21
 Tiles, Jim and Mary, 239n2
 Trachtenberg, Zev, 23n26
 Tversky, Amos, 68n10

Van Inwagen, Peter, 205n2

Vincent, Francis, 157

Wedgewood, Ralph, 205n2

Whitcomb, Dennis, 35n10, 92n21

White, Roger, 206n5

Williams, Bernard, 35n11, 126n9, 164n4

Wittgenstein, Ludwig, 213

Wolff, Robert Paul, 7, 18, 102n4

Wolford, George, 115n10

Wooldridge, Adrian, 215n11

Zagzebski, Linda, 1, 31n2, 34n7, 49n30,

55n2, 77n1, 87n13, 89n18, 90n20,

92n23, 118n13, 147n5, 163n3, 171n9,

221n13, 226n17, 247n7

- Acts of the Apostles, 238
- Ad Lucilium Epistulae Morales* (Seneca), 185
- ad populum* fallacy, 72
- admiration, 87–93, 105, 147, 150
- agreement, 73, 226
- See also* disagreement
- akrasia*, 31
- Amish, Old Order, 176
- anarchism, epistemic, 7, 18
- animals, experiments on, 115, 157
- Anthropology* (Kant), 25
- antinomies, 207–208
- Apology* (Plato), 175
- assertions, 120–121, 129–130
- assurance model, 122
- atheism, 187
- authority
- account of, 106, 140
 - definitions of, 103
 - disappearance of, 4–7
 - epistemic, 1–3, 99–105
 - communal, 151–157
 - in communities, 140–158, 222
 - moral, 159–161
 - from the outside, 247–251
 - and practical, 136–139
 - rejection of, 1–28
 - testimony and, 131–132
 - and trust, 99–119
 - intellectual, 23–24
 - justification of. *See* Justification Theses
 - moral, 159–180
 - and emotions, 170–174
 - skepticism about, 162
 - and understanding, 174–178
 - over large populations, 140, 143
 - political, 4, 99, 138, 142, 144, 148
 - limits of, 140–144
 - practical, 141
 - religious, 2, 181–203
 - conscientious belief and, 198–203
 - of the self, 29–33
 - self-consciousness and, 234
 - will and, 233
- Authority Principles, Expansion of, 151–153
- autonomy, 2–3, 6–7, 18, 166–170, 229–254
- attacks on, 239–240
 - defined, 230
 - from inside, 235–236
 - intellectual, 23–28, 53, 73
 - Kantian view of, 19–23
 - from outside, 234–235
 - practical, 26
- Behemoth* (Hobbes), 23n26
- beliefs
- belief states, 63, 83–84
 - on command, 5
 - conflicting, 29–33, 213–221
 - conscientious, 199–203, 226–227
 - influences on, 24–25
 - moral, 159–162, 167, 170–174

- beliefs (*continued*)
- norms for, 101
 - reasons for, 45–50
 - self-trust and, 208
 - shared, 222, 226
 - true, 12, 13n13, 16, 54, 83, 92–93
 - acquisition of, 110, 117–118
 - trust in, 126–128
 - trustworthiness of, 132, 218
 - universality of, 68–73
 - See also* religious beliefs
- Bell Curve, The* (Hernstein and Murray), 214–215
- Beyond Justification* (Alson), 39
- Bible, 17n17, 195
- Catholic Church. *See* Roman Catholic Church
- chain model, 193–194, 202
- Christianity, 183
- on conscience, 179
 - tradition in, 191–198
- circularity
- emotional, 82–83
 - epistemic, 38–47, 49, 80n6
- coercion, intellectual, 24, 100, 139, 234, 247
- collective subconscious, 224
- commanding belief, 101–102, 139, 249
- common consent argument, 71–72, 97, 185–188
- See also consensus gentium*
- common sense, 71–72, 85
- communication, difficulties of, 11–12
- communities
- authority of, 236
 - beliefs of, 154
 - breakdown of trust within, 237, 239
 - challenges from outside, 223–224
 - disagreement and, 212, 221–228
 - epistemic authority in, 97–98, 140–158
 - as extended self, 153, 155–156, 237
 - moral truth and, 178–180
 - properties of, 222
 - purposes of, 144–145
 - reasons for joining, 145
 - religious, conflict and, 221–223
 - religious beliefs and, 220–221
 - self-reflection in, 223
- compromise, 206–207
- conflict. *See* disagreement
- conscience, 179–180
- conscientiousness, 61–63, 104–105, 155, 184, 204, 209
- in communities, 222
 - emotional, 86
 - epistemic, 48–50, 55–63, 86, 91
- consciousness
- forms of, 29
 - philosophical investigation of, 229–230
 - reflective, 130
- consensus gentium*, 185–188
- contempt, 79, 84, 88n16, 93–95
- content-independence, 106–107, 116–117, 132
- conversion, 221, 246
- cynicism, 239
- daimonion*, 175–176, 179
- deliberation, moral, 178–179
- democratic structures, 156
- Dependency Thesis
- for authority of testimony, 132
 - Raz's thesis, 107–109, 132–133
- desires, 75–76, 232
- conflicting, 29–31
 - natural, 201–202
- determinism, 207
- dignity, 163–164
- disagreement
- communities and, 212, 221–228
 - first-person, deliberative reasons and, 210–211
 - in moral beliefs, 160
 - reasonable, 111–112
 - antinomy of, 204–209
 - religious, 181–183, 221–223
 - resolution of, 211–221
 - trust and, 73, 204–228
 - See also* peer disagreement
- Discourse on Method* (Descartes), 16–17

- disdain, 173–174
 disgust, 81–82, 172–173
 dissonance in self, 29–33, 43–46,
 231–237, 243–245, 251
 avoidance of, 242n5
 communities and, 224–225
 disagreement and, 211, 213
 resolution of, 216–217, 247
 distrust, political purposes of, 237
 Divine Command Theory, 21n24, 22, 103n6
 doubt, 16–18, 37
 egalitarianism, 6–7, 27, 208–210
 conflict with self-reliance, 204–205
 intellectual, 164
 methodological, 205, 230
 moral, 163–165
 moral-epistemic, 164–165
 and primacy of self, 210
 ego, transcendental, 208
 egocentrism, epistemic, 61–63
 egoism
 emotional, 173, 180
 epistemic, 52–63, 135, 189
 communal, 221–228
 extreme, 204
 extreme vs. standard, 52–59, 66–67,
 74, 119, 223–224, 227
 moral, 160
 religious, 181–184
 ethical, 52, 60, 73
 logical (Kant), 25–26
Emma (Austen), 166–169
 emotion, 75–98
 appropriateness of, 87
 cognitive components, 77, 83
 dispositions, 76–81, 160
 circular justification for, 93
 as grounds for belief, 25n28
 imitation of, 95–96
 and moral belief, 170–174
 nature of, 76, 83
 objects of, 76
 of others, trust in, 93–96
 skepticism about, 84–87, 177
 transmission of, 174
 trustworthy and untrustworthy, 84–87
 See also circularity, emotional
 empiricism, 15
 Enlightenment, 2, 198
epistēmē, 13, 175–176
 epistemic authority
 defined, 5
 See also authority
 epistemology, 1–2, 143, 204
 on authority, 5–6
 of religion, 202
 on testimony, 171
 equality, 4–7
 equals, epistemic, 205, 207
 “Essays on the Intellectual Powers of
 Man” (Reid), 127n10
eudaimonia, 252
 evidence, 109–110, 119, 124–125, 132
 for beliefs, 57–58
 empirical, 243
 theoretical, 238, 240–241, 243
 trust in, 214–215
 of trustworthiness, 61–62
 truth and, 46–49
 evolutionary biology, 240
 exemplars, 87–93, 108–109, 132, 150,
 157, 165
 epistemic, 158, 161, 218, 227
 imitation of, 220–221
 Expansion of Authority Principles,
 151–153
 Expansion of Trust Principle, 97, 127
 experiences, 64–65
 common human, 222
 difference in, 246
 experts, 5, 152
 moral, 164–165
 faculties
 loss of trust in, 240
 reliability of, 10–17
 trust in, 36–45, 48–50, 52–63
 faith, 189–190
 religious, 202–203
 fallacies, 241–242
 falsehoods, 143, 158

- fear, 37, 41, 76–77, 80–81, 95, 173, 243
 feelings. *See* emotion
 first-order beliefs, 217–218
 first-order judgments, 250
 first-person reasons, 240–241
 firsthand experiences, 194–195, 202
 foundationalism, 17, 39, 42, 47
 free will, 207
 freedom, political, 140–141
 future self, 156, 158, 221, 231–233, 241, 253
- gaslighting, 62n8
 General Expansion of Authority Principle, 151
 GJA (General Justification of Authority Thesis), 148–150, 250
- God
 authority derived from, 105
 authority grounded in, 20–21
 belief in, 185–188
 communication from, 183
 executive self and, 33n5
 self-reliance of, 8n8
 testimony of, 189–190
- Golden Rule, 72
 Gospels, 183, 194, 197
 gossip, 100
 Greeks, ancient, 20
Groundwork (Kant), 163n3
- happiness, 252
 harmony, 30–33, 46, 60, 148, 247, 251
 Hebrew scriptures, 191–192, 195–197, 222
 heteronomy of will, 24–26, 234–235
 historical accuracy, 238
 history, development of, 253
 holiness, 202
 Holy Spirit, 194, 197
 Hopkins, Robert, 164n4, 167–168
 humility, 157, 246
- I and Thou* (Buber), 194
 imitation, 89–90, 178
 imperfection, epistemology of, 157–158
 independence, and common consent arguments, 185–186
- indignation, 85
 intelligence, atheism and, 187
 intuitions, 27, 39, 65, 218
Invention of Autonomy, The (Schneewind, J. B.), 19
 Islam, 191–192
- JAB (Justification Theses for authority of belief), 110–112, 117–119, 133, 135
 JAMT (Justification Theses for authority of moral testimony), 161–162, 180
 JAT (Justification Theses for authority of testimony), 133, 135–138, 144, 149–151, 247
 JCEA (Justification Theses for communal epistemic authority), 155–156, 161, 178, 180
 JCEA2 (Justification Theses of religious epistemic authority), 200–201
 “Joint Declaration on the Doctrine of Justification” (Lutheran Federation and Catholic Church), 226n18
- Judaism, 191–197
 Judeo-Christian ideas on authority, 20
- judgment
 conscientious, 137–138
 higher-order, 236
 moral, 170–174, 180
 reflective, 31–32, 144
- justification
 of beliefs, 39–42, 45
 of emotion, 173
 of understanding, 177–178
- Justification Theses
 for authority of belief (JAB), 110–112, 117–119, 133, 135
 for authority of moral testimony (JAMT), 161–162, 180
 for authority of testimony (JAT), 133, 135–138, 144, 149–151, 247
 for communal epistemic authority (JCEA), 155–156, 161, 178, 180
 General Justification of Authority Thesis (GJA), 148–150, 250

- of religious authority, 201
 - of religious epistemic authority (JCEA2), 200
 - third-person theses, 137–139, 248–251
- Kabbalah, 191–192
- knowledge, nature of, 12–16
- “Lectures on Religion” (Wittgenstein), 213
- Letter Concerning Toleration, A* (Locke), 5
- liberalism, political, 106, 140
- lifestyles, morality of, 163
- loop of reason, 39
- Lottery Paradox, 30
- luck, 232
- Lutheran Church, 226
- Meditations* (Descartes), 7n7, 16–17, 143
- Metaphysics* (Aristotle), 34n6, 71
- middle ground, 207
- miracles, 238–239
- “Of Miracles” (Hume), 71
- mistakes, 157, 241–242
- monasticism, 146–147
- moral authority, 159–180
- moral codes, 72
- moral-epistemic egalitarianism, 164–165
- moral motivation, 170–174
- morality, self-governance and, 19–20
- Nature of the Gods, On the* (Cicero), 185
- Need to Resolve Conflict Principle, 224–226
- Nicomachean Ethics* (Aristotle), 71
- No Difference Thesis, 109
- noninterference, 235–236
- Normal Justification Thesis (NJ thesis), 137, 144–150
- Raz’s thesis, 107–110
- See also Justification Theses
- normative propositions, 217–218
- norms, higher-order, 230
- obeying authority, 249
- objects, inanimate, 119
- offense, 78
- “On Self-Reliance” (Emerson), 27
- oppression, epistemic, 142–143
- oral traditions, 195–197
- Papal encyclicals, 168
- Paradox of the Preface, 30
- paradoxes, 30
- peer disagreement, 205n2
- philosophy, 1–2, 149–150, 251–252
- pluralism, 213n10
- point of view, 230
- political authority, justification for, 4
- political beliefs, 212
- political communities, disagreement between, 221
- political principles, 222
- Politics* (Aristotle), 71
- preemption, evidence and, 113–117
- Preemption Thesis for Epistemic Authority, 107, 113–114
- Principle of Epistemic Trust in Others, 68, 211, 214, 216–219
- proof, 42
- Protestantism, American, 198
- prudence, 21n23
- Puritanism, English, 198
- Pyrrhonians, 34, 40n20, 45
- Qur’an, 192
- Rational Recognition Principle, 226
- rational will, 19–23, 233
- rationality, 22, 45, 51, 73, 97, 113, 138, 189
- communities and, 224, 226
 - dissonance and, 217, 231–233
 - emotions and, 43, 202
 - models of, 30–32, 45
 - of religious beliefs, 202
- reason
- authority grounded in, 20–23
 - divine, 21–23
 - self-governance and, 233
- reasons
- for belief, 45–50

- reasons (*continued*)
- deliberative, 190, 240–241, 245
 - deliberative vs. theoretical (first person vs. third person), 63–68, 114, 128–131, 210–211
 - first-order reasons, 145–147
 - noncircular, 76
 - preemptive, 102, 136, 139, 155
 - prima facie*, 57–58, 62, 68–70, 91, 94, 103, 127, 164
 - pro tanto*, 58n5
 - second-order, 146–147
- Recognition of Conscientiousness Principle, 226
- Recognition of Trustworthiness Principle, 226
- reductionism, 128–129
- reflection
- conscientious, 126, 158, 169
 - conscientious communal, 212
 - See also* self-reflection
- religious beliefs
- conflict over, 212, 218–221
 - evolutionary hypothesis, 244–245
 - truth of, 244
- responsibility, in testimony, 123–126, 128
- revelation, models of, 191–198
- ridicule, epistemic, 85
- Roman Catholic Church, 2, 102, 176, 201–202, 226
- Descartes and, 17
 - tradition, 192, 197–199
- rudeness, 78
- Rule of Faith, 197
- Rule of St. Benedict*, 146–147, 252
- scriptures, 189, 191–198, 222
- second-order judgments, 250
- second-order normative propositions, 217–218
- second person model, 122
- self
- as empirical object, 208
 - executive, 22, 33, 63, 168–170, 209, 210, 228, 234–236
 - extended, 153, 155–156, 200
 - harmonious, 31–32, 251
 - nature of, 131
 - philosophical inquiry into, 252
 - prereflective, 37–39, 42
 - primacy of, 205, 210
 - reflective, 38–43
 - self-consciousness, 229–230
 - reflective, 26–27, 29
 - self-fulfillment, 251–254
 - self-governance, 19–21, 230, 233–235, 239–240
 - See also* autonomy
 - self-reflection, 2, 157, 210, 230–233
 - capacity of, 208
 - conscientious, 117, 151, 162, 188, 203, 214–215, 235, 248
 - norm of, 229–233, 240
 - critical, 228
 - value of, 14
 - self-reliance, 25–28, 253
 - Cartesian doubt and, 16–18
 - conflict with egalitarianism, 204
 - epistemic, 3
 - as ideal, 7–11, 52, 55
 - knowledge and, 12–16
 - truth and, 117–119
 - value of, 27
 - self-trust, 27, 29–51, 186–187, 200, 208–209, 246
 - attacks on, 240
 - debunking arguments, 237–247
 - disagreement and, 211–221
 - emotional, 75–84, 93
 - epistemic, 3, 29–51, 56
 - levels of, 49
 - reflection and, 42–45
 - Seneca, *Ad Lucilium Epistulae Morales*, 185
 - sincerity, 10n9, 126–128, 132, 134
 - skepticism, 34–35, 39–42, 45, 47, 205
 - about emotion, 84–87, 177
 - about moral truth, 162
 - slaves, representation of, 206–207
 - South Pole expedition, 87–88
 - sovereign, 141
 - speaking, 134
 - See also* testimony
 - subconscious mind, 243–244
 - Subjection of Women, The (Mill)*, 142n4

- submission, 19–20, 22–23, 27
 sympathy, 85
- taboos, 81n7
- Talmudic *avot*, 191–193
- teachers, 146, 149–150, 175
- telling, 121–125, 139
 - epistemic authority and, 131
 - testimony as, 128
- testimony
 - assurance model, 122
 - authority of, 111, 120–140
 - chains of, 193–194
 - in communities, 154
 - conscientious, 120–128
 - definitions of, 121
 - deliberative vs. theoretical reasons, 128–131
 - divine, 189–199
 - evidence, 123, 128–129, 135–136
 - evidence model of, 189
 - limits of, 170–176
 - models of, 124
 - moral, 159
 - trust and, 126–128, 130, 189–190
 - trustworthiness of, 132
- theism, 185–188
- Torah, 191–192, 195–197
- tradition, 191–198
- Trinity, doctrine of, 212–213
- trust
 - attacks on epistemic trust, 239
 - and authority, epistemic, 99–119
 - components of, 36–38
 - in conscientiousness, 112
 - conscious reflective, 96
 - disagreement and, 204–228
 - in emotions of others, 93–96
 - in evidence, 214
 - expansion of, 151
 - Expansion of Trust Principle, 97, 127
 - feeling component of, 44, 60
 - justification of, 253
 - nonepistemic, 127
 - in normative principles, 217–218
 - in others, 3, 10–12, 52–74, 96–98, 184–188, 209–212, 227
 - prima facie*, 122
 - Principle of Epistemic Trust in Others, 68, 211, 214, 216–219
 - and proof, 42
 - in religious communities, 199–203
 - in theoretical evidence, 238
 - See also* self-trust; trustworthiness
- trustworthiness
 - of beliefs, 132, 218
 - of evidence, 61–62
 - Recognition of Trustworthiness Principle, 226
 - untrustworthiness of others, 10–12
- truth
 - communal search for, 157–158
 - conscientious self-reflection and, 48–50
 - desire for, 45–47
 - disagreement and, 213
 - epistemic self-reliance and, 117–119
 - moral, 162
 - natural desire for, 33–43, 183
 - religious, 183–184
- tyranny, 140–143
- understanding
 - illusory, 177
 - and moral authority, 174–178
- uniqueness thesis, 206n5
- universalism, epistemic, 58, 69–73
 - religious, 184–188
- untrustworthiness of others, 10–12
 - See also* trust; trustworthiness
- virtue, 150, 175–176
- voluntarism, 21n24
- will
 - autonomy of, 23–26, 250
 - coercion of, 234
 - divine, 21–23
 - heteronomy of, 24–26
 - rational, 19–23, 233
- wisdom, 91–92, 95, 150, 253
- wishful thinking, 180, 242