

ABSTRACT

Title of Document: MORAL INTUITIONS AND THEIR ROLE IN JUSTIFICATION

Ryan Fanselow, Ph.D, 2011

Directed By: Professor Samuel J. Kerstein, Philosophy

Moral intuitions play a vital role, not only in ordinary moral thought, but also in how philosophers choose between competing normative theories. The standard view about how intuitions ought to be used in moral theory is John Rawls' method of reflective equilibrium, according to which an agent ought to work back and forth between her intuitions, the principles that systematize them, and other background beliefs, revising each until all of her judgments are consistent. My dissertation addresses two problems with the standard view. First, the method makes use of moral intuitions but offers no account of why these judgments have the epistemic credibility to play a role in choosing between normative theories. Second, when we find an inconsistency between an intuition and a moral principle, the method tells us to revise either the principle or the theory. However, this leaves the interesting question unaddressed. Simple norms of consistency tell us that we ought to revise either the principle or the theory; the interesting question is which should we revise.

I argue that both of these problems can be solved simultaneously by conjoining the method of reflective equilibrium with an account of belief revision. Accordingly, I formulate and defend what I call a contributionist account of belief revision, according to which, when faced with a conflict between beliefs, one revises so as to preserve the belief that makes the greatest overall contribution to the coherence one's set of beliefs. This account, I argue, not only solves the second challenge by making the method of reflective equilibrium more determinate. It also explains why those intuitions that survive the reflective equilibrium process have the requisite epistemic credibility. These intuitions have this credibility in virtue of the contribution they make to the coherence of one's overall set of beliefs.

MORAL INTUITIONS AND THEIR ROLE IN JUSTIFICATION

By

Ryan Fanselow

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2011

Advisory Committee:
Professor Samuel J. Kerstein, Chair
Professor Patricia Greenspan
Professor Christopher W. Morris
Professor Dan Moller
Professor Kurt Gray
Professor Karol Soltan

© Copyright by
Ryan Fanselow
2011

Acknowledgements

I would like to thank my parents, Mary and Michael Fanselow for their love and support.

I also owe a debt of gratitude to Lynn Imai for her love and patience during the process of completing this dissertation. In addition, my advisor Samuel Kerstein provided many comments on many drafts of this material. Sam's feedback greatly benefitted both the presentation and substance of what follows. Finally, the members of my committee also provided valuable feedback during my dissertation defense.

Table of Contents

Acknowledgements	ii
Table of Contents	iii
Introduction	1
Chapter 1: Foundationalism and Coherentism in Moral Epistemology	7
<u>1.1 Justification</u>	7
<u>1.2 When a Belief is Justified</u>	9
<u>1.3 Moral Foundationalism</u>	16
<u>1.4 Disagreement</u>	20
<u>1.5 Disagreement and Moral Foundationalism</u>	30
<u>1.6 Non-Moral Normative Foundationalism</u>	34
<u>1.7 Descriptive Foundationalism</u>	37
<u>1.8 Quasi-Foundationalism</u>	39
<u>1.9 Conclusion</u>	41
Chapter 2: Skeptical Doubts about Moral Intuitions	42
<u>2.1 Moral Psychology</u>	42
<u>2.2 Intuitions</u>	44
<u>2.3 Intuitions and Reliability</u>	49
<u>2.4 The Reliability of Non-moral Intuitions</u>	50
<u>2.5 The Reliability of Moral Intuitions</u>	55
<u>2.6 Some Possible Responses on Behalf of Moral Intuitions</u>	61
<u>2.7 The Good News about Intuitions</u>	71
<u>2.8 Conclusion</u>	74
Chapter 3: Reflective Equilibrium in Ethics	75
<u>3.1 Introducing Reflective Equilibrium</u>	75
<u>3.2 Narrow Reflective Equilibrium</u>	75
<u>3.3 Wide Reflective Equilibrium</u>	77
<u>3.4 From Intuitions to Considered Moral Judgments</u>	80
<u>3.5 Against The Independence Constraint</u>	82
<u>3.6 A Coherentist Theory of Justification</u>	85
<u>3.7 The First Consideration - In Ethics, Intuitions are Unavoidable</u>	86
<u>3.8 The Second Consideration - Intuitions are Sometimes Unreliable</u>	88
<u>3.9 The Quinean Consideration</u>	88
<u>3.10 From Methodology to Epistemology</u>	91
<u>3.11 Conclusion</u>	92
Chapter 4: Reflective Equilibrium, Coherence, and Moral Intuitions	94
<u>4.1 The Worry about Intuitions</u>	94
<u>4.2 The First Proposed Solution – The Analogy to Non-moral Observation Reports</u> ..	94
<u>4.3 The Second Proposed Solution – The Corrective Reply</u>	97
<u>4.4 Credence</u>	100
<u>4.5 Coherence</u>	101
<u>4.6 Coherence and Reflective Equilibrium</u>	106
<u>4.7 Time Slice Justification vs. Historical Justification</u>	110

4.8 Moral Coherence	111
4.9 The Dogmatist Revisited	119
4.10 The Epistemology of Moral Intuitions	121
4.11 The Moral Center	124
4.12 Ideal and Non-Ideal Moral Reasoning	128
Chapter 5: In Defense of Reflective Equilibrium: A Response to Some Objections	132
5.1 Reflective Equilibrium and Its Critics	132
5.2 Reflective Equilibrium and Moral Conservatism	132
5.3 Reflective Equilibrium and Moral Relativism	133
5.4 Reflective Equilibrium and Moral Truth	136
5.5 Reflective Equilibrium and Moral Skepticism	138
5.6 Reflective Equilibrium and Deductive Moral Arguments	142
5.7 The Charge of Vacuity	143
Chapter 6: Aid to the Worst Off: A Case Study	146
6.1 Reflective Equilibrium, Moral Conservatism, and Deeply Held Moral Intuitions, Again	146
6.2 Starting Points	147
6.3 Singer's Argument as an Application of the Method of Reflective Equilibrium ..	150
6.4 Kant on Imperfect Duties	153
6.5 Kamm on Distance	159
6.6 Slote on Empathy	162
6.7 Giving up the Pond	168
6.8 Concession and Mitigation	172
6.9 The Limits of Beneficence	173
6.10 Fair Share Views	176
6.11 Conclusion	179
Conclusion	180
Bibliography	184

Introduction

Act-utilitarianism is the view that one ought to do that action which maximizes overall happiness. There is a familiar argument against this view that begins by asking us to think about a case.

You [are a doctor with]... five patients in the hospital who are dying, each in need of a separate organ. One needs a kidney, another a lung, a third a heart, and so forth. You can save all five if you take a single healthy person and remove his heart, lungs, kidneys and so forth, to distribute to these five patients. Just such a healthy person is in room 306. He is in the hospital for routine tests. Having seen his test results, you know that he is perfectly healthy and of the right tissue compatibility. If you do nothing, he will survive without incident; the other patients will die, however. The other patients can be saved only if the person in Room 306 is cut up and his organs distributed. In that case, there would be one dead but five saved.¹

Almost everyone who hears of this case has the intuition that it would be wrong to kill the one to save the five. However, assuming all else is equal and there is nothing especially unique about the one (e.g. he is not on the verge of curing cancer), act-utilitarianism yields the result that you ought to kill the one and save the five. Thus, act-utilitarianism is false.

This argument is an example of a type of argument that is very common in philosophy. These arguments have the following structure.

¹ Gilbert Harman, *The Nature of Morality* (Oxford: Oxford University Press, 1977), pp. 3-4.

- 1) In case C, I/we/most people have the intuition that P
- 2) Theory T yields the result $\sim P$ in C
- 3) Therefore, T must be rejected

This argument, as I have formulated here, is invalid. Nothing logically follows from 1) and 2) about whether or not one should accept T. It seems, thus, that this argument needs a further premise. However, it is no easy task to determine what the missing premise is.

The following premise would do the trick: for all moral theories and all intuitions, if a theory conflicts with an intuition, then that theory must be rejected. No philosopher that I know of accepts this universally quantified premise. Indeed, as I will argue in chapter 2, our intuitions are not entirely consistent with each other. Therefore, accepting this premise would commit one to a kind of incoherence.

Obviously, we should weaken the universal quantifiers. However, once we do this, our argument no longer yields the result that T must be rejected. If all we know is that on some occasions when a theory conflicts with an intuition that theory must be rejected, we still need to know that the case of act-utilitarianism's conflict with our intuitions in the transplant case is one of those occasions in order to be justified in accepting the argument's conclusion. The question that we need an answer to is the following: when is conflict with an intuition grounds for rejecting a theory?

It was perhaps with an eye toward answering this question that John Rawls developed the method of reflective equilibrium.² We will have occasion to spell out the method in detail in chapter 3 but for now the rough idea is the following. We begin by looking for the theory that best systematizes our intuitions. Presumably, the theory that

² John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971). See especially §§4, 9, & 87.

best systematizes our intuitions will still conflict with some of our intuitions. In this case, we work back and forth, sometimes modifying our theory further to fit our intuitions and sometimes modifying our intuitions to fit the theory.

This use of this method is widespread in philosophy but it is far from uncontroversial. Here, I want to consider two objections. The first objection notes that intuitions play a crucial role in justifying moral theories in the method of reflective equilibrium. It then asks what gives intuitions the epistemic credibility necessary to play this role. Often this objection is motivated by the fact that there is considerable ground for skepticism about the reliability of moral intuitions. If there are legitimate grounds for doubt about the reliability of a judgment, how can we use it to adjudicate between competing theories?³

Most philosophers do not share the kind of skepticism about intuitions that underlies this objection. Nonetheless, as I will argue in chapter 2, this skepticism cannot be dismissed out of hand. Merely attempting to shift the burden of proof or offering a promissory note that one day we may be able to explain the epistemic credibility of intuitions is unsatisfying.

³ Statements of this objection can be found in R.M. Hare, "Rawls Theory of Justice" *Philosophical Quarterly* 23 (1973): 144-55. Reprinted in Daniels (ed.), *Reading Rawls*, pp. 81-107; Peter Singer, "Sidgwick and Reflective Equilibrium," *Monist* 58 (1975): 490-517. Reprinted in *Unsanctifying Human Life* (Oxford: Blackwell, 2002), pp. 27-50; R.B. Brandt, *A Theory of the Good and the Right* (Oxford: Oxford University Press, 1979), Ch. 1; David Copp, "Considered Moral Judgments and Moral Justification: Conservatism in Moral Theory," in David Copp and Michael Zimmerman (eds.), *Morality, Reason, and Truth* (Totowa, NJ: Rowman and Allanheld, 1984); Stefan Sencerz, "Moral Intuitions and Justification in Ethics," *Philosophical Studies* 50 (1986): 77-95; Robert Cummins, "Reflections on Reflective Equilibrium," in Michael DePaul and William Ramsey (eds.), *Rethinking Intuitions* (Oxford: Rowman & Littlefield, 1998), pp. 113-28; Allen Wood, *Kantian Ethics* (Cambridge, UK: Cambridge University Press, 2008), pp. 49-51. Whenever a reprint of a work is listed in the references, any page numbers will given refer to the reprint.

There is a second kind of objection that does not rely on any kind of general skepticism about intuitions. Instead, one pressing this objection may grant that intuitions do have the kind of credibility to perform the task that the method of reflective equilibrium assigns them. Still, the objector insists, the method does not have the resources to solve the problem that we began with. This objection is pressed by Kwame Anthony Appiah when he writes,

Indeed, one could be forgiven for thinking that reflective equilibrium is really another name for the problem rather than the solution to it... As you might fear the procedure for reaching reflective equilibrium is less than determinate.⁴

Appiah goes on to point out that given a conflict between theory and an intuition, the method of reflective equilibrium tells us to revise either the theory or the intuition. However, this does not answer the question that needs answering: should we reject the theory or the intuition?

Both of these problems are troubling. Nonetheless, I will argue in this dissertation that they can be solved in, more or less, a single swoop. The trick is to motivate and defend a theory of belief revision within the method of reflective equilibrium. Accordingly, I formulate and defend what I call a contributionist account of belief revision, according to which when faced with a conflict between beliefs one revises so as to preserve the belief that makes the greatest overall contribution to the coherence one's set of beliefs. This theory, I will argue, not only solves Appiah's challenge by making the method of reflective equilibrium more determinate. It can also explain why those intuitions that survive the reflective equilibrium process have the requisite

⁴ *Experiments in Ethics* (Cambridge, MA: Harvard University Press, 2008), p. 78.

epistemic credibility. These intuitions have this credibility in virtue of the contribution they make to the coherence of one's overall set of beliefs.

This theory is motivated by the kind of concerns that motivate coherentist theories of justification generally. I begin, therefore, by discussing these considerations in chapter 1. There, I argue that coherentist theories in moral epistemology have a decisive advantage over rival foundationalist theories.

In chapter 2, I turn to moral intuitions directly. There, I examine the evidence for skeptical claims about the reliability of moral intuitions that underlie the first objection to the method of reflective equilibrium discussed above.

In chapter 3, I discuss in detail the method of reflective equilibrium and the considerations that motivate it. In chapter 4, I take up the two objections mentioned above and attempt to solve them by formulating and defending the contributionist account of belief revision.

The two objections mentioned above are, in my view, the two most challenging objections to the method. They are, however, far from the only objections. In chapter 5, I take up a range of other objections and show that the method has the resources to meet these objections.

In chapters 4 and 5, I argue against a common objection to the method, that it is conservative in one or both of two senses. In the first sense, the method is charged to be conservative because it cannot require, or perhaps even allow for, the revision of deeply held intuitive judgments. In the second sense, the method is alleged to be conservative because it would merely reinforce the moral status quo. I argue that facts about the

structure of the method of reflective equilibrium show that the method is not conservative in either of these senses.

Nonetheless, I recognize that abstract arguments from the structure of the method will not be satisfying to all. Many will retain their doubts until they see that, in practice, the method can require revisions to deeply held moral intuitions and status quo morality. Therefore, in chapter 6, I turn to an actual moral problem in order to show that the method can, in fact, be useful in practice and that it need not merely reinforce deeply held intuitions or status quo morality.

Chapter 1: Foundationalism and Coherentism in Moral Epistemology

1.1 Justification

Things would be better if we, like God, had perfect access to truth.

Unfortunately, we are not in such a position. Descartes famously argued that we are trapped behind a veil of experience; we only have access to our perceptions of the world and not the world itself. Even if Descartes is correct, I suspect that this is the least of our problems. For even if we suppose that experience is a reliable guide to truth, we are limited in what we can observe. Furthermore, we are often forced to make judgments when our evidence is less than complete and even when we have good evidence we are prone to a range of cognitive biases that make us less than ideal judges.

Cognitively imperfect beings like us need an account of when it is epistemically appropriate to believe a proposition. I suspect that this need is a deep, permanent feature of our epistemic condition. The concept of epistemic justification fills this need. To say that a proposition is justified is roughly to say that an agent can believe it rationally or without being epistemically irresponsible.

These brief remarks are not intended to provide a complete account of the concept referred to by the word 'justification.' Nonetheless, a few brief remarks about why I understand 'justification' this way are in order. As a methodological matter, I think that the best way to understand a concept is to think about why we need it. I have suggested that we need such a concept as justification because creatures like us are imperfect epistemic agents in need of some account of when it is appropriate to believe a proposition.

As minimal as this understanding of justification is, there is at least one group of philosophers that may object to it. Currently in epistemology there is a dispute between internalists, who think that what justifies a belief must be in some sense accessible, or internal, to the believer, and externalists who hold that what justifies a belief may be something that is inaccessible, or external, to the agent.¹ If externalism is true, an account of justification may hold that whether or not a belief is justified is inaccessible to believers. An externalist account of justification may, for this reason, not provide an account of justification that is useful to actual epistemic agents deciding what is appropriate or rational to believe. Thus, the externalist may object that the account of justification given here assumes internalism.

This is not the place to settle the debate between internalism and externalism. To some degree, I doubt that such a debate can be settled. A term like ‘justification’ is broad enough to admit of more than one analysis, and thus, it is at least possible that the internalist and the externalist are both providing plausible accounts of ‘justification.’

What I want to point out here is that even if externalists have the correct account of ‘justification,’ the need that we imperfect epistemic agents have for an account of when one can rationally believe a proposition does not disappear. Thus, even if the only plausible analysis of ‘justification’ is an externalist analysis, we would still need a (different) concept to fulfill this need. If externalists think that there is a better term for this concept than ‘justification’ they can feel free to substitute their own word for my ‘justification.’ I think that both internalist and externalist accounts may be plausibly called accounts of ‘justification.’

¹ A useful collection of papers on this debate is Hillary Kornblith (ed.), *Epistemology: Internalism and Externalism* (Oxford: Blackwell, 2001).

One more point about my understanding of the concept of justification before we can move on to the issue of when a belief is justified. Whether or not it is rational to adopt a belief, or whether or not a belief is justified, surely depends on what evidence we have available to us. This means that we should be prepared to accept that if I have different evidence than you, a proposition could be justified for me but not you. To take a concrete example, I have a belief that I am wearing a green t-shirt today. I am justified in accepting this belief because I see that I am wearing a green t-shirt. Now imagine that it spontaneously occurs to my mother, who has not seen me in months, that I am wearing a green t-shirt today. She would not be justified in believing that I am wearing a green t-shirt today because she has no evidence to suggest that I am. In this case, the same belief is justified for me but not for my mother. We can, and should, accept this consequence while denying any kind of general relativism about truth.

1.2 When a Belief is Justified

Having addressed, albeit briefly, what justification is, we can move on to the issue of when a belief is justified. Ordinarily, when I am asked why I am justified in believing a particular proposition B, I will cite some further proposition (or set of propositions) C.² Let us say that propositions like B which are justified by another proposition are *inferentially justified*. If I claim that B is justified by C, one could reasonably ask why I am justified in believing C, since if C is unjustified, it is doubtful that C could justify B. Usually I will answer that I am justified in believing C because of some further proposition (or set of propositions) D. Most philosophers believe that this process cannot

² It is worth making a brief remark about the relationship between beliefs and propositions; a belief is a mental state whose content is a proposition.

continue *ad infinitum*.³ Furthermore, if the chain of inferential justification from B to C to D were to circle back around such that D was ultimately justified by B, it seems that we would be caught in a vicious circle, since B would ultimately be justified by itself. Since this chain of justification cannot be circular or infinite, it follows that it must end somewhere. If it were to end in an unjustified proposition, then it seems doubtful that B is justified since, as we noted earlier, an unjustified proposition cannot justify a further proposition. Therefore, if B is justified at all, then B must be part of a chain of inferential justification that ends with a proposition that is justified independently of any other proposition.

This argument, which traces back to Aristotle, has led many philosophers to adopt a position called *foundationalism*, the view that a belief is justified if and only if it is a *basic belief*, a belief that is justified non-inferentially, or it is part of a chain of inferential justification that ends with a basic belief.⁴ Foundationalism has a fair amount of intuitive appeal, but it faces a serious problem. Many philosophers worry that it collapses into *skepticism*, the view that there are no justified beliefs. To avoid skepticism, the foundationalist must hold that there is at least one basic belief. But how could a belief be basic? How could a belief be justified without being supported by another proposition? One might worry that no belief is justified without the support of any further propositions. If this is true, then there are no basic beliefs. Since, for the foundationalist all justified beliefs are either basic or inferentially justified by a basic belief, if there are

³ For the exception that proves the rule see Peter Klein, "Foundationalism and the Infinite Regress of Reasons," *Philosophy and Phenomenological Research* 58 (1998): 919-26

⁴ For Aristotle's original statement of this argument, see *Posterior Analytics*, book I, chapters 2-3.

no basic beliefs, then there are no justified beliefs. Hence, the foundationalist is committed to skepticism.

One way that traditional foundationalists have responded to this argument is by claiming that some beliefs guarantee their own truth.⁵ The foundationalist can then avoid skepticism by identifying basic beliefs as just those beliefs that guarantee their own truth.

But let us consider this claim. Take an ordinary true empirical belief such as ‘the Eiffel Tower is made of iron.’ This belief is true but does not guarantee its own truth. It is only true in virtue of further facts about the world. What would it be for a belief to guarantee its own truth? Such a belief would have to be true just in virtue of the fact that the believer had it. Perhaps, there is such a belief, something like the belief that I have a belief. If there were such a belief, then the foundationalist would have avoided skepticism, in that she would be able to hold that there is at least one belief that is justified. But the foundationalist is still committed to a kind of mitigated skepticism because she only has shown that one belief is justified. To hold that at least some of our ordinary beliefs about middle-sized objects are justified, she has to be able to infer them from these basic beliefs. To avoid mitigated skepticism the foundationalist must hold that there are some beliefs that guarantee their own truth and are robust enough to inferentially justify our ordinary beliefs about middle-sized objects.

In response to this worry, the foundationalist could retreat from the claim that basic beliefs guarantee their own truth and make a weaker claim. Perhaps basic beliefs are those that are very likely to be true. However, the fact that a belief is likely to be true

⁵ This appears to be Descartes’ view in the *Meditations on First Philosophy*. For a contemporary defense see Timothy McGrew, “A Defense of Classical Foundationalism” in Louis Pojman (ed.), *Theory of Knowledge: Classical and Contemporary Readings* (Belmont, CA: Wadsworth, 1998).

seems to be a further fact about the belief. If a belief B is justified because it is likely to be true, then it seems that B is justified by appeal to a further proposition, namely the proposition that B is likely to be true. But if B is justified because of a further proposition, then B is not basic at all. The foundationalist seems to have avoided skepticism only by denying foundationalism.

Worries like these about foundationalism have led some philosophers to adopt *coherence theories of justification* according to which a belief is justified if and only if it coheres with certain other beliefs.⁶ To illustrate clearly what coherentism is, it is useful to think about how the coherentist might respond to the Aristotelean argument that we began this section with. This argument notes that when asked why we are justified in believing a proposition, we usually cite a further proposition. When asked how that further proposition is justified we usually cite yet another proposition. There seem to be three distinct possibilities: this process of citing a further proposition continues *ad infinitum*, the process is circular (e.g. B is justified by C which is justified by D which is justified by B), or the process ends somewhere. The first option is implausible. It seems unlikely that there could a chain of beliefs that stretches onto infinity without circling back on itself. The second option seems to allow for vicious circularity. Surely we cannot invoke B to justify B. Thus, by process of elimination we are left with the third foundationalist option.

The coherentist ultimately argues that we should, in some sense, adopt the second option. But how can the coherentist avoid the problem of vicious circularity? In the face

⁶ This view is defended by W.F. Sellars in *Science, Perception, and Reality* (London: Routledge & Kegan Paul, 1963); Bruce Aune in *Knowledge, Mind, and Nature* (New York: Random House, 1967); and Gilbert Harman in *Thought* (Princeton, NJ: Princeton University Press, 1973).

of this worry, the coherentist makes a radical move. The Aristotelian argument relies on a conception of justification that is essentially *linear*. A comparison may be useful here. According to this conception, the structure of justification is, in some respects, like that of a building. A building rests on a foundation, which supports the first floor, which in turn supports the second floor, and so on. Each floor *requires* the previous floor for its structural support.⁷ Therefore, if one were to remove the foundation, then the whole building would collapse. Suppose S is a foundational belief and that it justifies T, which in turn justifies U, and so on. On the linear conception of justification, a justified S is required for T's justification. If S turns out to be unjustified, the whole chain of beliefs is unjustified.

The coherentist's radical move in face of the problem of circularity is to deny the linear picture of justification on which the Aristotelian argument rests. Instead of looking at a building as the appropriate model for the structure of justification, we should look to a spider's web. A particular thread derives support from nearby threads while the nearby threads derive support from each other and that particular thread. According to the coherentist non-linear picture of justification, a belief S could be justified because of its inferential relationship to T, U, and other propositions, while T is justified because of its inferential relationship to S, U, and other propositions.

⁷ The building metaphor is a bit oversimplified. While the first floor of a building may depend on the foundation for *all* of its structural support, a defender of the linear picture of justification need not claim that *all* of a belief's justification is derived from a foundational belief. Instead, the linear view only requires the claim that epistemic support from a foundational belief is necessary for an inferentially justified belief to be justified. For more discussion see Robert Audi, *Epistemology: A Contemporary Introduction* 2nd ed. (London: Routledge, 1998), pp. 210-211.

It should be clear that the worry about circularity arises only if we assume a linear picture of justification. If B requires C for its justification and C requires D for its justification while D requires B for its justification, then we can only show that B is justified by assuming it in the first place. But if justification is not linear in this way, then we are not left with this result. B can be justified by its relationship to C, D and the other propositions the agent believes, while C is justified by its relationship to B, D and the other propositions the agent believes.

Rejecting the linear picture of justification would answer the worry about circularity. But should we reject this picture in the first place? After all, it seems rather intuitive. As we noted at the beginning of this section, when someone asks me why I am justified in having a belief, I cite a further proposition. If someone asks me about this further proposition, I will cite yet another proposition, and so on.

The coherentist can agree that justification appears linear. But it appears linear only because we are considering the justification of a small number of beliefs. If I am asked why I am justified in believing a proposition, the right thing to do, whether we have a linear or non-linear picture of justification, is to cite nearby beliefs. According to the coherentist, the mistake of a linear picture of justification is to assume nothing changes when we are asking about our entire set of beliefs or a large subset thereof. In the latter case, we can cite the coherence of the set of beliefs. We do not need to claim that some of these beliefs constitute a foundation upon which all other knowledge depends.

These are obviously deep theoretical issues that require a lot more discussion. My intent here is not to resolve these issues but simply to make clear what the basic

commitments and difficulties are for foundationalism and coherentism. Note that even if the coherentist can successfully defend the rejection of the linear conceptions of justification, the coherentist about perceptual beliefs is not entirely out of trouble. There are many other objections that a coherentist about perceptual beliefs must contend with. Let me mention one more here. Suppose that the coherentist is right and justification is purely a matter of internal coherence between beliefs. Suppose further that it is coherent with my overall set of beliefs to believe that you will not be in your office today. Perhaps I believe that today is your day off and that you never come into your office on your day off. Now let us imagine that I come to your office and I see you there. In this case, it may be more coherent for me to continue to hold that you did not come to your office today despite the fact that I see you in your office. In some cases, coherentism about perceptual beliefs may have the consequence that I should just ignore the evidence from my experience of the world.

In light of these difficulties, many philosophers have elected to go moderate. Thus, contemporary epistemologists like Robert Audi and James Pryor have abandoned some central Cartesian claims and opted for a moderate foundationalism.⁸ Laurence Bonjour and Keith Lehrer have allowed some elements of foundationalism to play a role in their coherence theories of justification,⁹ and finally Susan Haack has adopted a view

⁸ Robert Audi, *The Structure of Justification* (Cambridge, UK: Cambridge University Press, 1993) and James Pryor, "The Skeptic and the Dogmatist," *Nous* 34 (2000): 517-49.

⁹ Laurence Bonjour, *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1985) and Keith Lehrer, *Theory of Knowledge* (Boulder, CO: Westview Press, 2000). Bonjour has since abandoned the moderate coherentism he defends in this book.

that she calls foundherentism which seems to occupy a conceptual space between the moderate versions of foundationalism and coherentism.¹⁰

For our purposes we can leave open the question of which of these views, if any, is correct. This is because our project is moral epistemology and I want to leave open the question of whether or not the correct account of moral justification is the same as the correct account of non-moral justification.

1.3 Moral Foundationalism

In the rest of this chapter I will argue that whatever its plausibility in the non-moral case, foundationalism fails as an account of moral justification. I do this in order to motivate the search for a coherentist theory of moral justification that will occupy the rest of this dissertation. Let me begin by making clear what foundationalism about moral justification is committed to and how such a view might be developed. *Foundationalism about moral justification* holds that a moral belief is justified if and only if it is a basic belief or it is part of a chain of justification that ends in a basic belief. There are several different possible versions of foundationalism about moral justification, each of which takes a different set of beliefs to be basic. *Moral foundationalism* holds that the basic beliefs that justify all other justified moral beliefs are themselves moral beliefs. *Non-moral normative foundationalism* holds that the basic beliefs that justify all justified moral beliefs are non-moral normative beliefs. *Descriptive foundationalism* holds that the basic beliefs in question are descriptive beliefs. In this chapter, I will argue that all three versions of foundationalism about moral justification are untenable.

¹⁰ Susan Haack, *Evidence and Enquiry: Towards Reconstruction in Epistemology* (Oxford: Blackwell, 1993).

I begin with moral foundationalism. I will devote more space to moral foundationalism than any of its rivals because moral foundationalism is the version of foundationalism about moral justification that has received the most attention. In fact, moral foundationalism was perhaps the dominant view among moral philosophers in the early twentieth century.¹¹ At some point in the early to mid-twentieth century it became the subject of serious criticisms that were thought to have shown the view to be implausible. However, moral foundationalism is currently going through something of a revival in the contemporary literature.¹² Contemporary defenders argue that many of the old criticisms were based on misunderstandings about the commitments of the view.

At first, moral foundationalism can appear puzzling. How could a moral belief be epistemically basic? There are two sort of answers moral foundationalists give to this question. The standard version of moral foundationalism claims that some moral propositions are self-evident. *Self-evident* propositions are those true propositions such that “if one adequately understands them, then by virtue of that understanding one is

¹¹ For early twentieth century defenses of moral foundationalism see Henry Sidgwick, *The Methods of Ethics* 7th Ed. (London: MacMillan, 1907); G.E. Moore, *Principia Ethica* (Cambridge, UK: Cambridge University Press, 1903); H.A. Prichard, “Does Moral Philosophy Rest on a Mistake?” *Mind* 21 (1912): 21-37; and W.D. Ross, *The Right and the Good* (Oxford: Oxford University Press, 1930).

¹² For contemporary defenses of moral foundationalism see Robert Audi, “Intuitionism, Pluralism, and the Foundations of Ethics,” in Walter Sinnott-Armstrong and Mark Timmons (eds.), *Moral Knowledge? New Readings in Moral Epistemology* (Oxford: Oxford University Press, 1996); Robert Audi, “Moderate Intuitionism and the Epistemology of Moral Judgment,” *Ethical Theory and Moral Practice* 1 (1998): 15-44; Jeff McMahan, “Moral Intuition,” in Hugh LaFollette (ed.), *Blackwell Guide to Ethical Theory* (Oxford: Blackwell, 2000), pp. 92-101; Phillip Stratton-Lake (ed.), *Ethical Intuitionism: Re-evaluations* (Oxford: Oxford University Press, 2002); Russ Shafer-Landau, *Moral Realism: A Defense* (Oxford: Oxford University Press, 2003); Robert Audi, *The Right in the Good: A Theory of Intuition and Intrinsic Value* (Princeton, NJ: Princeton University Press, 2004); and Michael Heumer, *Ethical Intuitionism* (New York: Palgrave, 2005), Ch. 5.

justified *in* believing them.”¹³ Moral foundationalists, in this sense, disagree about which propositions are self-evident. Henry Sidgwick holds that it is self-evident that “the good of any one individual is of no more importance from the point of view (if I may say so) of the universe than the good of another” and that “as a rational being I am bound to aim at good generally – so far as it is attainable by my efforts – not merely a particular part of it.”¹⁴ According to G.E. Moore, it is self-evident that certain things (e.g. pleasure, aesthetic appreciation, beauty) are intrinsically good.¹⁵ W.D. Ross and Robert Audi claim that it is self-evident that we have certain *prima facie* duties.¹⁶ While these theorists disagree about which propositions are self-evident, they agree that some moral principles are self-evident.

Two points of clarification about this version of moral foundationalism are in order. First, self-evidence should not be confused with obviousness. The elementary principles of logic are good candidates for being self-evident but, as any teacher of logic can attest, they are not obvious to students who are attending to them for the first time. Second, one can be justified in believing a self-evident moral principle without believing that the principle is self-evident. Thus, these principles are justified independently of one’s further beliefs, including one’s beliefs about which, if any, principles are self-evident.

A different version of moral foundationalism holds that foundational moral beliefs are not self-evident moral principles but the propositional content of seemings,

¹³ Audi, *Epistemology*, p. 94, emphasis in original.

¹⁴ Sidgwick, *The Methods of Ethics*, p. 382.

¹⁵ Moore, *Principia Ethica*.

¹⁶ Ross, *The Right and the Good* and Audi, *The Good in the Right*.

appearances, or intuitions.¹⁷ According to these views, we have a certain mental state with propositional content that is called ‘a seeming’ or ‘an appearance.’¹⁸ We often describe these mental states by using phrases like ‘it seems that...’ or ‘it appears that...’ We have perceptual seemings (“it seems that Mary is taller than Jane”), seemings from memory (“I seem to remember that she was at the party”), and seemings from introspection (“it seems to me that I am in pain”). We also have mathematical seemings (“it seems that $7+5=12$ ”) and, particularly important for our purposes, we have moral seemings (“it seems wrong to push the fat man”). When these seemings are not derived from perception, memory, or introspection, we might call them ‘intuitions.’

On this view we are justified in believing the propositional contents of these seemings, appearances, or intuitions independently of any other beliefs. It is important to forestall a common misunderstanding. Suppose I have the intuition that murder is wrong. It is important to note that what is foundational, according to this view, is the belief that murder is wrong, *not* the belief that *it seems that* murder is wrong. This is important, because I may never form any beliefs about the seeming. The justification of the proposition ‘murder is wrong’ is foundational, it need not be derived from any belief about what seems to me to be the case.

My argument against these views proceeds in two stages. I begin by defending the following thesis: given our awareness of disagreement among epistemic peers regarding our moral beliefs, we are not justified in believing any moral proposition unless

¹⁷ This view is defended by Huemer, *Ethical Intuitionism*, Ch. 5. My exposition of this view follows Huemer’s, especially pp. 99-106.

¹⁸ For a useful account of these mental states see William Tolhurst, “Seemings,” *American Philosophical Quarterly* 35: 293-302.

we have a further belief that serves as a reason to believe that moral proposition. I then argue that if this thesis is true, moral foundationalism is in trouble.

1.4 Disagreement

This section defends the thesis that, given our awareness of disagreement among epistemic peers regarding our moral beliefs, we are not justified in believing any moral proposition unless we have a further belief that serves as a reason to believe that moral proposition.

As it stands, the thesis to be argued for is in need of quite a bit of clarification. Much of this clarification will have to be done as we go along but one term is in need of defining at the outset. As I understand the term, an agent's 'epistemic peer' is one who has roughly the same evidence, cognitive powers (e.g. intelligence, creativity, etc.), and motivation to arrive at the truth as the agent. In what follows, I argue that awareness of peer disagreement regarding our moral beliefs can change the epistemic status of these beliefs.

This idea that peer disagreement can change the epistemic status of our moral beliefs goes back to an oft-quoted passage from Sidgwick:

If I find any of my judgments, intuitive or inferential, in direct conflict with a judgment of some other mind, there must be error somewhere: and if I have no more reason to suspect error in the other mind than in my own, reflective comparison between the two judgments necessarily reduces me temporarily to a state of neutrality.¹⁹

¹⁹ Sidgwick, *Methods of Ethics*, p. 342.

This idea persists.²⁰ Consider the following passage from Russ Shafer-Landau.

It is true that awareness of disagreement regarding one's moral endorsements may serve as a defeater. It will do so if one has nothing to say on behalf of one's moral views, after receiving or conceiving of a challenge from a dissenter whose conflicting views are themselves coherent, compatible with the non-moral evidence, etc. Crucially, one is in a different epistemic position before and after confronting such disagreement. Prior to this sort of confrontation, one may be justified in one's belief simply because of having understood a self-evident proposition. But after the challenge is issued, one is required to defend oneself.²¹

The crucial idea in Shafer-Landau's passage is that awareness of moral disagreement serves as a defeater for the moral beliefs about which there is disagreement. A defeater for a belief B is a further belief that cancels or overrides an agent's justification for believing B.²² In the above passage, Shafer-Landau is noting that a further belief, a belief that there is disagreement regarding one's moral endorsements, can cancel or override one's justification for holding one's moral views. While one may have been justified in holding a particular moral belief without having further beliefs that serve as reasons for that belief, after encountering peer disagreement, one now requires further reason in order to be justified in holding one's belief.

²⁰ In addition to the passage from Shafer-Landau quoted below, see Walter Sinnott-Armstrong, "Moral Relativity and Intuitionism" *Philosophical Issues* 12: 305-82 and Sarah McGrath, "Moral Disagreement and Moral Expertise" in Russ Shafer-Landau (ed.) *Oxford Studies in Metaethics* Vol. 3 (Oxford: Oxford University Press, 2008), pp. 87-108.

²¹ Shafer-Landau, *Moral Realism*, p. 265.

²² This definition is intended to be stipulative. I believe that it accords well with Shafer-Landau's usage. See *Moral Realism*, p. 289.

It is now widely accepted that in the non-moral case, peer disagreement can serve as a defeater in the way Shafer-Landau suggests. Consider the following well-known example.

Suppose that five of us go out to dinner. It's time to pay the check, so the question we're interested in is how much we each owe. We can all see the bill total clearly, we all agree to give a 20% tip, and we further agree to split the whole cost evenly, not worrying over who asked for imported water, or skipped desert, or drank more of the wine. I do the math in my head and become highly confident that our shares are \$43 each. Meanwhile, my friend does the math in her head and becomes highly confident that our shares are \$45 each. How should I react, upon learning of her belief?²³

Let us suppose that I know that my friend is genuinely my epistemic peer. We might, "suppose that my friend and I have a long history of eating out together and dividing the check in our heads, and that we've been equally successful in our arithmetic efforts: the vast majority of times, we agree; but when we disagree, she's right as often as I am."²⁴

It seems to most people who consider this case that I should suspend judgment about whether or not we owe \$43 each. If the proper attitude towards the proposition 'we owe \$43 each' is suspension of judgment, then it follows that I am not justified in believing that we owe \$43 each. I may have been justified in holding this belief before I learned of my friend's disagreement but now that I am aware of my friend's disagreement I should adopt a different attitude toward the proposition that we owe \$43 each. The way

²³ Taken from David Christensen, "Epistemology of Disagreement: The Good News," *Philosophical Review* 116: 187-217, p. 193.

²⁴ *Ibid.*

this case seems to function is that my awareness of peer disagreement defeats my justification for believing that we owe \$43 each. I may become justified in holding this belief again, if, for example, the three other persons at the table perform the appropriate calculations and come to the conclusion that we owe \$43 each. However, what has happened in this case is that I have adopted a further belief that serves as a reason to believe that we owe \$43 each. It seems that a further belief that serves as a reason to believe that we owe \$43 each is *necessary* for my belief that we owe \$43 to regain justification.²⁵

I will argue that the moral case is relevantly similar. Suppose I am eating lunch with my friends Judith and Shelly, when Judith tells us the following story:

George is on a footbridge over the trolley tracks. He knows trolleys, and can see that the one approaching the bridge is out of control. On the track back of the bridge there are five people; the banks are so steep that they will not be able to get off the track in time. George knows that the only way to stop an out-of-control trolley is to drop a very heavy weight into its path. But the only available, sufficiently heavy weight is a fat man, also watching the trolley from the footbridge. George can shove the fat man onto the track in the path of the trolley, killing the fat man; or he can refrain from doing this, letting the five die.²⁶

Upon hearing this story, I immediately judge that it would be wrong to push the fat man off the bridge. I form the belief that this action is wrong and perhaps this belief is

²⁵ A further belief that serves as a reason to believe that we owe \$43 each is necessary but perhaps not sufficient. My belief that I am a reliable at dividing the check in my head is a further belief that is a reason to believe that we owe \$43 each but it may not be sufficient to justify me in believing that we owe \$43 each in cases of peer disagreement.

²⁶ Judith Jarvis Thomson, *Rights, Restitution, and Risk* (Cambridge, MA: Harvard University Press, 1986), p. 82.

justified. However, suppose Shelly has exactly the opposite reaction, he immediately judges that it would be morally permissible to push the fat man off the bridge.

There are many things that could explain why my friend had a different reaction to the case than I did. Perhaps my friend misunderstood the case, subscribes to a false moral theory, or reasoned poorly about the case. However, as Sidgwick notes, if we are truly epistemic peers it is just as likely that I made an error as it is that my friend made an error.

If we believe that my friend's disagreement in the restaurant in the case served as a defeater for my belief that we owe \$43 each, then we should accept that Shelly's disagreement in this case serves as a defeater for my judgment that it is morally impermissible to push the fat man off the bridge. Of course, this does not mean that I cannot become justified once again in believing that it was impermissible to push the fat man off the bridge. I could learn of the doctrine of double effect and note that this doctrine would forbid pushing the fat man. Perhaps, inferring my judgment about pushing the fat man from the doctrine of double effect could render my judgment justified. However, this requires a further belief in order to render my judgment about the fat man case justified.

So far, our discussion of moral disagreement has been restricted to cases where I make an immediate judgment, (i.e. cases where my belief is not based on further belief that serves as a reason for my belief). However, suppose my belief is based on further beliefs that serve as reasons for my belief. Does my awareness of peer disagreement still function as a defeater?

We need not take a stand on this issue here because the thesis of this section

regarding disagreement only requires that my awareness of peer disagreement function as a defeater when I have no further beliefs that serve as reasons for my belief. This thesis leaves open what, if anything, is required of me when I encounter peer disagreement about a belief for which I have further reasons. Thus, the claim made here is rather weak. It is only that peer disagreement functions as a defeater for a belief when I have no further beliefs that serve as reasons for holding that belief.

The weakness of this claim about disagreement is worth emphasizing. Currently there is a dispute in epistemology about what peer disagreement requires of us. Some think that peer disagreement requires us to radically alter our confidence in our beliefs. For instance, Richard Feldman writes that in cases where there is widespread peer disagreement, “suspension of judgment is the proper attitude. It follows that in such cases we lack reasonable belief and so, on standard conceptions, knowledge.”²⁷ In contrast, Thomas Kelly writes, “disagreement does not provide a good reason for skepticism or to change one’s view.”²⁸ The position argued for here is merely that awareness of disagreement is a defeater in certain cases, namely, ones in which one does not have a further belief that serves as a reason for their belief. Since most of our beliefs are supported by further reasons, we can, at least most of the time, agree with Kelly that disagreement does not give us a reason to change our view or become skeptics.

I have been arguing that awareness of peer disagreement is a defeater for a moral

²⁷ Richard Feldman, “Epistemological Puzzles about Disagreement,” in Stephen Hetherington (ed.), *Epistemology Futures* (Oxford: Oxford University Press, 2006), pp. 216-236, p. 217.

²⁸ Thomas Kelly, “The Epistemic Significance of Disagreement,” in Tamar Gendler and John Hawthorne (eds.), *Oxford Studies in Epistemology* (Oxford: Oxford University Press, 2006), pp. 167-196, p. 170.

belief and that to overcome this defeater requires at least a further belief that serves as a reason for the moral belief in question. However, this only supports my general thesis that we require further reasons for all of our moral beliefs if it turns out that peer disagreement is widespread enough to provide a defeater for all of our moral beliefs. In what follows, I argue that this is the case.²⁹

Moral philosophy provides us with ample evidence. Deontologists disagree with consequentialists about which moral theory is correct. Practical ethicists disagree about whether abortion and euthanasia are morally permissible. In the extreme case, moral nihilists such as J.L. Mackie and Richard Joyce hold that no first order moral propositions are true.³⁰ This is especially important, for it means that there is a potential defeater for all of our moral beliefs.

One line of response is to deny that moral nihilists are our epistemic peers. One who responded this way would not deny that there is widespread moral disagreement, but could at least resist my claim that there is moral disagreement about every moral proposition. One who is tempted toward this response should keep in mind how I have characterized epistemic peers. Our epistemic peers are those who have cognitive powers, motivation to arrive at the truth, and evidence that is roughly equivalent to ours. I unfortunately cannot prove that Mackie, Joyce, and other nihilists are our peers in this sense. Nonetheless, it seems to me likely that they are.

²⁹ A qualification: I take it that there is no peer disagreement about the truth of certain conditionals (e.g. ‘if killing is always wrong, then killing your child is wrong.’). This fact is of no help to any actual moral foundationalist because no moral foundationalist holds that it is only these sorts of conditionals that are foundational.

³⁰ J.L. Mackie, *Ethics: Inventing Right and Wrong* (New York: Penguin, 1977), Ch. 1 and Richard Joyce, *The Myth of Morality* (Cambridge, UK: Cambridge University Press, 2001).

Another possible line of response is to concede that there is quite a bit of disagreement about what one morally ought to do all things considered but that there is little disagreement that we have certain *prima facie* duties.³¹ For example, it may not be particularly controversial that one has *prima facie* duty not to lie. To assess this claim we must keep in mind that to say that we have a *prima facie* duty not to lie is not merely to imply that we often or usually have a duty not to lie. To claim that we have a *prima facie* duty not to lie is to claim that we have an all things considered duty not to lie unless lying is required by another duty.³² This latter claim *is* controversial. It is denied by nihilists like Joyce and Mackie, but also by non-nihilists such as Jonathan Dancy.³³ Many utilitarians are also committed to its denial. To see this, imagine a case where telling a lie is not required by any duty and produces exactly as much overall utility as not lying. In such a case the claim that we have a *prima facie* duty not to lie entails that we have an all things considered duty not to lie, since, by hypothesis, lying is not required by any other duties. However, the act-utilitarian will hold that we have no duty not to lie since, by hypothesis, lying produces exactly as much utility as not lying. Note that since the claim that we have a *prima facie* duty not to lie entails that, in this case, we have an all things considered duty not to lie, utilitarians who hold that we do not have an all things considered duty not to lie are committed to denying that we have a *prima facie* duty not to lie as well.

A third line of response is to concede that it *appears* that moral disagreement is

³¹ I thank Robert Audi and Bruce Russell for pressing me on this point.

³² This is Ross's understanding of *prima facie* duty. See Ross, *The Right and the Good*, p. 19.

³³ Jonathan Dancy, *Moral Reasons* (Oxford: Blackwell, 1993), Ch. 6

extremely widespread, but hold that the appearances are misleading. In many cases the sentences one person utters seem to contradict the sentences uttered by another person. However, it does not follow that the propositions one person intends to assert contradict the proposition another person intends to assert. This is because the relationship between the sentences we utter and the propositions we intend to assert is complicated. Many times it appears that two sentences are contradictory but only because terms are being used imprecisely. For example, suppose I believe that there is no omniscient, omnipotent, and perfectly moral being who created the universe. I come across a friend who utters the sentence 'God exists' and I utter in reply 'God does not exist.' In this case, the sentences we utter make it appear that we disagree. However, suppose my friend believes that God is love and only means to be asserting that love exists. In this case we may not actually disagree, since the proposition my friend is intends to assert is 'love exists' and that proposition is consistent with my atheism.

The sentences we utter are indeed imperfect guides to the propositions that we intend to assert. However, this does not show that there is less disagreement than there appears to be. I may *assent* to the sentence you utter only because I am mistaken about which proposition you intend to assert. Perhaps if I really knew what proposition you were trying to assert, I would dissent rather than assent to your statement. If this happens often, there may actually be *more* disagreement than there appears to be. Thus, the fact that the sentences we utter are imperfect guides to the propositions we intend to assert does not show that there is less moral disagreement than there appears to be.

I take it, then, that there is strong *prima facie* reason to think that peer disagreement about morality is indeed widespread enough to provide a defeater for all of our moral

beliefs. This *prima facie* reason is not defeated by either the objection from *prima facie* duties nor the fact that the sentences we utter and the propositions that we intend to assert come apart. If we are aware of extremely widespread peer disagreement about morality and such disagreement functions as a defeater in the way I have suggested, then it follows that we (those of us aware of widespread moral disagreement) must have a further reason for our moral beliefs in order for them to be justified.

There is, however, an objection to this line of thought that must be considered. Suppose Peter believes it is wrong to torture a puppy for amusement but David does not. However, Peter knows that David was bitten by a dog as a child and has had an irrational disdain for canines ever since. Peter does not have a further reason to believe that it is wrong to torture dogs for amusement, however it does not seem that David's disagreement in this case in any way undermines Peter's justification for holding his belief. The way that this case functions is that while normally David's disagreement would be a defeater for Peter's belief, this defeater is itself defeated by Peter's knowledge that David has an irrational disdain for canines. Generalizing from this case, one might hold that there is a set of cases in which one need not hold a further belief that serves as a reason for their beliefs even in the face of disagreement. These cases occur when one knows that the disagreement itself is motivated by beliefs that are irrational, delusional, etc.

In order to accommodate this objection, the thesis must be qualified somewhat. Let us understand 'reasons to believe that p' in a very broad sense to include not only evidence for p but also defeaters for potential evidence for $\sim p$. I recognize that it is a bit unnatural to call a defeater for potential evidence for $\sim p$, a reason to believe that p,

however, I am using ‘reason to believe that p’ in a somewhat technical sense that is slightly divorced from its ordinary usage. In this technical sense, one has a reason to believe p if and only if one has evidence for p or a defeater for would-be evidence that \sim p.

1.5 Disagreement and Moral Foundationalism

In this section, I argue that the thesis defended in the last section raises serious problems for moral foundationalism. I begin by restating what moral foundationalism is

Moral Foundationalism: A moral belief is justified if and only if it is a *basic moral belief*, a moral belief that is non-inferentially justified, or it bears an appropriate inferential or evidential relationship to a foundational moral belief.

This formulation intentionally leaves some things open to allow for a variety of different views to count as instances of moral foundationalism. For instance, it leaves open what sort of inferential or evidential relationships non-foundational moral beliefs must bear to moral beliefs in order to be justified. Some older foundationalist views held that a non-foundational belief had to be deductively inferable from a foundational belief. However, contemporary foundationalists are much more liberal about what sorts of inferential and evidential relationships can justify a non-foundational belief and this formulation allows for that.

This formulation also leaves open the question of whether or not basic moral beliefs are indefeasibly justified or only defeasibly justified.³⁴ A defender of the former view might hold that we are indefeasibly justified in believing a self-evident proposition so long as we understand it or they might hold that we are indefeasibly justified in believing

³⁴ For an especially clear overview of this distinction and its consequences see Audi, “Moderate Intuitionism and the Epistemology of Moral Judgment.”

the propositional content of our seemings. This view is in direct conflict with the argument from section 1.4. For on this view there are some moral propositions for which our justification could not be defeated. Fortunately, this view is implausible and rarely defended today. On this view a self-evident proposition could be radically incoherent with my other beliefs, yet I would still be justified in believing the proposition so long as I understood it.

Contemporary moral foundationalists defend the view that foundational moral beliefs are only defeasibly justified. On this view, understanding a self-evident proposition renders us *prima facie* justified in believing the proposition. This *prima facie* justification could be defeated if, for example, this proposition fails to cohere with our other beliefs. Similarly, one could hold that I am *prima facie* justified in believing the content of my moral intuitions. However, this justification could be defeated. These views are strictly compatible with the thesis of section I. This is because they allow that there could be a defeater for foundational moral beliefs. However, I will argue that these views can only accommodate the thesis of section I at a high cost.

The first thing to notice is that while one can be *prima facie* justified in believing a moral proposition non-inferentially, one cannot be all things considered justified in believing a moral proposition unless they have a further belief that serves as a reason to believe it. This is because, as argued in section 1.4, our awareness of extremely widespread peer disagreement about morality serves as a defeater of our moral beliefs unless we have a further reason for them.

Notice also that these further beliefs that serve as reasons to believe foundational moral beliefs must themselves be justified. This is because an unjustified belief cannot

override or undermine a defeater. Thus, even foundational moral beliefs must ultimately be supported by further justified beliefs.

This fact calls into question the regress argument that we began with. For a belief to stop the regress that belief must have all things considered justification. *Defeated prima facie* justification cannot stop the regress. However, we have seen that for foundational moral beliefs to have all things considered justification they must be supported by further beliefs and that these further beliefs must themselves be justified beliefs. We can then sensibly ask, what makes these further reasons justified. The regress begins again.

It follows from these considerations that no moral belief can stop the regress. This is because our awareness of extremely widespread peer disagreement about the truth of our moral beliefs serves as a defeater that can only be overcome or undermined by a further reason that is itself a justified belief. Thus, to achieve all things considered justification, a moral belief always requires a further justified belief.

If no moral belief can stop the regress, the threat of skepticism looms. For it may seem that unless something can stop the regress, given the implausibility of an infinite regress of justified beliefs, we may be forced to accept that none of our moral beliefs are justified. However, as we have already seen there is another way of avoiding the regress argument without positing regress stoppers.

This alternative is to deny the linear picture of justification on which the regress argument rests. However, for the foundationalist to take this route is to make a major concession to the coherentist. This combined with the earlier concession, that all moral

beliefs require further justified beliefs in order to be justified, produces a view that is difficult to distinguish from coherentism.

It is worth spelling out explicitly what the moral foundationalist would have to give up. This argument does *not* show that there are no self-evident moral propositions, so long as self-evidence is understood as conferring only defeasible justification on an agent's beliefs.³⁵ What the argument does show is that our *prima facie* justification for believing self-evident moral propositions, if there are any, is defeated by our awareness of extremely widespread moral disagreement. Thus, in practice, self-evident propositions will function much like non self-evident propositions, in that they require evidential or inferential support from other beliefs in order to be justified.

This point is worth dwelling on because it is one that moral foundationalists have traditionally denied. For instance G.E. Moore claims that it is self-evident that certain things are intrinsically good and he defines a self-evident proposition as one that is "evident or true by *itself* alone; that it is not an inference from some proposition other than itself."³⁶ W.D. Ross holds that the proposition that certain acts are *prima facie* right

³⁵ Thus, this argument is distinct from the standard objection that holds that no moral belief can be self-evident because persons disagree about the truth of allegedly self-evident moral beliefs. The standard objection has been widely discussed by moral foundationalists (see Philip Stratton-Lake, "Introduction," pp. 1-28; and Audi, *The Good in the Right*, pp. 60-68). The argument discussed here is less often discussed. It does not claim that the fact of widespread moral disagreement shows that there are no self-evident moral propositions. Instead, this argument claims that even if there are self-evident moral propositions, the *prima facie* justification of these propositions is defeated. This sort of argument has received little attention from moral foundationalists. An exception is Russ Shafer-Landau, *Moral Realism*, pp. 252-265 who distinguishes between these two arguments and discusses versions of both of them. However, Shafer-Landau does not discuss the fact that once the moral foundationalist concedes that further beliefs are required in order for a moral belief to be justified, the regress begins again.

³⁶ Moore, *Principia Ethica*, p. 143, emphasis in original

“is self-evident; without any need of proof, or of evidence beyond itself.”³⁷ And Robert Audi claims that it is self-evident that we have certain *prima facie* duties where “a self-evident proposition is knowable without our relying on any inferential *ground* for it.”³⁸ If the thesis of the last section is correct, however, one cannot be justified in believing a moral proposition unless one has a further belief that serves as a reason to believe the proposition.

As I noted earlier, this raises the question of how one can stop the regress. I suggested one plausible way of stopping the regress is to deny the linear picture of justification on which the regress argument depends. However, to deny the linear picture of justification is to deny much of what makes foundationalism a distinctive position. As we have seen, once one adopts a non-linear picture of justification there is no need to posit foundational beliefs on which other beliefs rest. We have seen, then, that in order to account for the epistemic obligations we incur as a result of moral disagreement, the moral foundationalist has to give up exactly those claims that make their view distinct from moral coherentism. I now move on to discuss another version of foundationalism in moral epistemology, non-moral normative foundationalism.

1.6 Non-Moral Normative Foundationalism

The non-moral normative foundationalist agrees with the moral foundationalist that a moral belief is justified if and only if it can be inferentially justified by appeal to a basic belief. However, the non-moral normative foundationalist disagrees with the moral foundationalist about which beliefs are basic. Whereas the moral foundationalist claims

³⁷ Ross *The Right and the Good*, p. 29.

³⁸ Audi, *The Good in the Right*, p. 51, emphasis in original

that the basic beliefs that justify all justified moral beliefs are themselves moral beliefs, the non-moral normative foundationalist insists that they are non-moral normative beliefs.

To get an idea of how non-moral normative foundationalism might be fleshed out, consider the moral theories of R.B. Brandt and David Gauthier.³⁹ Both theorists begin by developing theories of practical rationality. They then argue from these theories of rationality to particular moral theories. One might, inspired by such views, adopt the position that some non-moral normative claims, such as claims about practical rationality, are self-evident and that moral claims are justified by being part of an inferential chain that terminates in a self-evident non-moral normative belief

Neither Brandt nor Gauthier claim that any part of their theories of rationality is epistemically basic. Therefore, Brandt and Gauthier cannot be said to have endorsed any kind of foundationalism.⁴⁰ In fact either theorist could, consistently with the main thrust of their writings, adopt coherentist positions according to which moral beliefs and beliefs about rationality are justified by being part of a coherent set of beliefs that may include beliefs about rationality and morality, as well as beliefs about psychology, economics,

³⁹ As defended by Brandt in *A Theory of the Good and the Right* and Gauthier in *Morals by Agreement* (Oxford: Clarendon Press, 1986).

⁴⁰ For this reason Dale Jamieson claims in his “Method and Moral Theory” in Peter Singer (ed.), *A Companion to Ethics* (Oxford: Blackwell, 1991), that the views of Brandt and Gauthier are *neither* coherentist nor foundationalist. I think that it is best to say that such views are *either* coherentist or foundationalist. They are foundationalist if they take non-moral normative beliefs to be basic and coherentist if they take them to be justified in the way described in the main text.

and perhaps even sociology. Such a view would be coherentist and, I think, highly plausible.⁴¹

Still, even if no one has explicitly defended non-moral normative foundationalism, a foundationalist frustrated by the failure of moral foundationalism may be tempted to take up such a view. Let me attempt to block this temptation by noting that non-moral normative foundationalism faces the same problem as moral foundationalism. The argument against moral foundationalism can be run *mutatis mutandis* against non-normative moral foundationalism

One may object that there is not such widespread disagreement about what is rational. I grant that there may be claims about practical rationality that are uncontroversial. But what is needed for non-moral normative foundationalism is, not just uncontroversial claims about practical rationality, but uncontroversial claims about practical rationality from which one can derive moral requirements.⁴² I conjecture that there is peer disagreement about any such claim.

⁴¹ For more on the epistemological commitments of views like Gauthier's see Christopher W. Morris, "A Contractarian Account of Moral Justification" in Sinnott-Armstrong and Timmons (eds.), *Moral Knowledge*, pp. 214-242. Morris notes that Gauthier and other contractarians are committed to the view that certain non-moral normative beliefs have some epistemic priority over moral beliefs. This may be incompatible with some radical coherentist views but a plausible moderate coherentism can allow for some notion of epistemic priority. To be coherentist such views would have to reject the kind of radical epistemic priority that foundationalists are committed to. However, it does not seem to me that contractarian views like Gauthier's presuppose any radical foundationalist notion of epistemic priority. Thus, contractarian views like Gauthier's can be either coherentist or foundationalist.

⁴² For skepticism about whether one can derive moral requirements from practical rationality see James Drier, "Humean Doubts about the Practical Justification of Morality" in G. Cullity and B. Gaut (eds.) *Ethics and Practical Reason* (Oxford: Oxford University Press, 1997), pp. 81-100 and Mackie, *Ethics*, Ch. 1.

If my conjecture is correct and disagreement functions as a defeater in the way I suggested in section 1.4, then it follows that those beliefs about practical rationality from which moral requirements can be derived are unable to stop the regress in the same way that moral beliefs are unable to stop the regress. If this is right, then those theorists who hold that moral requirements can be derived from the requirements of practical rationality would be better off being coherentists.

1.7 Descriptive Foundationalism

Descriptive foundationalism is also a version of foundationalism about moral justification in that it holds that a moral belief is justified if and only if it is inferentially justified by appeal to a basic belief. However, it differs from moral and non-moral normative foundationalism because it takes some descriptive belief(s) to be the basic belief(s) from which all moral beliefs are derived.

The problem with descriptive foundationalism is that it violates Hume's law, according to which one cannot derive an 'ought' from an 'is.'⁴³ The fact that God commands me not to murder only entails that I ought not to murder if we assume that I ought to do what God commands. Similarly, the fact that a certain experiment would cause babies pain only entails that I ought not to conduct the experiment if we assume that I ought not to cause babies pain. It seems that descriptive 'is' claims do not, by themselves, tell me what I ought to do. To arrive at a claim about what I ought to do, I must also assume a further 'ought' claim.

⁴³ Hume appears to state Hume's law in *A Treatise of Human Nature*, Book III, Part I, Section I (final paragraph). Some interpreters (e.g. A.C. MacIntyre, "Hume on 'Is' and 'Ought,'" *Philosophical Review* 68 (1959): 451-68) doubt that Hume actually intended to state what is today called 'Hume's law.' In any case, the philosophical plausibility of the law is independent of whether or not Hume affirmed it.

Some philosophers have employed logical tricks that actually allow them to derive an 'ought' from an 'is.' Take the following examples:

Argument A:

A1) Bob is married and not married.

A2) Therefore, you ought not to lie.

A2 follows from A1 since A1 is contradiction and anything follows from a contradiction.

Argument B:

B1) Grass is green

B2) Therefore, grass is green or you ought not to lie.

While B2 follows from B1, in this case it is not so obvious that B2 is an 'ought.' Perhaps B2 is an 'is.' But consider the following argument

Argument C:

C1) Grass is green or you ought not to lie

C2) Grass is not green

C3) Therefore, you ought not to lie.

If the disjunction, 'grass is green or you ought not to lie' is an 'is,' then argument C derives an 'ought' from an 'is.' If the disjunction is an 'ought' then argument B derives an 'ought' from an 'is.'

I am willing to concede that strictly speaking one can derive an 'ought' from an 'is.' But these arguments are of no help to the descriptive foundationalist because for descriptive foundationalism to be true, there must be an argument that derives 'ought' from an 'is' in a way that confers epistemic justification on the 'ought.' These arguments do not do this. This is most obvious with the first argument, since the first argument can

only work if the first premise is a contradiction, and all contradictions are false. Surely, a false claim cannot confer justification on an ‘ought.’ The second and third arguments depend on ad hoc gerrymandered disjunctions, and thus are also unlikely to serve as an epistemic foundation for moral ‘ought’s.

This has not been anything like a complete survey of attempts to derive an ‘ought’ from an ‘is.’⁴⁴ Attempting such a survey would take us too far afield. However, no attempt to derive an ‘ought’ from an ‘is’ has yet gained widespread acceptance among philosophers. I suspect that this is for good reason.

1.8 Quasi-Foundationalism

I want to close by considering quasi-foundationalism, a view suggested, though not endorsed, by Mark Timmons.⁴⁵ The quasi-foundationalist believes in a coherence theory of justification for non-moral beliefs. The quasi-foundationalist then picks out a particular non-moral belief (or set of non-moral beliefs) and argues that all moral beliefs must be justified inferentially by this non-moral belief (or set of non-moral beliefs).

Strictly speaking the quasi-foundationalist is not a foundationalist. This is because she denies that there are any basic beliefs. According to her, moral beliefs are justified inferentially by appeal to a non-moral belief that is justified on coherentist grounds. However, the quasi-foundationalist also disagrees with the moral coherentist. This is because the moral coherentist thinks that moral beliefs are justified by being part of a coherent set of beliefs. As such, moral beliefs are ultimately justified, for the moral

⁴⁴ For a more thorough survey, see Walter Sinnott-Armstrong, *Moral Skepticisms* (Oxford: Oxford University Press, 2006), Ch. 7.

⁴⁵ Mark Timmons, “Foundationalism and the Structure of Ethical Justification,” *Ethics* 97 (1986): 596-609.

coherentist, non-linearly. However, the quasi-foundationalist holds that moral beliefs (but not non-moral beliefs) are justified linearly, by appeal to a non-linearly justified non-moral belief.

This view, if plausible, would constitute a serious rival to coherentism about moral justification. So, is this view plausible? Notice that if the non-moral belief in question is a purely descriptive belief, then quasi-foundationalism will violate Hume's law and be implausible for the same reason that descriptive foundationalism is implausible.⁴⁶

But what if the non-moral belief in question is itself a normative belief? Hume's law does not say that we cannot derive an 'ought' from another 'ought.' I do not have a knockdown argument against this form of quasi-foundationalism. However, I want to note that there is something peculiar about it.

On this view some non-moral normative beliefs would be part of a coherentist web with a foundationalist building sticking out of it. These non-moral normative beliefs are justified in a non-linear fashion while justifying moral beliefs in a strictly linear fashion. The obvious question is, why should the non-moral normative beliefs be part of the web and not the building? In other words, why should some normative beliefs be justified non-linearly while moral ones are justified linearly?

If the quasi-foundationalist gives us no answer, then the claim that moral beliefs must be justified linearly is purely ad-hoc. Thus, one who holds the quasi-foundationalist view must hold that there is some difference between the moral and the rest of the

⁴⁶ Timmons acknowledges this in a footnote (p. 609 n27). However, he seems to be more sanguine about the possibility of deriving an 'ought' from an 'is' (in a way that confers epistemic justification on the 'ought') than I am.

normative that compels us to adopt a different epistemology for the moral than the rest of the normative. It is difficult for me to see what this difference could be.

As I said earlier, there is no knockdown argument against the quasi-foundationalist here. Until the view is further developed, the most I can do is make a burden-shifting move. The quasi-foundationalist holds that the non-moral normative is justified non-linearly while the moral is justified linearly. One who defends this claim inherits the burden of explaining why the moral and the rest of the normative should be treated so differently. I am doubtful that such a burden can be met. If I am right that it cannot be, then quasi-foundationalism preserves the idea that moral beliefs must be justified linearly only by making an ad hoc maneuver.

1.9 Conclusion

In this chapter I have argued that three versions of foundationalism in moral epistemology are implausible. I also considered a view that, while not strictly foundationalist, does pose a threat to coherence theories of ethical justification. I argued that this view has a peculiar entailment in need of defense.

I have criticized these views in order to motivate the search for an alternative. Beginning in chapter three, I develop and defend a coherentist account, which I argue avoids the pitfalls of the views discussed in this chapter. Before we do that, however, we will take a quick detour out of moral epistemology and into moral psychology. Concerns from moral psychology will force us to rethink some traditional motivations for coherence accounts and hopefully allow us to develop a novel defense of coherentism in moral epistemology

Chapter 2: Skeptical Doubts about Moral Intuitions

2.1 Moral Psychology

If one's only exposure to moral discourse and practice were moral philosophy, one might think that moral judgments about actions are arrived at by conscious inference from abstract principles and facts about the actions in question. In fact, this was the dominant view of how mature moral reasoning worked in moral psychology for most of the twentieth century.¹

Things changed quite a bit in the twenty first century when the psychologist Jonthan Haidt asked participants in a study about the following case:

Julie and Mark are brother and sister. They are traveling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it might be interesting and fun if they tried making love. At the very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making love, but they decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that? Was it OK for them to make love?²

Haidt found that,

Most people who hear the above story immediately say that it was wrong for the siblings to make love, and they then begin searching for reasons. They point out the dangers of inbreeding, only to remember that Julie and Mark used two forms of birth control. They argue that Julie and Mark will be hurt, perhaps emotionally, even though the story makes it clear that no harm befell them.

¹ For examples see Jean Piaget, *The Moral Judgment of the Child* (New York: Free Press, 1932) and Lawrence Kohlberg *The Philosophy of Moral Development* Vol. 1 (San Francisco: Harper & Row, 1981).

² Haidt, "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment" *Psychological Review* 108 (2001): 814-34.

Eventually, many people say something like, “I don’t know, I can’t explain it, I just know it’s wrong.”³

If people arrived at moral judgments via conscious inference, we might expect them to take some time to arrive at a conclusion about the moral permissibility of Mark and Julie making love. Yet Haidt makes it clear that most of the participants in his study responded immediately. People are able to state a reason for judging Mark and Julie’s lovemaking impermissible, but when that reason is shown not to be relevant to this case, people typically respond by inventing a second or perhaps a third reason. Haidt observed that the typical participant behaved like “a lawyer trying to build a case rather than a judge searching for the truth.”⁴ When all the arguments ran out, many participants declared that they just knew the lovemaking was wrong, even though they could not provide a reason. Surely, this is not what we would expect if they held their belief on the basis of conscious inference from a moral principle.

In order to explain his observations, Haidt proposed an alternative to the conscious inference model. On Haidt’s model, people have an immediate judgment, an intuition, which they arrive at without conscious inference. People then engage in an *ex post facto* search for reasons to justify their intuition. This model explains why people are able to make moral judgments so quickly, even about unfamiliar cases. It also explains why the conscious inference model was able to dominate moral psychology for so long. When subjects are asked why they made a particular moral judgment, they are able to easily invent reasons for their intuition. But Haidt argues that these reasons are *ex post facto* and not really the cause of the judgment. This explains why people continue to

³ Ibid. p. 814.

⁴ Ibid.

hold on to their judgments even if their reasons are shown not to be relevant. Their judgments were not actually made as the result of the reasons given. Instead, these judgments are made intuitively.⁵

As Haidt's research makes clear, intuitions play a vital role in ordinary moral judgment. In this chapter, I offer a characterization of what intuitions are and a preliminary diagnosis of their reliability.

2.2 Intuitions

In the last section, we saw that intuitions play a crucial role in our moral judgment. But what are these intuitions? This is the question that I hope to answer in this section.

One of the earliest attempts to explicitly state what an intuition is can be found in Descartes' *Rules for the Direction of Our Native Intelligence*.

By 'intuition' I do not mean the fluctuating testimony of the senses or the deceptive judgment of the imagination as it botches things together, but the conception of a clear and attentive mind, which is so easy and distinct that there can be no room for doubt about what we are understanding. Alternatively, and this comes to the same thing, intuition is the indubitable conception of a clear and attentive mind which proceeds solely from the light of reason. Because it is

⁵ It might be objected that rather than having an intuition, perhaps subjects consciously inferred their judgment from a general principle: incest is wrong. While this is a possible explanation, I do not think it is the best. First, when asked why what Mark and Julie did was wrong, subjects tended to cite ways in which Mark and Julie's act could have harmed someone (Mark, Julie, or a potential offspring) rather than a general prohibition against incest. Second, Haidt has found similar results using other stories (Jonathan Haidt, Silvia Koller, Maria Dias, "Affect, Culture and Morality, or Is It Wrong to Eat Your Dog?," *Journal of Personality and Social Psychology* 65 (1993): 613-28) including a story in which a woman cleans her bathroom with an American flag, a story in which a family eats their dog after it has been run over by a car, and a story in which a man eats a chicken carcass after using it to masturbate. While it seems likely that the subjects subscribe to a general prohibition against incest, it is much less likely that subjects subscribe to a general prohibition against eating the family dog when it is deceased or eating animal carcasses that have been used for masturbation.

simpler, it is more certain than deduction, though deduction, as we noted above, is not something a man can perform wrongly (AT 368).⁶

Descartes goes on to contrast intuition with “deduction, by which we mean the inference of something as following necessarily from some other propositions (AT 369).”

Hence we are distinguishing mental intuition from certain deduction on the grounds that we are aware of a movement or a sort of sequence in the latter but not in the former (AT 370)

For Descartes, to call something an ‘intuition’ is to make some psychological claims about it as well as some epistemic claims. Let us begin with the psychological claims. Descartes says that intuitions are distinct from perceptual judgments, judgments arrived at through deduction, and imaginings. Furthermore, intuitions come to us when our minds are “clear and attentive.” As far as the epistemic features of intuitions, Descartes claims that intuitions are “indubitable” and “more certain than deduction” even though deduction is “not something a man can perform wrongly.” Later, he calls intuitions “self-evident (AT 369).”

It is a somewhat peculiar feature of Descartes understanding of ‘intuition’ that ‘intuition’ is *defined* by both psychological and epistemic features. Compare this to perceptual judgments. Most philosophers agree that perceptual judgments have certain psychological and epistemic features (though they disagree about what these features are) but one typically defines perceptual judgments by their psychological features and then argues for their epistemic features. It would seem to beg all kinds of question in epistemology to just define perceptual judgments as having epistemic as well as

⁶ All page references to Descartes will be included in the main text. The page numbers give are those of the Adam and Tannery pagination found in the margins of most translations. The translation I have used is from *The Philosophical Writings of Descartes* (trans. by John Cottingham, Robert Stoothoff, and Dugald Murdoch) (Cambridge, UK: Cambridge University Press, 1985).

perceptual features. Why does Descartes appear to do just this when he is defining intuitive judgments?

We can find the answer when we look at the passage above. Descartes seems to be saying that the psychological features of intuitions *entail* that they also have certain epistemic features. Descartes argues that because intuitive judgments are more simple and clear than judgments arrived at by deduction, that these judgments are more certain than judgments arrived at by deduction. But why think that more simple or clear judgments are more likely to be true? In other words, why think that these psychological features of intuitions always coincide with a certain epistemic status? The objection I am raising here is essentially the same objection Mersenne raised about one of Descartes later works, the *Meditations*.

How can you establish with certainty that you are not deceived, or capable of being deceived, in matters which you think you know clearly and distinctly? Have we not often seen people deceived in matters where they thought their knowledge was as clear as sunlight? Your principle of clear and distinct knowledge thus requires a clear and distinct explanation, in such a way as to rule out the possibility that anyone of sound mind may be deceived on matters which he thinks he knows clearly and distinctly (AT 126).

Descartes' reply is intellectually honest but unsatisfying. He says, "In such cases, I simply assert that it is impossible for us to be deceived (AT 144)." Simple assertion will not do here. We need an account of why the psychological features of intuitions always coincide with a particular epistemic status.⁷

⁷ In fairness to Descartes, he does later give a further argument from God's good nature (AT 144). But many commentators think that to invoke God here is viciously circular since clear and distinct perceptions are used to prove God's existence in the first place. This objection is known as the Cartesian Circle. For references and further discussion see section 6 of Lex Newman, "Descartes' Epistemology", *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/descartes-epistemology/>.

The lesson to be learned from this dialogue between Descartes and Mersenne is that the epistemic features of intuitions are something to be argued for, not assumed at the outset. Thus, in discussing what an intuition is, I will focus on the psychological features of intuitions. What epistemic features these intuitions have will be a central topic of the rest of this dissertation.

In regard to the psychological features of intuitions, Descartes suggests two very important points. The first point is that intuitive judgments are a distinct cognitive process from perceptual judgments and imaginings. If I believe that there is a computer in front of me because I see it, then it would be odd to say that I have the ‘intuition’ that there is a computer in front of me. Intuitions are distinct from imaginings in that imaginings do not aim at truth in the way that intuitions do. Descartes also claims that when we have intuitions we are not “aware of a movement or a sort of sequence (AT 370)” the way we are when we come to know something by deduction. I take the claim that we are not aware of “a movement” when we have an intuition to mean that intuitions are not arrived at by conscious inference.⁸ Thus, if I judge that killing Bob was wrong because I consciously infer that it was from my belief that killing is always wrong, then I have not had an intuition that killing Bob was wrong.

⁸ The qualifier ‘conscious’ is important here because it is at least possible that most of the judgments we make are the result of inference at some non-conscious level. For example, I now judge that there is a computer in front of me. This judgment appears to me to be completely spontaneous. But it is possible that at some non-conscious level this judgment is the product of an inference from my brute sensations and my knowledge of what a computer looks like. I believe that there is some reason to think that intuitions are arrived at by non-conscious inference in the sense that my judgment that there is a computer in front of me is arrived at by non-conscious inference. The reason to suspect that intuitions may be the product of non-conscious inference will become clear in the discussion of heuristics below. See note 21.

The second lesson to draw from Descartes' account of intuitions is that intuitions strike us as being clear and obvious. When we make an intuitive judgment we understand clearly what our intuitive judgment asserts. Furthermore, the assertion has a certain feel of obviousness to it. Borrowing a phrase from Hume, we might say that intuitions have a certain "degree of force and vivacity"⁹

I take these to be the two most crucial features of intuitions. Let me now make two further points about intuitions. First, an intuitive judgment need not be a judgment that is made quickly or automatically. It may be that one only has an intuition after thinking deeply about certain features of a case. As John Rawls puts it:

By the term 'intuitive' I do not mean the same as that expressed by the terms 'impulsive' and 'instinctive.' An intuitive judgment may be consequent to a thorough inquiry into the facts of the case, and it may follow a series of reflections on the possible effects of different decisions.¹⁰

Holding that an intuition can be the product of reflection may seem to blur the line between intuitions and other sorts of judgments. However, a crucial difference remains. Though an intuition can be the product of reflection, this reflection cannot involve conscious inference. As soon as the agent makes a conscious inference, the resulting judgment can no longer be called an 'intuition.'

A second further point about intuitions is that intuitions can be very particular or very general in scope and application. One can have an intuition about a very general principle such as "the right action is that which produces the most overall good" and a particular case such as "it was wrong for Bob to lie to John."

⁹ *A Treatise of Human Nature*, Book I, Chapter 1, Section I, paragraph 3.

¹⁰ "Outline of a Decision Procedure for Ethics" *Philosophical Review* 60 (1951): 177-97. Reprinted in *Collected Papers* (ed. by Samuel Freeman) (Cambridge, MA: Harvard University Press, 1999), pp. 1-19.

To summarize, intuitions have two crucial features. First, they are a type of judgment that is distinct from perceptual judgments, imaginings, and judgments arrived at by conscious inference. Second, intuitions have a certain degree of clarity and obviousness to them. An intuition need not be arrived at impulsively or instinctively; it can be a product of careful reflection, so long as this reflection does not involve conscious inference. In addition, an intuition can be a very general judgment about an abstract principle or a very particular judgment about an actual or hypothetical case.

2.3 Intuitions and Reliability

Now that we have some idea of what an intuition is, we can turn to the issue of how often these intuitions are reliable. This issue is highly contentious with some philosophers taking the view that intuitions are largely reliable and others taking positions that are much more skeptical.¹¹ My own position is a moderate one. I will argue that our intuitions are subject to a range of biases. These biases infect our judgment even when our thinking is not distorted by self-interest or cognitive impairment.¹² In this and the two following sections, I examine the case for this skeptical

¹¹ For examples of philosophers who think that moral intuitions are reliable see Ross, *The Right and the Good*, pp. 39-41; William Tolhurst, "On the Epistemic Value of Moral Experience" *Southern Journal of Philosophy* 29 supp. (1990): 67-87; and F.M. Kamm, *Morality, Mortality* (Oxford: Oxford University Press, 1998), pp. 7-9. For examples of philosophers who are more skeptical see R.M. Hare, "The Argument from Received Opinion" in *Essays on Philosophical Method* (Berkeley: University of California Press, 1972), pp. 117-135; Peter Singer, "Sigwick and Reflective Equilibrium" *Monist* 58 (1975): 490-517. Reprinted in *Unsanctifying Human Life* (Oxford: Blackwell, 2002), pp. 27-50; and R.B. Brandt, *A Theory of the Good and the Right*, Ch. 1

¹² This is what differentiates my position from that of the optimists. I take it that the optimists cited in the last footnote would agree that moral intuitions sometimes give us the wrong answer (though they are not always explicit about this) but that this rarely happens when someone is thinking clearly. My view, instead, is that even when we are thinking clearly, our intuitions are often unreliable.

claim. Some philosophers think that the skeptical case is easily dismissed. I reject these dismissals in 2.6. Nonetheless, I think it is an open question how widespread these biases are and, in section 2.7, I make the case that there are grounds for thinking that the impact of these biases is limited.

2.4 The Reliability of Non-moral Intuitions¹³

The psychologists Amos Tversky and Daniel Kahneman gave the following scenario to a group of 88 undergraduates and the University of British Columbia.¹⁴

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of social justice and also participated in anti-nuclear demonstrations.¹⁵

Participants were then asked to rank the probability that Linda was associated with various occupations and avocations. Among the possibilities were “Linda is active in the feminist movement (F),” “Linda is a bank teller (T),” and “Linda is a bank teller and active in the feminist movement (T +F).” Eighty-five percent of the students said that T+F was more likely than T. Of course, it is a well-known fact of probability theory that the probability of a conjunction cannot be higher than the probability of a single conjunct.

¹³ I became aware of some of the literature discussed in this and the next section after reading Tamar Gendler, “Philosophical Thought Experiments, Intuitions, and Cognitive Equilibrium,” *Midwest Studies in Philosophy* 31 (2007): 68-89 and Walter Sinnott-Armstrong, “Framing Moral Intuitions” in Walter Sinnott-Armstrong (ed.) *Moral Psychology Volume 2* (Cambridge, MA: MIT Press, 2008), pp. 47-76. Though I am not sure that either author endorses my philosophical conclusions, their thinking was a large influence on this and the following section.

¹⁴ Amos Tversky and Daniel Kahneman, “Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment,” *Psychological Review* 90 (1983) : 293-315.

¹⁵ Ibid, p. 297

Tversky and Kahneman next tested a group of first year graduate students in psychology, education, and medicine who had taken one or more courses in statistics as well as a group “of doctoral students in the decision science program of the Stanford Business School who had taken several advanced courses in probability, statistics, and decision theory.”¹⁶ They found that “there is no effect of statistical sophistication on either the direct or indirect tests.”¹⁷

What could explain these findings? Tversky and Kahneman argue that instead of calculating probability and using what Tversky and Kahneman call “extensional reasoning,” the participants in this study engaged in “intuitive reasoning.” Kahneman describes intuitive reasoning as being “fast, automatic, [and] effortless.”¹⁸ He goes on to say “a defining property of intuitive thoughts is that they come to mind spontaneously.”¹⁹ In other words, Kahneman’s ‘intuitive thoughts’ are not produced by any kind of conscious inference.²⁰

¹⁶ Ibid, p. 298.

¹⁷ Ibid.

¹⁸ Daniel Kahneman, “Maps of Bounded Rationality: A Perspective on Intuitive Judgment and Choice,” in Tore Frangsmyr (ed.) *Nobel Prizes 2002* (Almqvist & Wiksell International, 2003), pp. 449-489.

¹⁹ Ibid, p., 452.

²⁰ I mention this because if it turned out that psychologists were using the term ‘intuition’ differently than I was, I would be guilty of an equivocation. However, I think that psychologists appear to use the term roughly as I do. In general, they seem to agree that intuitions are not made by conscious inference. For example, Haidt, when explaining what intuitions are, claims that the process that generates intuitions is automatic and inaccessible (Haidt, “The Emotional Dog and Its Rational Tail,” p. 818). This entails that intuitions are not made by conscious inference. Haidt and Kahneman are hardly idiosyncratic here. Gerd Gigerenzer also holds that “the difference between moral intuition and moral deliberation is that the reasons underlying moral intuitions are

According to Tversky and Kahneman, intuitive reasoning, rather than using computationally expensive calculations, proceeds by the use of quick and dirty heuristics. The heuristic at work in the Linda case is the representativeness heuristic in which one judges the probability of an event “by the degree to which that event is representative of an appropriate model.”²¹ In Linda’s case T+F is more representative of Linda than T, even though probability theory shows us that T+F cannot be more probable than T.

Heuristics have the virtue of being computationally cheap. They do not require our brains to engage in any sophisticated mathematics. However, because heuristics use simple judgments that only approximate more computationally complex judgments, they also often err. If Tversky and Kahneman are correct, and many of our intuitive judgments are made by heuristics, this gives us good reason to doubt the reliability of our intuitive judgments.

Probability is far from the only domain where our intuitions lead us astray. For an example of where our intuitions about logic go wrong, consider the Wason selection task.²² Participants are given four cards, each of which has a letter on one side and a number on the other. For example, they may be given the following four cards:

typically unconscious (*Gut Feelings: The Intelligence of the Unconscious* (New York: Penguin, 2007), p. 192)”

²¹ Ibid, p. 295. This helps make clear why it is important to stipulate that intuitions are not derived from *conscious* inferences. Judgments made by the representative heuristic involve inference at some level. At the very least, they involve inferences like ‘T+F is more representative of Linda than T.’ However, these inferences are not typically accessible to the person making the judgment. Indeed, Kahneman argues that these judgments come to mind spontaneously (Kahneman, “Maps of Bounded Rationality,” p. 452.)

²² First described in P.C. Wason, “Reasoning,” in B.M. Foss, *New Horizons in Psychology* (Harmondsworth, UK: Penguin, 1966).

T S 5 8

They are then asked which cards they would have to turn over to determine the truth of the material conditional ‘if there is a T on one side of the card, then there is a five on the other.’ The correct answer is ‘T’ and ‘8’ but “typically, fewer than 10% of participants give this answer, however, and the most common responses are T alone or T and 5.”²³

The psychologist Jonathan Evans argues that the reason for this is a simple matching bias, “reasoners may simply be choosing the cards that are named in the conditional.”²⁴

Matching bias and the representative heuristic are two out of a sea of cases where psychologists have found that intuitive reasoning gets the wrong answer.²⁵ Rather than undertake a lengthy review of this enormous literature let me confine myself to one more example. This example is particularly interesting because the error in our intuitions was not discovered by psychologists but by philosophers.

Recently, Stacey Swain, Joshua Alexander, and Jonathan Weinberg put Keith Lehrer’s truetemp case to a group of undergraduates at Indiana University.²⁶ The case, a familiar one from the epistemology literature, is reprinted below:

²³ Jonathan Evans, “Matching Bias in Conditional Reasoning: Do We Understand it After 25 Years?” *Thinking and Reasoning* 4 (1998): 45-82.

²⁴ *Ibid*, p. 51.

²⁵ A useful starting point for this large literature is Rudiger Pohl (ed.) *Cognitive Illusions* (Sussex, NJ: Psychology Press, 2004). See also Daniel Kahneman, Paul Slovic, and Amos Tversky (eds.) *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge, UK: Cambridge University Press, 1982); Daniel Kahneman and Amos Tversky (eds.) *Choices, Values, and Frames* (Cambridge, UK: Cambridge University Press, 2000); and Thomas Gilovich, Dale Griffin, Daniel Kahneman (eds.) *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge, UK: Cambridge University Press, 2002).

²⁶ Stacey Swain, Joshua Alexander, and Jonathan M. Weinberg, “The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp,” *Philosophy and Phenomenological Research* 56 (2008): 138-155.

Suppose a person, whom we shall name Mr. Truetemp, undergoes brain surgery by an experimental surgeon who invents a small device which is both a very accurate thermometer and a computational device capable of generating thoughts. The device, call it a tempucomp, is implanted in Truetemp's head so that the very tip of the device, no larger than the head of a pin, sits unnoticed on his scalp and acts as a sensor to transmit information about the temperature to the computational system of his brain. This device, in turn, sends a message to his brain causing him to think of the temperature recorded by the external sensor. Assume that the tempucomp is very reliable, and so his thoughts are correct temperature thoughts. All told, this is a reliable belief-forming process. Now imagine, finally, that he has no idea that the tempucomp has been inserted in his brain, is only slightly puzzled about why he thinks so obsessively about the temperature, but never checks a thermometer to determine whether his thoughts about the temperature are correct. He accepts them unreflectively, another effect of the tempucomp. Thus, he thinks and accepts that the temperature is 104 degrees. It is. Does he know that it is?²⁷

Swain and colleagues presented truetemp along with several other cases. They found that students who were presented the truetemp case after a clear case of non-knowledge are more likely to attribute knowledge to Mr. Truetemp than students who were presented the truetemp case after a clear case of knowledge. In other words, the order in which cases were presented has an impact on how people judged the case in question. Swain and colleagues wrote, "The fact that people's intuitions about particular thought-experiments vary based on what other things they have been thinking about recently is troubling. Philosophers who rely on thought-experiments should be especially concerned about findings that indicate that, at least in some cases, subjects' intuitions are easily influenced."²⁸

Given that our intuitions have these defects in non-moral contexts, it would be somewhat surprising if, when we focused on moral goodness or badness or rightness or

²⁷ Lehrer, *Theory of Knowledge*, pp. 163-164.

²⁸ Swain et. al, "The Instability of Philosophical Intuitions," pp. 144-145.

wrongness, our intuitions suddenly became very reliable. However, because surprise in psychology is possible, I now turn to an investigation of moral intuitions.

2.5 The Reliability of Moral Intuitions

Let us begin, once again, with Kahneman and Tversky.²⁹ Consider the following scenario.

Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences are as follows:

If Program A is adopted, 200 people will be saved.

If Program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved.³⁰

Kahneman and Tversky found that 72% of participants chose program A over program B.

Another set of participants were given the same scenario but instead of program A and B they were asked to choose between program C and D.

If Program C is adopted, 400 people will die.

If Program D is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die.³¹

Given these options, 78% of participants chose program D over program C. Note however, that C is equivalent to A and D is equivalent to B.

This result is explained by Kahneman and Tversky's prospect theory according to which people are generally risk averse when it comes to potential gains but risk seeking

²⁹ Daniel Kahneman and Amos Tversky, "Choices, Values, and Frames," *American Psychologist* 39 (1984): 341-350.

³⁰ Ibid, p. 343.

³¹ Ibid.

when it comes to avoiding potential losses.³² What counts a gain or a loss is always relative to what the agent perceives as the status quo and that can be manipulated using simple wording. Program A and B implicitly assume a status quo according to which 600 people are going to be killed. Program A represents a certain gain over the status quo whereas Program B is a risky gain over the status quo. Since A and B represent potential gains from the status quo, participants are generally risk-averse. Program C and D implicitly assume a status quo according to which no one has been killed yet. Both C and D represent losses from the status quo and thus participants are risk seeking.³³

There are two lessons that can be learned from our look at the Asian disease case. First, the familiar biases that distort our intuitive judgment in non-moral cases are still present when we turn our attention to moral cases. Second, when the same case is described using different words, peoples' intuitions about the case change.

³² For a more formal statement see Daniel Kahneman and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* 47 (1979): 263-91.

³³ In her paper "Philosophical Intuitions and Psychological Theory," *Ethics* 108 (1998): 367-385, Tamara Horowitz argues that some widely held moral intuitions about various cases can be explained by prospect theory. Cass Sunstein agrees:

In short, I believe that some philosophical analysis, based on exotic moral dilemmas, is inadvertently and even comically replicating the early work of Kahneman and Tversky by uncovering situations in which intuitions, normally quite sensible, turn out to misfire. The irony is that where Kahneman and Tversky meant to devise cases that would demonstrate misfiring, some philosophers develop exotic cases with the thought that the intuitions are likely to be reliable and should form the building blocks for sound moral judgments ("Moral Heuristics," *Behavioral and Brain Sciences* 28 (2005): 531-573).

For an alternative view see F.M. Kamm, "Moral Intuitions, Cognitive Psychology, and the Harming-Versus-Not-Aiding Distinction," *Ethics* 108 (1998): 463-488.

The latter result was also found in a study by Lewis Petrinovich and Patricia O'Neill.³⁴ Petrinovich and O'Neill gave over 400 undergraduates at my alma mater, the University of California, Riverside, 21 different versions of a familiar case from the moral philosophy literature. Below is the standard version of the case.³⁵

A trolley is hurtling down the tracks. There are five innocent people on the track ahead of the trolley, and they will be killed if the trolley continues going straight ahead. There is a spur of track leading off to the side. There is one innocent person on that spur of the track. The brakes of the trolley have failed and there is a switch that can be activated to cause the trolley to go to the side track. You can throw the switch, saving the five innocent people, which will result in the death of one innocent person on the side track. What would you do?³⁶

Participants were sometimes given two options that were phrased in Kill wording:

- 1) Throw the switch, which will result in the death of one innocent person on the side track.
- 2) Do nothing, which will result in the death of the five innocent people.³⁷

Other times, participants were given two options phrased in Save wording:

- 1) Throw the switch, which will result in five innocent people on the main track being saved
- 2) Do nothing, which will result in the one innocent person being saved.³⁸

In each case, participants were asked whether they agreed or disagreed that one should throw the switch. Petrinovich and O'Neill found that,

³⁴ Lewis Petrinovich and Patricia O'Neill, "Influence of Wording and Framing Effects on Moral Intuitions," *Ethology and Sociobiology* 17 (1996): 145-171.

³⁵ This case was invented by Philippa Foot, "The Problem of Abortion and the Doctrine of Double Effect," *Oxford Review* 5 (1967): 5-15.

³⁶ Petrinovich and O'Neill, "Influence of Wording and Framing Effects on Moral Intuitions," p. 149.

³⁷ Ibid.

³⁸ Ibid.

A tendency was found for a number of individuals to shift from disagree to agree when the answer was in the Save compared to the Kill wording, and this tendency was evident for 39 of the 40 questions. In summary, the Save wording resulted in a greater likelihood that people would agree and the level of agreement was stronger. The manner in which the questions were worded, therefore, had a considerable systematic influence on the decision of most of the individuals in this sample.³⁹

The fact that our moral intuitions often judge an action permissible when a case is described one way and impermissible when the same case is described differently is strong evidence against the reliability of our intuitions. However, it is far from the only evidence. I now want to examine a study in which participants tended to have different judgments about cases that are similar but not identical. The problem is that the difference between the cases is clearly not morally relevant.

An example of this is a recent study by Eric Uhlmann, David Pizarro and colleagues.⁴⁰

Participants received one of two scenarios involving an individual who has to decide whether or not to throw a large man in the path of a trolley (described as large enough that he would stop the progress of the trolley) in order to prevent the trolley from killing 100 innocent individuals trapped in a bus. Half of the participants received a version of the scenario in which the agent could choose to sacrifice an individual named “Tyrone Payton” to save 100 members of the New York Philharmonic, and the other half received a version in which the agent could choose to sacrifice “Chip Ellsworth III” to save 100 members of the Harlem Jazz Orchestra.⁴¹

Uhlmann and colleagues found that participants were significantly less likely to sacrifice Tyrone than Chip. Uhlmann and colleagues suspect that this is due to

³⁹ Ibid, p. 153.

⁴⁰ Eric Luis Uhlmann, David A. Pizarro, David Tannenbaum, and Peter H. Ditto, “The Motivated Use of Moral Principles,” (m.s.).

⁴¹ Ibid, p. 10

overcompensating for racist attitudes. Further analysis showed that this tendency manifests almost exclusively in participants who identified as politically liberal.⁴²

In any case, it does not seem likely that the names (or ethnicities) of the people involved is a morally relevant consideration in this case. Indeed, in a separate study Uhlmann and colleagues asked participants “If Chip’s (or Tyrone’s) race was different than what you imagined, do you think this fact would change the way you responded to the questions asked above (Yes or No)?”⁴³ Ninety-two percent of subjects said no,⁴⁴ yet this feature causes peoples intuitions about the case to shift significantly.

Let us now examine another type of distortion in our intuitions. Recall that in the last section we found that subjects’ intuitions about the truetemp case vary depending on what cases are presented before it. Unfortunately, the same effect can be found in the moral case.

Petrinovich and O’Neill gave participants a set of forms designed to test for ordering effects.⁴⁵ Form 2 included three trolley cases. The first case was the standard version of the trolley case, reprinted above, which Petrinovich and O’Neill call “pull switch.” The second case on the form was a trolley case called “push button” in which there is a trolley speeding toward five people and this trolley can only be stopped by pushing a button which would cause a ramp to emerge from underneath the trolley, thereby directing the trolley onto a higher a set of tracks where it would run over one

⁴² Ibid, p. 12

⁴³ Ibid, p. 13.

⁴⁴ Ibid, p. 15.

⁴⁵ Petrinovich & O’Neill, “Influence of Wording and Framing Effects on Moral Intuitions,” pp. 153-158.

person. The third trolley case on form 2 was a case Petrinovich & O'Neill called "push person" in which the only way to stop a trolley is to push a large person off of a footbridge over the tracks.⁴⁶ Form 2 was given to 30 persons. An additional 29 participants were given form 2R which contained the same cases in reverse order. In each case participants were asked to choose between action (i.e. pulling the switch, pushing the button or person) or inaction. Petrinovich and O'Neill found that "people more strongly approved of action when [pull switch] appeared first in the sequence than when it appeared last."⁴⁷ Furthermore, people approved of the pushing the button in the push button dilemma more strongly when it followed pull switch rather than when it followed push person.⁴⁸ Ordering effects have also been found in studies by Jonathan Haidt and Jonathan Baron⁴⁹ and Uhlmann and colleagues.⁵⁰

In summary, we have seen that people's moral intuitions can vary when the same case is described in different words, when morally irrelevant changes are made in cases, and when cases are presented in a different order. We have also seen that some of the biases that are prevalent in non-moral intuitions do not go away when we contemplate morality.

⁴⁶ "Push person" is usually called "fat man" or "footbridge" in the moral philosophy literature. It was invented by Judith Jarvis Thomson, "Killing, Letting Die, and the Trolley Problem," *The Monist* 59 (1976): 204-17.

⁴⁷ Petrinovich & O'Neill, "Influence of Wording and Framing Effects on Moral Intuitions," p. 157.

⁴⁸ Ibid.

⁴⁹ "Social Roles and the Moral Judgment of Acts and Omissions," *European Journal of Social Psychology* 26 (1996): 201-218.

⁵⁰ Uhlmann et. al. "The Motivated Use of Moral Principles," pp. 19-20.

What this shows us is that our moral intuitions systematically go astray. Indeed the fact that these intuitions are subject to wording and ordering effects shows that these judgments are inconsistent. (In the Asian flu case it is inconsistent to believe that A is preferable to B and that D is preferable to C when A just is C and D just is B).

These considerations suggest two related arguments that conclude that the process by which we form our intuitions is somewhat unreliable.⁵¹ The most direct argument begins with the premise that if a judgment forming process is subject to wording and ordering effects, then that judgment forming process is unreliable. Since, as we have seen, our intuitive judgments about morality are subject to these sorts of distortions, our intuition forming process is somewhat unreliable.

A more indirect argument might appeal to consistency. If a judgment forming process delivers results that are inconsistent, then that process is somewhat unreliable. Since the process at which we arrive at our moral intuitions does produce inconsistent results, that process is somewhat unreliable. This point favors taking a skeptical view of the reliability of moral intuitions.

2.6 Some Possible Responses on Behalf of Moral Intuitions

In this section I consider three possible responses to the argument of the last section. I want to point out at the outset that philosophers who are inclined to these responses often do not realize how much they implicitly concede to the skeptical view of intuitions defended in the last section. Indeed, all of these responses have a similar structure; they all argue that certain types of (or certain people's) intuitions are left

⁵¹ A judgment is reliable, I stipulate, just in case it tends to produce the correct answer. So understood, reliability is a matter of degree. I speak of our intuition forming process as being "somewhat unreliable" because it seems to me that these arguments are inconclusive about the degree to which these processes are unreliable.

untouched by the empirical considerations in the last section. Even if such responses work they leave many of our intuitions open to the empirically motivated attacks of the last section. It is also worth stating at the outset that all of these responses make substantive claims that can be tested empirically, though not all of them have. However, even in the absence of empirical testing, these substantive empirical claims should not just be assumed to be true. With these introductory remarks out of the way, let me now take up the perspective of one such response.

The Reflective Response: In section 2.2 we noted Rawls' claim that 'intuitive' need not mean 'impulsive' or 'instinctive.' An intuitive judgment can be made on the basis of thorough reflection on the facts, so long as no conscious inference involved. Perhaps the subjects of the experiments discussed above failed to adequately reflect on the cases at hand. If they had, perhaps they would have noticed that in the Asian flu case program A and program C were equivalent or perhaps they would have been less susceptible to ordering effects. The research cited above does not show that intuitions that are the product of thorough reflection are unreliable.

Reply: The first thing to notice is that this response only attempts to defend a subset of our intuitions, those that have survived a certain amount of reflection. The second thing to notice is that the claim that reflection would improve our intuitions (e.g. make us less subject to biases) is a substantive empirical claim. Indeed, it is one that has been investigated by psychologists.

According to Robyn LeBoeuf and Eldar Shafir,⁵² the results are somewhat mixed. A study by Miller and Fagley found that framing effects were reduced on the Asian flu problem when subjects were asked to provide a rationale, but their study was also unable to replicate some already well-known framing effects among the control group that was not asked to provide a rationale.⁵³ Another study showed that framing effects were reduced when participants were asked to justify their choices on the Asian flu problem, but only after forcing participants to spend 50 minutes contemplating their choice.⁵⁴

Other studies considered participants with a high degree of what psychologists call Need for Cognition (NC), a construct that measures “differences among individuals in their tendency to engage in and enjoy thinking.”⁵⁵ Some studies showed that individuals with high NC were able to make more consistent judgments.⁵⁶ But LeBeouf and Shafir have argued persuasively that these studies do not show that high NC participants are not subject to framing effects; they only show that high NC participants

⁵² Robyn LeBoeuf and Eldar Shafir, “Deep Thoughts and Shallow Frames: On the Susceptibility to Framing Effects,” *Journal of Behavioral Decision Making* 16 (2003): 77-92.

⁵³ P.M. Miller and N.S. Fagley, “The effects of framing, problem variations, and providing rationale on choices,” *Personality and Social Psychology Bulletin* 17 (1991): 517-22.

⁵⁴ W. Sieck & J.F. Yates, “Exposition Effects on Decision Making: Choice and Confidence in Choice,” *Organizational Behavior and Human Decision Processes* 70 (1997): 207-19.

⁵⁵ J.T. Cacioppo and R.E. Petty, “The Need for Cognition,” *Journal of Personality and Social Psychology* 42 (1982): 116-131. Quoted in LeBoeuf and Shafir, “Deep Thoughts and Shallow Frames,” p. 79.

⁵⁶ S.M. Smith & I.P. Levin, “Need for cognition and Choice Framing Effects,” *Journal of Behavioral Decision Making* 9 (1996): 283-90 and LeBeouf and Shafir, “Deep Thoughts and Shallow Frames.”

are able to correct for consistency. LeBeouf and Shafir found that when given the program A and B version of the Asian flu problem, high NC participants typically chose A, as prospect theory predicts, but then when given the program C and D version of the problem after they had been given the A and B version, high NC participants chose C, contra the predictions of prospect theory.

So far, so good. However, when high NC participants were given the program C and D version of the Asian flu problem first they chose D, as prospect theory predicts. When they were then given the program A and B versions of the Asian flu problem, they chose B, contra prospect theory

What is going on here? It appears that when high NC participants look at the first question they have the intuition that prospect theory predicts. Thus, when they look at the program A and B version of the Asian flu problem first they choose A and when they look at the program C and D version of the Asian flu problem first they choose D. However, when they look at a second problem they are able to override their intuition and correct for consistency. Thus, when looking at the program C and D version after the A and B version most participants chose A.

My explanation for what is going on here is that participants with high NC (i.e. those prone to reflection) have the same intuitions that other participants have. However, they are able to override inconsistent intuitions. This explains why they choose A if given the A and B version of the Asian flu problem first but tend to choose D if given the C and D version of the problem. It also explains why they act contra to prospect theory's predictions if given the other version of the problem after the first version.

If further confirmation is needed, a further study gave high NC participants the A and B version of the Asian flu problem and then, two weeks later, gave high NC participants the C and D version of the problem.⁵⁷ In these circumstances, high NC participants gave exactly the answers that prospect theory predicts.

It appears that people prone to reflection are still subject to the same biases on the Asian flu problem (though they do correct for consistency). Of course, more research could be done. Specifically, future research might inquire on the degree to which high NC participants are subject to ordering effects on trolley problems or the degree to which their judgments are altered in Chip/Tyrone cases. However, initial research gives us some ground for pessimism and in the absence of further psychological research it cannot just be assumed that high NC participants would do better on trolley cases or Chip/Tyrone cases.

The Elitist Response: Those whose moral intuitions were surveyed in the section above were mainly undergraduates. However, their intuitions are not the ones that we consider in philosophy. We consider the intuitions of trained philosophers. These are intelligent people with advanced degrees from respected institutions who have spent years pondering the deepest questions about us and our universe. The research cited above does not show that the intuitions of trained philosophers are unreliable.

⁵⁷ I.P. Levin, G.J. Gaeth, J. Shreiber, and M. Lauriola, "A New Look at Framing Effects: Distribution of Effect Sizes, Individual Differences, and Independence of Types of Effects," *Organizational Behavior and Human Decision Processes* 88 (2002): 411-29.

Reply: First, I want to distinguish between the elitist response and another response that is similar but not identical to the elitist response. One could concede that philosophers' intuitions are no better off than those of lay people but insist that philosophers have unique skills that allow them to correct their intuitions. This, I do not deny. However, the elitist response does not concede that philosophers' raw intuitions are no better than ordinary people. The elitist contends that some feature of philosophical training gives philosophers better raw intuitions.

As before, the elitist at least implicitly concedes that most peoples' intuitions have been shown by the above research to be unreliable. The elitist also makes a substantive empirical claim, a claim that philosophers would not be subject to the same distortions that the participants in the studies above are subject to. This claim cannot just be assumed to be true. It must be argued for.

Unfortunately, to my knowledge, the relevant empirical claim has not yet been tested, though I hear that studies are underway. I do think that there is some ground for doubt about this empirical claim. First, as mentioned above, even graduate students with advance training in statistics are not immune from the representative bias in the Linda case. Tversky and Kahneman found that in the standard version of the Linda case, statistical sophistication had little effect.⁵⁸ A meta-analysis conducted by Anton Kuhberger found that, in regards to reference point framing, the sort of framing that we find in the Asian flu case, "although students and experts may differ in a variety of ways,

⁵⁸ Tversky & Kahneman, "Extensional Versus Intuitive Reasoning," p. 298.

with respect to framing experts are not immune to the reference effect... or at least, experts are also influenced by framing but to a lesser degree than students.”⁵⁹

The fact that experts in other domains are still subject to the same distortions as undergraduates creates a presumption that moral philosophers will also be subject to these distortions. However, no research that I have reviewed has looked specifically at the intuitions of moral philosophers. Thus, I concede that further research may overturn this presumption.

Let me turn autobiographical for a moment and reflect, somewhat unscientifically, on how my training in moral philosophy has impacted my judgment of the Asian flu case. I can see that programs A and C are equivalent as are programs D and B. However, even with this realization, I still feel intuitively pulled toward A, when I read the A and B version of the case, and pulled toward D when I read the C and D version of the case.

Perhaps, I am being idiosyncratic here but my reaction is strikingly similar to that of Stephen Jay Gould. Discussing the Linda case, he writes, “I know that [T+F] is least probable, yet a little homunculus in my head continues to jump up and down, shouting at me – ‘but she can’t just be a bank teller; read the description.’”⁶⁰ If Gould and I are representative, then perhaps both a graduate education and awareness of the cognitive bias in question does not improve our raw intuitions.

⁵⁹ Anton Kuhberger, “The Influence of Framing on Risky Decisions,” *Organizational Behavior and Human Decision Processes* 75 (1998): 23-55.

⁶⁰ Stephen Jay Gould, *Bully for Brontosaurus: Reflections on Natural History* (New York: Norton, 1991), p. 469.

The Principled Response: All of the research cited above deals with our intuitions about particular cases. However, as noted in section 2.2, intuitions need not be about particular cases, they can be about abstract principles. The research cited above does not show that our intuitions about abstract principles are unreliable.

Reply: This response, like the other two we have discussed, implicitly concedes that many intuitions, our intuitions about particular cases, are still vulnerable to the objections discussed in the last section. However, it holds that our intuitions about abstract principles are not vulnerable to these distortions.⁶¹ In holding the later claims, it makes a substantive empirical assumption, that such principles would not be subject to the type of distortions discussed in the previous section.

Unfortunately, there has been little empirical work on our intuitive judgments about principles. I am aware of one study that found that American students were more likely to endorse the claim that “From time to time the tree of liberty must be watered with the blood of tyrants” when they were told that it was written by Thomas Jefferson

⁶¹ This appears to be Peter Singer’s view. Singer rejects appeal to case intuitions as a way of choosing between normative theories on the ground that such judgments may be unreliable while endorsing a view, which he attributes to Sidgwick, that one should start from “self-evident moral axioms (“Sidgwick and Reflective Equilibrium, p. 47)”. He goes on to state, “While judgments of self-evidence may be regarded as appeals to intuition, they are intuitions about fundamental principles – like the principle already mentioned, that the good of one individual is of no more significance than the like good of any other – and not moral intuitions about what ought to be done in specific situations (“A Response” in Dale Jamieson (ed.), *Singer and his Critics* (Oxford: Blackwell, 1999), p. 314).” As I hope the main text makes clear, I am somewhat perplexed as to why Singer is so confident about our intuitions about principles while being so dismissive about our case intuitions.

than when they were told that it was written by Vladimir Lenin.⁶² However, I wonder if some of the same results from the above section could be replicated using abstract principles.

For instance, consider the kill/save effect that Peterinovich and O'Neill observed. They found that participants were more likely to agree with certain options in trolley cases when 'save' rather than 'kill' wording was used. Could we replicate the experiment using abstract principles? Consider the following:

A1) It is permissible to cause a few persons to die if doing so will prevent a larger number of deaths

A2) It is permissible to save a large number of persons even if doing so will result in a few deaths

I suspect that we would find that A1 is more widely endorsed than A2.

I also suspect that the Tyrone/Chip effect can be replicated using abstract principles:

B1) It is permissible to sacrifice a few black persons to save a larger number of white persons

B2) It is permissible to sacrifice a few white persons to save a larger number of black persons.

I suspect that political liberals may be more likely to endorse B2 than B1.

I also suspect that our judgment about abstract principles may be subject to ordering effects. Consider the following principles

C1) It is morally impermissible for a wealthy person not to donate to charity.

C2) It is morally impermissible to steal a paperclip from one's employer.

⁶² S.E. Asch, "The Doctrine of Suggestion, Prestige, and Imitation in Psychology," *Psychological Review* 55 (1948): 250-76.

C3) It is morally impermissible to torture a baby.

I suspect that if C1 is presented after trivial moral violations like C2 people will be more likely to agree than if C1 is presented after serious moral violations like C3. I suspect that this is because after hearing about serious moral violations like C3, C1 will not sound so bad but after hearing about trivial moral violations like C2, C1 will seem morally wrong.

Of course, in the absence of empirical research these are mere speculations.

However, my goal here is not to definitively answer the principled response but to create a *prima facie case* against it.

It may be objected that the principles I have considered above are not abstract enough. This would be an extreme version of the principled reply that implicitly concedes not only that our intuitions about cases are unreliable; our intuitions about concrete principles are also unreliable.

In response to such a reply I claim that once principles get so abstract, many of us only have intuitions about them when we think about how they might apply in actual cases.⁶³ From my own perspective, I cannot generate an intuition about whether or not one ought to always do that action which produces the most overall good unless I think about how such a principle might apply in actual cases. But these sorts of intuitions are

⁶³ In his paper "Thinking About Cases," *Social Philosophy & Policy* 18 (2001): 44-63, Shelly Kagan makes the interesting suggestion that our judgments about principles are just judgments about *types* of cases rather than *particular* cases. This claim, if it is true, could explain why it is difficult to generate intuitions about very abstract principles whose implications for cases are unclear. This claim would also imply that it would be a mistake to hold that case judgments are unreliable while also claiming that principle judgments are reliable. Since, according to Kagan's suggestion, principle judgments are, in some sense, a species of case judgment.

dependent on our case intuitions. When it comes to extremely abstract principles many of us find it difficult to generate intuitions that are independent of our case intuitions.

2.7 The Good News about Intuitions

So far the results of our inquiry have been primarily negative. In the last section, we saw that these negative results cannot be easily dismissed. Nonetheless, our conclusion that moral intuitions sometimes deliver incorrect results leaves open the question of how often they do so. It is compatible with what we have argued so far that our intuitions give us the right result 90% of the time.

The problem, however, is that it seems that we have no way of knowing when we are experiencing an incorrect intuition or when we are experiencing a correct intuition. Incorrect intuitions do not seem to be accompanied by a distinct phenomenology. Thus, it seems that any of our intuitions could be in error and we have no way of knowing when this is so.

There is, however, some good news about intuitions. In a recent study, Jennifer Wright tested various cases for ordering effects.⁶⁴ Wright hypothesized that when subjects were very confident about their intuitive judgments these judgments would not be subject to ordering effects. She first replicated Swain et. al's result that participants judgments about the TrueTemp case varies depending on the cases that precede it. She then found, however, that other epistemic cases were not subject to any ordering effect at all. Here are two cases where no ordering effect was found:

Testimony: Karen is a distinguished professor of chemistry. This morning, she read an article in a leading scientific journal that mixing two common floor disinfectants, Cleano Plus and Washaway, will create a poisonous gas that is

⁶⁴ Jennifer Wright, "On Intuitional Stability: the Clear, the Strong, and the Paradigmatic," *Cognition* 115 (2010): 491-503.

deadly to humans. In fact, the article is correct: mixing the two products does create a poisonous gas. At noon, Karen sees a janitor mixing Cleano Plus and Washaway and yells to him, “Get away! Mixing those two products creates a poisonous gas!”⁶⁵

Coin Flip: Dave likes to play a game with flipping a coin. He sometimes gets a “special feeling” that the next flip will come out heads. When he gets this “special feeling”, he is right about half the time, and wrong about half the time. Just before the next flip, Dave gets that “special feeling”, and the feeling leads him to believe that the coin will land heads. He flips the coin, and it does land heads.⁶⁶

Participants overwhelmingly thought that Karen had knowledge in Testimony but that Dave did not in Coin Flip. Unlike in Truetemp, these results did not change no matter which cases preceded them. Perhaps even more interestingly, participants reported that they were much more confident in their judgments in Testimony and Coin Flip than they were in Truetemp just as Wright’s hypothesis predicts.

Importantly for our purposes, Wright found similar results in moral cases. She gave the following three vignettes to participants.

Break-Promise: Fred promises his girlfriend that he will meet her for lunch at 12 pm on Wednesday at their favorite café. Wednesday at 11:45 am, on his way to the café, Fred runs into his grandfather, who is out for a stroll. They exchange hellos, and then suddenly Fred’s grandfather clutches his chest and falls to the ground unconscious. An ambulance arrives minutes later to take Fred’s grandfather to the hospital. Fred accompanies his grandfather to the hospital, even though he knows that doing so means that he will be breaking his promise to have lunch with his girlfriend.

Sell-iPod: Laura and Suzy are roommates. Laura asks Suzy if she has seen her new iPod, which she had worked an extra job over the summer to be able to afford. Suzy did recently see it under a pile of papers on the bookshelf. But Suzy lies to Laura, telling her that she hasn’t seen it. She thinks that if Laura doesn’t find it on her own in a day or two, she can take it down to the pawnshop and get \$100 for it, which would provide her with beer money for the week.

⁶⁵ Ibid, p. 502.

⁶⁶ Ibid, p. 501.

Hide-Bombers: Martha hides her Jewish neighbors in her basement during the Nazi occupation of France. A German soldier comes to her door one afternoon and asks her if she knows where her neighbors have gone. Martha knows that her neighbors are wanted by the Germans or bombing a German-only schoolyard and killing several children, injuring others. Martha lies to the soldier, telling them no, she hasn't seen them recently, but she believes that they fled the country.⁶⁷

No matter what order the cases were presented in, participants overwhelmingly agreed Fred's action in Break-Promise was not wrong and the Laura's action in Sell-iPod was. When participants read Hide-Bombers after Sell-iPod, 55% of them judged Martha's action to be wrong. However, when participants read Hide-Bombers after Break-Promise, only 33% of participants judged Martha's action wrong. In other words, the intuitive reactions of participants to Break-Promise were subject to ordering effects. As Wright's hypothesis predicts, subjects were more confident about their judgments about Break-Promise and Sell-iPod than their judgments about Hide-Bomber.

There are two optimistic conclusions we can draw from Wright's research. First, some intuitions do not seem to be subject to ordering effects. Second, the intuitions that we are most confident in are least likely to be subject to ordering effects. Can we infer from this that the intuitions that we are most confident about are correct? Not quite. While the fact that a judgment forming process is inconsistent entails it is somewhat unreliable, consistency does not entail reliability. It is possible that our intuitions could consistently track falsehood. Nonetheless, Wright's research does show that the effects of a certain kind of bias are limited and do not affect our most deeply held intuitions.

⁶⁷ Ibid, p. 503.

2.8 Conclusion

In this chapter I have offered an account of what moral intuitions are and the role they play in ordinary moral judgment. I have also defended a kind of moderate position about their reliability.

Thinking about the conclusions we have reached in this chapter naturally brings to mind further questions. What does this mean for the role of intuitions in moral philosophy? What does this imply about the epistemic status of moral intuitions? The answers to these questions are by no means obvious and will require us to think deeply about difficult issues in epistemology and ethics. In the following chapters, I turn to these issues with the hope of answering these questions

Chapter 3: Reflective Equilibrium in Ethics

3.1 Introducing Reflective Equilibrium

At the conclusion of Chapter 1, I said that this dissertation would defend a coherentist theory of justification that would avoid the pitfalls of the foundationalist views discussed in that chapter. Roughly, that coherentist theory is that a moral belief is justified if and only if it would be the output of a version of the procedure that John Rawls has called “the method of reflective equilibrium.”¹

In this chapter, I discuss what the method of reflective equilibrium is and how best to formulate it. This is the project of sections 3.2-3.5. Once we have seen what reflective equilibrium is, we will be able to see the considerations that motivate it. Drawing out these considerations is the project of sections 3.7-3.9.

3.2 Narrow Reflective Equilibrium

The term ‘reflective equilibrium’ is introduced by Rawls in his most well known work, *A Theory of Justice*.² There, Rawls notes that “there are several interpretations of reflective equilibrium (p. 49).” In this section I focus on the interpretation that is commonly called “narrow reflective equilibrium.” In the next section, I consider an alternative interpretation, which is commonly called “wide reflective equilibrium.”³

¹ I follow a tradition established in the literature on this topic by using the phrase ‘reflective equilibrium’ to refer to the state that one’s beliefs would be in had she completed the reflective equilibrium process and using the phrase ‘method of reflective equilibrium’ to refer to the process itself.

² All page references in the main text refer to this work.

³ Rawls introduces this terminology in “The Independence of Moral Theory,” *Proceedings and Addresses of the American Philosophical Association* 48 (1974): 5-22. Reprinted in *Collected Papers* ed. Samuel Freeman, pp. 286-302.

The method of narrow reflective equilibrium begins with our intuitions as described in section 2.2.⁴ We then filter out those intuitions that were formed in circumstances in which we are likely to err. For example, we filter out those judgments “made with hesitation, or in which we have little confidence. Similarly, those given when we are upset or frightened or when we stand to gain one way or another can be left aside (p. 47).” The judgments we are left with after this filtering process are our *considered moral judgments*.

We then look for the set of principles that systematizes our considered moral judgments. Such a set of principles should not only yield the considered moral judgments that we already have, they should also be able to derive new judgments in cases where we are uncertain (p. 20).

If we are lucky we will be able to find a set of principles that systematizes all of our considered moral judgments. If so, then we have completed the procedure. Most likely, however, we will find conflicts between our considered moral judgments and the principles that we have adopted in order to systematize them. Rawls says that, “in this case, we have a choice (p. 20).” We can either modify our considered moral judgments to accommodate our principles or we can modify our principles to accommodate our considered moral judgments. We then work back and forth until we arrive at a consistent set of judgments and principles. Once we have achieved a consistent set of judgments and principles, our moral beliefs are in narrow reflective equilibrium.

⁴ Some textual evidence that the judgments that become our considered moral judgments are what I am calling ‘intuitions’ can be found in “Outline of a Decision Procedure for Ethics,” p. 6. There, Rawls says “It is required that the judgment be intuitive with respect to ethical principles. That is, that it should not be determined by a conscious application of principles so far as this may be evidenced by introspection.”

Before moving on, I want to correct a common misinterpretation of the method of reflective equilibrium. It is sometimes thought that our considered moral judgments are simply our judgments about particular cases. While it is true that our judgments about particular cases are among our considered moral judgments, they are not the only judgments that count as considered moral judgments. As Rawls himself puts it, “people have considered judgments at all levels of generality, from those about particular situations and institutions up through broad standards and first principles to formal and abstract conditions on moral conceptions.”⁵ Indeed, Rawls’ only concrete examples of considered moral judgments, our judgments that “racial discrimination and religious intolerance are unjust (p. 19),” are not case judgments.⁶ As Rawls’ comments illustrate, considered moral judgments can be fairly wide or narrow in scope.

3.3 Wide Reflective Equilibrium

Earlier, I noted that Rawls claims that there are several different interpretations of reflective equilibrium. While narrow reflective equilibrium is the version of reflective equilibrium that receives the most discussion in *A Theory of Justice*, Rawls makes it clear that he actually prefers a different interpretation, wide reflective equilibrium (p. 49).⁷ On this interpretation one considers not only those principles that best systematize her

⁵ “The Independence of Moral Theory,” p. 289.

⁶ Whether or not these judgments count as considered moral judgments depends on whether the agent arrived at them without conscious inference. Some agents arrive at judgments like these on the basis of inductive inference from cases or as a consequence of certain theoretical commitments. If so, then these judgments are not considered moral judgments for these agents. However, it seems to me that at least some people simply assent to these judgments intuitively. For such agents these judgments will count as considered moral judgments. In any case, this does not affect the general point that Rawls considered certain non-case judgments to be considered moral judgments.

⁷ Also see “The Independence of Moral Theory,” p. 289.

considered moral judgments, but also alternative sets of principles “together with all the relevant philosophical arguments for them (p. 49).”

Like narrow reflective equilibrium, the method of wide reflective equilibrium includes a set of considered moral judgments and a set of moral principles. However, wide reflective equilibrium introduces a third set, the premises of philosophical arguments relevant to determining the strengths and weaknesses of various moral principles. Following Norman Daniels, let us call this third set, a set of background theories.⁸ Background theories may be either moral or non-moral and may include claims

about the right-making properties of actions, policies, and institutions; about the conflict between consequentialist and deontological views; about partiality and impartiality and the moral point of view; about motivations, moral development, strains of moral commitment and the limits of ethics; about the nature of persons; about the role or function of ethics in our lives; about the implications of game theory, decision theory, and accounts of rationality for morality; about human psychology, sociology, and political and economic behavior; about the ways we should reply to moral skepticism and moral disagreement; and about moral justification itself.⁹

The method of wide reflective equilibrium thus begins with three sets: a set of considered moral judgments, a set of moral principles, and a set of background theories.

It is worth noting that these sets can be overlapping. For example, I may have the

⁸ Norman Daniels, “Wide Reflective Equilibrium and Theory Acceptance in Ethics,” *Journal of Philosophy* 76 (1979): 256-282. Reprinted in *Justice and Justification* (Cambridge, UK: Cambridge University Press, 1996). pp. 21-46.

⁹ Daniels, *Justice and Justification*, p. 6.

considered moral judgment that causing another person to have excruciating pain for my amusement is *prima facie* wrong and this may be among the principles that best systematizes my other considered moral judgments. If so, this belief will be a member of both the set of considered moral judgments and the set of moral principles. I may also have the considered moral judgment that rational beings have a special moral status. This judgment may also be among the background theories that I am committed to. If so, this judgment will be included in both the set of considered moral judgments and the set of background theories.

Most of us will find conflicts between our considered moral judgments and moral principles as well as between our considered moral judgments and background theories (recall that we have considered moral judgments at all level of generality). For instance, we may have the considered moral judgment that it is wrong to torture unloved animals and infants while holding a Kantian background theory according to which only rational beings are worthy of moral consideration. In this case, we have a conflict between a background theory and a considered moral judgment.

Having found such conflicts, we then begin the process of revising for consistency. We may choose to revise a moral principle to accommodate a considered moral judgment. Of course, such a revision may cause our moral principles to conflict with our background theories. In such a case we may choose to modify the considered moral judgment, leaving the moral principles and background theories as they were, or we may modify our background theory. We may also modify a background theory because it directly conflicts with a considered moral judgment and vice-versa. When we have made all three sets consistent we have achieved wide reflective equilibrium.

3.4 From Intuitions to Considered Moral Judgments

There is one question about the way in which both the methods of wide and narrow reflective equilibrium have been formulated that I have not answered. Recall that only some of our intuitions, those that count as considered moral judgments, play a role in the method of reflective equilibrium. Why are some intuitions filtered out at the beginning of the process?

Rawls claims that we filter out the intuitions that were formed in situations in which we are likely to err (p. 47). However, Michael DePaul has argued convincingly that this suggestion has two problems.¹⁰ The first problem is that this suggestion would require us to formulate an account of when our moral intuitions are likely err and such an account is bound to be controversial.

There is a second and perhaps more serious problem. Suppose we had such an account. We can now say that at the beginning of the wide reflective equilibrium process, the agent must rule out all those judgments made in circumstances where she is likely to err according to this account. But now suppose that the agent herself has no reason to accept this account. Maybe this account claims that she is likely to err when she is frightened but has reason to believe that one can gain genuine moral insight when frightened and has no evidence to the contrary. The second problem with Rawls' formulation is that it would force the agent to make changes to her set of moral beliefs that she herself may have no reason to make. Indeed, her evidence may dictate that she not make these changes.

¹⁰ Michael DePaul, *Balance and Refinement: Beyond Coherence Methods of Moral Inquiry* (London: Routledge, 1993), p. 17

As DePaul recognizes, solving this problem is simple. Instead of ruling out those judgments made when the agent is likely to err, we rule out those judgments made when the agent herself believes that she is likely to err.¹¹ If we interpret the initial filtering out process this way, we are not guilty of imposing constraints on agents that they themselves have no reason to accept. Of course, the judgments ruled out by this filtering process are those that presumably would not survive long in reflective equilibrium anyway. Dropping them early on is partly a matter of convenience. It is also to ensure that those judgments that are not trustworthy by the agent's own standards do not play a role in constraining the moral principles that she adopts.

This modification of Rawls' description of the method of reflective equilibrium has another important consequence, it allows us to expand the scope of which judgments are allowed to play a role in reflective equilibrium. Martha Nussbaum and others have worried that Rawls' initial filtering procedure would rule out "information given us by our fear, or grief, or love" from the reflective equilibrium process.¹² However, the modified filtering procedure described in this section will not necessarily have this result. As long as the agent herself believes that she is not likely to err when she is afraid, grieving or in love, then information gained from these emotions will play a role in the method of reflective equilibrium.¹³

¹¹ Ibid, p. 18.

¹² Martha Nussbaum, *Love's Knowledge* (Oxford: Oxford University Press, 1990), p. 175.

¹³ This point is also made by Henry Richardson in *Practical Reasoning about Final Ends* (Cambridge, UK: Cambridge University Press, 1994), pp. 183-190.

3.5 Against The Independence Constraint

Daniels has argued that the method of reflective equilibrium, as formulated above, is not complete. It requires one further modification.¹⁴ I do not accept this modification.¹⁵

Daniels worries that the set of background theories in the method of wide reflective equilibrium may just be reformulations of the set of considered moral judgments. This claim gains some plausibility if we recognize that many of the background theories we accept are generalizations of our considered moral judgments. For example, consider the well-known case discussed in the introduction in which a surgeon can save five lives by killing a healthy patient and harvesting his organs. Most of us have the considered moral judgment that this action would be wrong. One might infer, from cases like this, the background theory that there are deontic constraints on what agents may do to maximize the good. Note however, this background theory seems to be a sort of generalization of our case judgments.

Suppose that all our background theories were like this. If this were so, then wide reflective equilibrium would be no different than narrow reflective equilibrium. For those of us who think that wide reflective equilibrium is preferable to narrow reflective equilibrium (a thesis I will argue for in section 3.8), this appears problematic.

Daniels attempts to solve this problem by introducing two sets of considered moral judgments: one to constrain the set of moral principles and another to constrain the set of background theories. Furthermore, he insists that these two sets “be to some

¹⁴ Daniels, “Wide Reflective Equilibrium and Theory Acceptance in Ethics,” pp. 22-23.

¹⁵ DePaul also rejects the independence constraint (*Balance and Refinement*, pp. 21-22). His argument is similar but not identical to mine.

significant degree disjoint.”¹⁶ This constraint is what he refers to as the independence constraint. The *independence constraint* is a requirement that the set of background theories and the set of considered moral judgments are constrained by two different sets of considered moral judgments and that these two sets must be to some significant degree disjoint. If these sets are somewhat disjoint then the set of background theories is constrained in a way that the set of considered moral principles is not and vice-versa. This, Daniels thinks, will ensure that the contents of these sets is significantly different and thus that the agent’s beliefs in wide reflective equilibrium are distinct from those in narrow reflective equilibrium.

I do not think that there is a problem. It would be an odd sort of agent who accepted only background theories that were merely reformulations of their considered moral judgments. Surely most people have some beliefs (e.g. beliefs about religion, folk psychology, politics and its relation to morality, perhaps even proto-metaethical beliefs)¹⁷ that would count as background theories and are not just mere re-hashes of their considered moral judgments. Among those who are less philosophically inclined, the set of background theories will be sparse enough that the beliefs they would hold in wide reflective equilibrium may not be that different from the beliefs that they would hold in narrow reflective equilibrium. But this does not seem to me to be especially problematic.

¹⁶ Daniels, “Wide Reflective Equilibrium and Theory Acceptance in Ethics,” p. 23.

¹⁷ The view that ordinary people have metaethical beliefs is defended by Mackie in *Ethics*, pp. 30-35, Joyce in *The Myth of Morality*, (especially chapters 2 and 6), Terence Cuneo in “Saying What we Mean: An Argument against Expressivism” in Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics* vol. 1 (Oxford: Oxford University Press, 2006), pp. 60-65, and David Copp, *Morality in a Natural World* (Cambridge, UK: Cambridge University Press, 2007), pp. 6-7.

In any case, whether or not there is a genuine problem here, there is a significant cost to introducing the independence constraint. Let us call the set of all of an agent's considered moral judgments $\{CMJ\}$. In the method of wide reflective equilibrium, as described in section 3.3, $\{CMJ\}$ constrains both the set of moral principles and the set of background theories. Call this option 1. Let me spell it out for clarity's sake.

Option 1: $\{CMJ\}$ constrains both the set of moral principles and the set of background theories (and these sets constrain $\{CMJ\}$).

What Daniels is proposing is that there be two sets, $\{CMJ_p\}$, which would constrain the set of moral principles, and $\{CMJ_b\}$, which would constrain the set of background theories. Call this Option 2:

Option 2: $\{CMJ_p\}$ constrains the set of moral principles and $\{CMJ_b\}$ constrains the set of background theories, where $\{CMJ_p\}$ and $\{CMJ_b\}$ are somewhat disjoint.

I think that option 1 is actually preferable to option 2. Recall that the independence constraint requires that we break $\{CMJ\}$ up into two sets, $\{CMJ_p\}$ and $\{CMJ_b\}$, which must be somewhat disjoint. This entails that either $\{CMJ_p\}$ or $\{CMJ_b\}$ will lack some judgments that there were present in $\{CMJ\}$. Suppose that $\{CMJ_p\}$ lacks some judgments that were present in $\{CMJ\}$. This means that in option 2, the set of moral principles will be constrained by fewer beliefs in option 2 than it was in option 1. Thus, when one achieves wide reflective equilibrium they may still have at least some considered moral judgments (those judgments in $\{CMJ\}$ that were not included in $\{CMJ_p\}$) that are inconsistent with the set of moral principles.

If $\{CMJ_p\}$ does not have fewer beliefs than $\{CMJ\}$, then by the requirement $\{CMJ_p\}$ and $\{CMJ_b\}$ be somewhat disjoint, $\{CMJ_b\}$ must have fewer beliefs than

{CMJ}. If this is the case, then we run into the same problem with {CMJ_b} and background theories that we did with {CMJ_p} and moral principles. The beliefs that are part of {CMJ} but not {CMJ_b} then may remain inconsistent with the set of background theories even when one has reached wide reflective equilibrium. I take this to be a decisive objection against the independence constraint.

3.6 A Coherentist Theory of Justification

Let me now re-describe the method of wide reflective equilibrium with the modification suggested in section 3.4. One begins the process with three sets:

- 1) A set of considered moral judgments, which are those intuitions that are not made in circumstances in which the agent believes she is likely to err.
- 2) A set of moral principles
- 3) A set of background theories, which are those beliefs that might serve as premises in an argument for or against a set of moral principles.

We then work back and forth to make these three sets consistent.

As stated at the outset of this chapter, I aim to defend the view that a moral belief is justified if and only if it is the output of this method. It should be clear that such a theory of justification is coherentist. An agent's justified moral beliefs are justified not because they are derived from self-evident truths but because of the way in which they cohere with the agent's other beliefs, namely because they are in wide reflective equilibrium with the other things an agent believes. As Rawls puts it, "A conception of justice cannot be deduced from self-evident principles or conditions on principles; instead, its justification is a matter of the mutual support of many considerations, of everything fitting together into one coherent view (p. 21)."

A natural question for the reader to ask at this point is ‘I see what the method of reflective equilibrium is, but why should I believe that it is the appropriate method for moral inquiry?’ The rest of this chapter is devoted to answering this question. In what follows, I give three considerations that, when taken together suggest that reflective equilibrium is the correct method for moral inquiry. These three considerations are not to be thought of as premises in a deductive argument with the conclusion that one ought to accept the method of reflective equilibrium. While I wish there were such an argument, I doubt that there is one to be found. However, I do not want to be too modest. I will argue that these three considerations, when taken together, do strongly suggest, even if they do not entail, that the method of reflective equilibrium is the appropriate method for moral inquiry. Moreover, the plausibility of these considerations is independent of the method of reflective equilibrium, or so I shall argue.

3.7 The First Consideration - In Ethics, Intuitions are Unavoidable

The first of these considerations is that, in ethics, intuitions are unavoidable. To see that this consideration is true ask yourself the following question: if ethical inquiry did not require intuitions, how would it get started?

Let us survey the alternatives. Many types of inquiry start with sensory perception. Could ethical inquiry start this way? In answering this question, we have to keep in mind that ethical inquiry deals essentially with ‘ought’s. We also must keep in mind Hume’s law that one cannot derive an ‘ought’ from an ‘is.’ I think it is fairly obvious that, while our senses give us much valuable information about the world, they do not directly tell us what we ought to do. Sensory perception deals exclusively in ‘is’

claims, and since we cannot derive an 'ought' from an 'is,' ethical inquiry cannot get started with sensory perception.

In section 2.2, I said that a judgment made by conscious inference is not an intuition. Perhaps, ethical inquiry could start with an inferred 'ought.' Hume's law shows us that such an inference must be from a further 'ought' claim. This means that every validly inferred 'ought' presupposes a further 'ought.' A regress threatens, and since we are concerned with how ethical inquiry could ever get started, it will not do to say there is an infinite regress of 'ought's or that this inference travels in a circle. For ethical inquiry to even get started, we need an un-derived 'ought' claim.

It is difficult to see where such an un-derived 'ought' could come from if not from intuition. We have already seen that sensory perception does not give us 'ought's. In the absence of any other alternatives, it seems that ethical inquiry must start with an intuition. Those who wish to avoid this conclusion must show that there is something other than intuition that could be the source of an un-derived 'ought' claim. If there is no other plausible starting place for ethical inquiry, then at least some appeal to intuition is unavoidable in ethics.

I recognize that talk of 'the starting point of ethical inquiry' is somewhat artificial. However, the main point here does not rely on the assumption that there is some moment in time that we can characterize as the starting point of ethical inquiry. It is simply that without some appeal to intuition, we would not be able to think about ethics at all.

This last point entails something rather stark that is worth emphasizing. Ethical inquiry cannot get off the ground without intuitions. This means that to repudiate all appeals to intuition is to give up moral inquiry entirely. It is not as if we face a choice

between doing ethics with intuitions or without any appeal to intuitions whatsoever. There latter option is not an option.

3.8 The Second Consideration - Intuitions are Sometimes Unreliable

The second consideration is that intuitions are sometimes unreliable. This was argued for back in chapter 2. I will not rehearse those arguments here. Instead I want to consider a possible objection. It may be thought that when taken together, the first two considerations leave us condemned to a kind of skepticism about ethical inquiry. If we have to rely on intuitions in ethics and intuitions are unreliable, then perhaps ethical inquiry is a non-starter to begin with.

This line of thought ignores an important possibility, that we can subject our intuitions to scrutiny. The claim that intuitions are unavoidable in ethics does not entail that we are forced to blindly accept every intuition we have. Instead, we can subject our intuitions to a kind of scrutiny and accept those intuitions that have passed scrutiny. This way we neither have to avoid intuitions (which the first consideration shows us is impossible) nor do we have blindly trust judgments that are sometimes unreliable

Of course, this response to the skeptic only works if we can provide some means of scrutinizing our intuitions. How could this be done? Our next consideration shows us that there is only one possible answer.

3.9 The Quinean Consideration

Throughout his philosophical writings, Quine frequently invokes the image of Neurath's boat. Quine argues that science is like "a boat which, if we are to rebuild it, we must rebuild plank by plank while staying afloat in it."¹⁸ Perhaps the best way to

¹⁸ W.V.O. Quine, *Word and Object* (Cambridge, MA: MIT Press, 1960), p. 3.

understand Quine's metaphor is to see what it is intended to rule out: the Cartesian project of abandoning everything we believe and starting again from scratch. In metaphorical terms, if we hope to modify or add on to the raft we must stand on another part of the raft. In literal terms, if we hope to modify our conceptual scheme, we must do so by appealing to other parts of our conceptual scheme.

This, I realize, is an ambitious claim. As a descriptive matter, I believe that it is actually one of the claims motivating Rawls' acceptance of the method of reflective equilibrium.¹⁹ As a prescriptive matter, I believe that there is much that can be said in defense of this claim. Suppose we had abandoned our entire conceptual scheme, as Descartes asks us to in the first Meditation. Now suppose we are deciding which belief we should adopt as our first. How would we scrutinize such a belief? In other words, how would we decide whether or not we should believe a first belief? We cannot see if it contradicts other things that we believe, since we have no other beliefs. Even rejecting a first belief on logical grounds would require a further belief, namely belief in the laws of logic. Thus, even Descartes' claim that his existence is certain because denying his existence involves contradicting himself requires a further belief in the law of non-contradiction. It seems that in the absence of further belief we have no means of scrutinizing our beliefs. In order to scrutinize our beliefs, we must do so from the standpoint of other beliefs. This is the lesson I wish to draw from Neurath's boat.

Returning to our main line of thought, in the last section we saw that we could avoid being led into skepticism if we could provide some way to scrutinize our moral

¹⁹ Rawls notes that his account of the method of reflective equilibrium "has benefited from the conception of justification found in W.V. Quine, *Word and Object*, ch. 1 (*A Theory of Justice*, p. 579n33)."

intuitions. The Quinean consideration shows us that there is only one way to scrutinize our beliefs and that is by bringing our other beliefs to bear on it. This entails that the most scrutiny that we can subject a belief to is scrutiny from the perspective of all of our relevant beliefs. Thus, the only way to scrutinize our moral intuitions is to see how they fare against the beliefs that are relevant to their truth or falsity.²⁰ To subject our moral intuitions to the most possible scrutiny is just to see how they fare against *all* the beliefs relevant to their truth or falsity. Of course, this just *is* the method of wide reflective equilibrium. In the method of wide reflective equilibrium we bring the full force of all of our relevant beliefs to bear on our moral intuitions.

This reasoning also suggests why we should prefer the method of wide reflective equilibrium to the method of narrow reflective equilibrium. The method of narrow reflective equilibrium puts our intuitions under much less scrutiny than the method of wide reflective equilibrium does. Subjecting our intuitions to more scrutiny can increase the credibility of those intuitions that survive this process. Unfortunately, a full defense of this claim will have to wait until the next chapter. For now, I can only suggest that if we take the worries about intuitions raised in chapter 2 seriously, we should not ignore those background theories that are relevant to the truth or falsity of our considered moral judgments. Instead, we should see what these background theories have to say about the plausibility of our moral intuitions.

²⁰ Of course, determining which beliefs are relevant to the truth or falsity of our moral beliefs will be a challenge. I assume that this can be done, though I have no unique suggestion for how best to accomplish it.

3.10 From Methodology to Epistemology

Allow me summarize the argument so far. Ethical inquiry requires some appeal to intuition and intuitions are sometimes unreliable. This suggests that intuitions must be subjected to scrutiny. This scrutiny can only come from the perspective of one's other beliefs. These considerations, when taken together, suggest that one should scrutinize their moral intuitions from the perspective of all their beliefs that bear on their moral intuitions. To do this is just to engage in the method of wide reflective equilibrium.

Let us suppose that this argument is successful. What does it show? Does it show that the method of wide reflective equilibrium is a correct method of moral inquiry? Does it show that the method of wide reflective equilibrium is the only correct method of moral inquiry? I believe that these three considerations suggest that the method of reflective equilibrium is the only correct method of moral inquiry. If the method of wide reflective equilibrium is the process which subjects our moral intuitions to the most possible scrutiny, then an alternative method must either eschew intuitions altogether or subject them to less scrutiny than the method of wide reflective equilibrium. If an alternative method eschews intuitions altogether, then it runs afoul of the first consideration. If an alternative method uses intuitions but subjects them to less or no scrutiny, then it fails to take the second consideration seriously.²¹

²¹ Nagel's method as described in *Mortal Questions* (Cambridge, UK: Cambridge University Press, 1979) is an example of subjecting one's intuitions to less scrutiny than the method of reflective equilibrium requires. Nagel writes, "I believe one should trust... intuitions over arguments... Given a knockdown argument for an intuitively unacceptable conclusion, one should assume there is probably something wrong with the argument that one cannot detect (p. x)." F.M. Kamm's method as described in *Morality, Mortality* vol. 1, pp. 7-9, seems to involve privileging intuitions about particular cases over theoretical considerations. This marks a departure from the method of reflective equilibrium in that in the method of reflective equilibrium, intuitions about particular cases do not have any privileged position over other sorts of intuitions or theoretical considerations.

Some readers will note that this argument is a methodological argument, in that if successful, this argument shows that the method of reflective equilibrium is the correct methodology to use in ethics. However, the coherentist position I mentioned in the section 3.1 is epistemological, it is an account of when a moral belief is justified. An objector might claim that even if the argument of the last few sections is correct, it establishes only a methodological claim and not an epistemological one.

In response, I want to note that this apparent distinction between an account of methodology and an account of when a belief is justified collapses if we understand ‘justification’ as I characterized it back section 1.1. There, I stated that to say that a proposition is justified is to say that it is rational or appropriate to believe it. The claim that method M is the correct method for inquiry I, seems to me to just be the claim that we should use M to decide what to believe about I. In other words, a claim that method M is the correct method for inquiry I entails that it is rational or appropriate to believe the results of M. Thus, if we understand ‘justification’ as I have suggested, claiming that the method of wide reflective equilibrium is the correct method for moral inquiry entails that the moral beliefs that are justified are just those that we would hold had we attained wide reflective equilibrium.

3.11 Conclusion

More needs to be said about the theory of justification explicated and defended here. Some central notions employed in theory require further clarification and perhaps modification. Furthermore, philosophers have proposed a range of objections against the sort of theory defended here and something must be said in reply to these objections. The

project of clarifying, modifying, and defending the theory discussed in this chapter will occupy the rest of this dissertation.

Chapter 4: Reflective Equilibrium, Coherence, and Moral Intuitions

4.1 The Worry about Intuitions

The juxtaposition of chapters 2 and 3 may leave the reader somewhat perplexed. In chapter 2, I argued that we should take certain skeptical worries about intuitions seriously. In chapter 3, I endorsed a theory of justification that made use of our intuitions. If there are serious grounds for worry about the reliability of intuitions, how can we accept a theory of justification in which they play a prominent role?

If the reader has this worry, she is not alone. The most enduring and troubling objection to the method of reflective equilibrium is that, in the form of considered moral judgments, intuitions play a prominent role in constraining the moral principles we adopt.

¹ However, we have no reason to think such intuitions are credible and some reason to think that they are not credible.

In this chapter, I begin by discussing two suggestions that have been made by other authors for solving this problem. In my view, neither of these suggestions is adequate as stated. I then argue for a modification to the method of reflective equilibrium as described in chapter 3. Finally, I argue that with this modification, we can answer the worry about intuitions.

4.2 The First Proposed Solution – The Analogy to Non-moral Observation Reports

The first attempt to solve the worry about intuitions is due to Norman Daniels.²

Daniels makes an analogy between moral and scientific inquiry. Scientific inquiry begins

¹ For references, see footnote 3 in the introduction.

² “Wide Reflective Equilibrium and Theory Acceptance in Ethics,” pp. 30-33.

with our observations. Eventually from a kind of non-moral reflective equilibrium we were able to derive an account of how these observations work that vindicates their role in scientific inquiry. This account is presumably the familiar one in which the external world impinges on our sensory organs which then send signals to our brains causing us to have our sensory experience of the external world. The important thing to note is that we started with just brute observations. However, by testing those observations against various principles that systematized and explained them, we were able to develop an account of the reliability of these brute observations.

Daniels speculates that something similar could be done in the moral case. By trying to bring our considered moral judgments into reflective equilibrium, we might be able to develop an account that vindicates our considered moral judgments, just as an attempt to bring our observation reports into reflective equilibrium allowed us to develop an account about the reliability of our observations. It is important to note that Daniels does not give us any hint of what such an explanation would look like; he merely suggests that there could be one.

I think that there is a genuine possibility such an account may emerge from reflective equilibrium. In chapter 2, we saw that our considered moral judgments sometimes lead us astray. However, our observations sometimes lead us astray (e.g. when it is dark, when we are looking at far away objects, or in familiar optical illusions). This gives us no reason to assume that we cannot give an account of how our moral intuitions are reliable in many cases, just as our observations are reliable in many cases.

However, another possibility looms. This is the possibility that the best explanation of our considered moral judgments is a debunking explanation, one that

shows that they are unreliable. Peter Singer suggests that our moral intuitions “are likely to derive from discarded religious systems, from warped views of sex and bodily functions, or from customs necessary for the survival of the group in social and economic circumstances that now lie in the distant past,”³ Joshua Greene has argued that many of our moral intuitions are based on emotional responses to morally irrelevant features of situations,⁴ and Peter Unger has also offered debunking explanations for many of our central moral intuitions.⁵

Ideally we would compare Daniels’ vindicating explanation to the various debunking explanations that have been offered and see which one best explains the data. However, we cannot do that since Daniels does not give us such an explanation, he only suggests that there could be one. Thus, at this point, I do not see any reason for preferring a Daniels-style vindicating explanation of considered moral judgments to a Singer/Greene/Unger-style debunking explanation of considered moral judgments. At best, Daniels’ reply is inconclusive. However, I think there is a more serious problem with Daniels’ reply. It makes an unnecessary concession to the critic of reflective equilibrium. Daniels’ reply concedes that the plausibility of the method of reflective

³ Singer, “Sidgwick and Reflective Equilibrium,” p. 47. The adequacy of Singer’s debunking explanation is discussed by Kagan in “Judgments about Cases”

⁴ Joshua Greene, R. Sommerville, L.E. Nystrom, J.M. Darley, J.D. Cohen, “An fMRI Investigation of Emotional Engagement in Moral Judgment.” *Science* 293 (2001): 2105-8; Joshua Greene and Jonathan Haidt, “How (and Where) Does Moral Judgment Work?” *Trends in Cognitive Sciences* 6 (2002): 517-33; Joshua Greene, “The Secret Joke of Kant’s Soul,” in Walter Sinnott-Armstrong (ed.), *Moral Psychology Volume 3: The Neuroscience of Morality* (Cambridge, MA: MIT Press, 2008). For an argument against Greene’s debunking explanation see Selim Berker, “The Normative Insignificance of Neuroscience,” *Philosophy and Public Affairs* 37 (2009): 293-329.

⁵ *Living High and Letting Die* (Oxford: Oxford University Press, 1996). For an argument against Unger’s debunking explanation see F.M. Kamm, *Intricate Ethics* (Oxford: Oxford University Press, 2007), Ch. 6.

equilibrium hinges on providing a general story about the reliability of our considered moral judgments. For reasons that will emerge in the next section, I emphatically deny that the plausibility of the method of reflective equilibrium hinges on any such story. Let me now consider a reply that does not make this concession.

4.3 The Second Proposed Solution – The Corrective Reply

Thus far, I have been talking of the reliability of considered moral judgments *simpliciter*. However, talking this way runs the risk of conflating two very distinct sets of judgments. One is our *initial moral judgments* – the set of considered moral judgments that we begin the reflective equilibrium process with. These judgments serve as inputs in the method of reflective equilibrium. Another distinct set of judgments is our *considered moral judgments in reflective equilibrium* – the set of considered moral judgments that one would have after completing the reflective equilibrium process. These judgments are outputs of the method of reflective equilibrium.

The corrective reply attempts to answer the worry about intuitions by noting how much the set of initial moral judgments and considered moral judgments in reflective equilibrium can come apart. Recall from section 3.3 the long list of things that may serve as background theories. We may have to alter our initial moral judgments quite radically to accommodate such background theories. Our initial moral judgments are put under further pressure by the requirement that they must be systematized by a set of principles. It is thus likely that our considered moral judgments will undergo substantial revision in reflective equilibrium. Indeed, under the twin pressures of systematization and accommodating background theories we may have to revise every single initial moral judgment.

The corrective reply holds that because the method of reflective equilibrium allows for such radical revision, there is no need to defend the reliability of initial moral judgments. Such judgments may be unreliable. However, the procedure itself functions as a corrective for these potentially unreliable judgments.⁶

Given the things that I said in sections 3.8-3.10, it should come as no surprise that I think that the corrective reply is on the right track. However, I do not think that it is adequate as stated. To show why this is, allow me introduce a new character to our discussion: the dogmatist.

The dogmatist holds some considered moral judgment C very strongly. In fact, she holds C so strongly that she would give up any of her other beliefs to preserve C. Furthermore, the dogmatist believes C independently of any reasons for C. She may give you some reasons why she believes C, but if those reasons turn out to support \sim C, then she will continue to believe C.

I take it that the dogmatist is a paradigm case of an epistemically irresponsible agent. Whether or not she is justified in believing C will of course depend on C's evidential and inferential relationships to other things she believes. That much is relatively uncontroversial. However, it would be a serious defect of a theory of justification if that theory had the implication that whatever C was, and whatever relationships it had to dogmatist's evidence, it would be justified. I worry, however, that the method of reflective equilibrium has that implication. Allow me to explain.

⁶ Very explicit statements of the corrective reply are hard to find, but many defenders of the method of reflective equilibrium come very close to stating it. See Rawls, "The Independence of Moral Theory," p. 289; Daniels, "Wide Reflective Equilibrium and Theory Acceptance in Ethics," pp. 26-28; and David Brink, *Moral Realism and The Foundations of Ethics*, (Cambridge, UK: Cambridge University Press, 1989), pp. 141-142.

C will clearly be one of the dogmatist's initial moral judgments. Let us just suppose, for the sake of argument, that C conflicts with other things that the dogmatist believes. When the dogmatist is going through the reflective equilibrium process she will uncover these conflicts and then, to paraphrase Rawls, she will have a "choice."⁷ She can either revise C or revise those beliefs of hers that conflict with C. Of course, we have already stipulated that the dogmatist would give up any of her beliefs to preserve C. Thus, the dogmatist will give up her beliefs that conflict with C. No matter what kind of evidence the dogmatist has or lacks for C, C will be the output of the method of wide reflective equilibrium. Even if the dogmatist has strong evidence for \sim C and no evidence for C, C will be the output of the wide reflective equilibrium process. This result is unacceptable.

We are left with this result because the way that the method of reflective equilibrium is commonly formulated it requires that agents revise their beliefs when faced with a conflict. However, the standard formulations say very little about how the agent is to go about deciding what to revise. What has been said makes it sound as if the decision about what to revise is entirely in the hands of the agent.

I suspect that it is this feature of the method of reflective equilibrium that is especially worrisome to its critics. Moreover, I think the critics have a legitimate cause to worry. The solution, I will argue, is to take the decision about what to revise out of the hands of the agent. In the rest of this chapter, I search for a way to do this.

⁷ *A Theory of Justice*, p. 18.

4.4 Credence

As I noted in the last section, defenders of the method of reflective equilibrium have said very little about how agents should go about revising their beliefs when they find conflicts. Ironically, the only philosopher who has described this process in any detail is a critic of reflective equilibrium, R.B. Brandt.⁸ On Brandt's account one assigns all of their beliefs a value based on how strongly one believes them. Call this value a *credence level*. When one finds conflicts between beliefs one revises so as to maximize the overall credence level of her set of beliefs.⁹

While this account does make the process of belief revision in reflective equilibrium more determinate, it does not solve the problem of the dogmatist nor the more general worry about intuitions (nor does Brandt think it does, his view is that worries of this sort should lead us to reject the method of reflective equilibrium). The dogmatist will presumably assign a very high value to C such that it will not make sense to revise C no matter what her overall set of evidence suggests. We are still left with the conclusion that the method of reflective equilibrium allows an agent to be justified in holding a moral belief independently of its evidential and inferential relations to other beliefs that she holds.

⁸ *A Theory of the Good and the Right*, pp. 17-20. Brandt's account draws on an account of non-moral justification described by Israel Scheffler, "Justification and Commitment," *Journal of Philosophy* 51 (1954): 180-190.

⁹ To see that there really is an issue about how one ought to revise her beliefs when she comes across conflicts in reflective equilibrium, note that Brandt's account and Rawls's account may come apart, since, one may choose to revise in ways that do not maximize the overall credence level of her set of beliefs. If an agent were to revise in such a way, would she be mistaken?

4.5 Coherence

Some readers will note that one notion that plays a central role in this dissertation has not been explicated, the notion of coherence. Eventually, I will argue that a better understanding of coherence can solve the specific worry about the dogmatist and the general worry about the credibility of our moral intuitions. Unfortunately, I cannot provide a complete account of coherence.¹⁰ However, I do want to offer a sketch of some of the things that are involved.

It is, I think, fairly obvious that coherence involves brute logical consistency. But it should also be clear that coherence involves more than logical consistency. Consider two sets of propositions, set A and set B. Set A consists of two propositions:

A1) There is life on Mars

A2) Given everything that we know, it is very unlikely that there is life on Mars.

Set B also contains two propositions

B1) There is no life on Mars

B2) Given everything that we know, it is very unlikely that there is life on Mars.

Set A and Set B are both logically consistent. However, set B is clearly more coherent than set A. The reason is that set A has what Laurence Bonjour calls “probabilistic inconsistency.”¹¹ While there is no logical contradiction between A1 and A2, the conjunction of A1 and A2 is very unlikely. Let us generalize and say that other things

¹⁰ In fact, Bonjour has suggested that a complete account of coherence is not possible “within the scope of any work of a manageable length (*The Structure of Empirical Knowledge*, p. 93).” He does not take this to be a problem for the coherentist since virtually all non-skeptical theories of justification (even foundationalist ones) make use of the concept of coherence.

¹¹ *The Structure of Empirical Knowledge*, p. 95.

being equal, the more probabilistically consistent a set of propositions is, the more coherent it is.

Probabilistic and logical consistency are not the only elements of coherence. This point is best brought out using another example.¹² Consider two sets of propositions set C and set D. Set C consists of three propositions

- C1) This chair is brown.
- C2) Electrons are negatively charged
- C3) Today is Thursday

Set D also consists of three propositions

- D1) All ravens are black
- D2) This bird is a raven
- D3) This bird is black

Set C and D are both logically and probabilistically consistent. However, set D is obviously more coherent than set C. The reason is that set D involves what Geoffrey Sayre-McCord calls “connectedness.”¹³

The connectedness of a set of beliefs is determined by the degree to which the beliefs in the set possess evidential or inferential relations to each other. As with ‘coherence,’ I will not be able to provide a complete account of ‘connectedness.’ I am not sure that such a thing is possible and if it were it would require a higher degree of formalization than is necessary or desirable here. However, I do want to say a bit more about what

¹² This example is taken from Ibid, p. 96.

¹³ “Coherentist Epistemology and Moral Theory,” in Sinnott-Armstrong and Timmons (eds.) *Moral Knowledge*, pp. 137-189.

connectedness consists in. Importing some suggestions from Bonjour (once again),¹⁴ I claim that the connectedness of a set of beliefs is determined by the number and strength of the inferential connections between beliefs. The greater the degree to which such connections are present, the more connected the set of beliefs is. When the set of beliefs contains unconnected subsets of beliefs, the connectedness of the set of beliefs is diminished.

In addition to connectedness and consistency, coherence also involves what Sayre-McCord calls ‘comprehensiveness.’¹⁵ I will, once again, illustrate ‘comprehensiveness’ using an example. Compare two sets of beliefs C and E. Set C contains three beliefs:

- C1) This chair is brown.
- C2) Electrons are negatively charged
- C3) Today is Thursday

Set E contains one belief

- E1) This chair is brown

Now suppose a believer encounters two further claims

- ~C2) Electrons are not negatively charged
- ~C3) Today is not Thursday

Note that a believer with the beliefs in set C will be able to pass judgment on these two claims but a believer whose belief system contained only the belief in set E would have nothing to say about ~C2) and ~C3).

¹⁴ *The Structure of Empirical Knowledge*, pp. 96-100.

¹⁵ “Coherentist Epistemology and Moral Theory,” pp. 166-167

Generalizing from this example, I think we can explain comprehensiveness in the following way. The comprehensiveness of a set of beliefs is determined by the degree to which the beliefs contained in that set allow a believer to pass judgment on a range of further propositions. A set of beliefs that enables a believer to pass judgments on a more varied set of further propositions is more comprehensive than a set of beliefs that is able to pass judgment on a less varied range of propositions.

Of course, this only raises further questions. What is it for a set of beliefs to pass judgment on a further belief? While I cannot fully answer these questions here, let me make a few remarks. The simplest way a set of beliefs can pass judgment on a further belief is if that set of beliefs entails that the further belief is true or false. However, there are other ways a set of beliefs can pass judgment on a further belief. A set of beliefs may imply that a further belief is likely to be true or likely to be false or a set of beliefs may imply that there is strong evidence for or against a further belief. A set of beliefs fails to pass judgment on a further belief when it simply has nothing to say about it, as set E has nothing to say about $\sim C2$) and $\sim C3$).

Let me now move to the question of what it is for a set of beliefs to be more varied. One may say a set of beliefs is more varied if it contains a higher number of beliefs. While this is often evidence that a set of beliefs is more varied it need not entail that a set of beliefs is more varied. Set C and set D above have the same number of beliefs, however, set C is clearly more varied than set D. In fact, I think that if we added one further belief about ravens to set D, set C would still be more varied. One could say a lot more about what is involved in 'varied' but an intuitive grasp will good enough for our purposes here.

I have not provided a complete account of either ‘connectedness’ or ‘comprehensiveness.’ Let me mention that even if I had, my account of ‘coherence’ would not be complete. First, I’m leaving open that there may be more elements to coherence than consistency, comprehensiveness, and connectedness, though I will not speculate on what they might be. Second, a complete account of coherence would require a function that specifies the contribution that consistency, connectedness, and comprehensiveness make to the overall coherence of a set of beliefs. My guess is that a simple function such as

$$\text{coherence} = \text{consistency} + \text{comprehensiveness} + \text{connectedness}$$

will not work. First, one may doubt that we could quantify consistency, comprehensiveness, and connectedness. Even supposing we could, there is a further problem. Imagine two sets of beliefs with the following values:

$$\text{A: consistency}=1, \text{comprehensiveness}=0.9, \text{connectedness}=0.01$$

$$\text{B: consistency}=1, \text{comprehensiveness}=.45, \text{connectedness}=.45$$

A has a higher overall value than B, however, since A has so little connectedness, I am inclined to say that B is more coherent than A. This raises an important point: maximizing coherence may not simply be a matter of maximizing consistency, comprehensiveness, and connectedness.

This leads to a third important point. It may be that in some cases the elements of coherence (i.e. consistency, connectedness, and comprehensiveness) conflict. For example it could be that making one revision increases connectedness while decreasing comprehensiveness. Does this revision make the overall set of beliefs more or less

coherent? A complete account of coherence would have to provide an answer but I offer none here.

Let me conclude by drawing some lessons from our discussion of coherence. First, as I hope it is clear, coherence is much more than simple consistency. Second, while consistency may be a matter of kind, connectedness and comprehensiveness are clearly a matter of degree. This means that coherence is also a matter of degree. We can therefore speak of a set of beliefs as being more or less coherent.

4.6 Coherence and Reflective Equilibrium

In discussing the dogmatist, we saw that defenders of the method of reflective equilibrium have said little about the process of belief revision in reflective equilibrium. We have so far surveyed two alternatives. The first, suggested by Rawls' use of the word 'choice' in describing the method of reflective equilibrium, is that it is completely up to the agent. On this view there are simply no restrictions about how one should go about revising her beliefs in reflective equilibrium. The second alternative, suggested by Brandt in his critique of reflective equilibrium, is that an agent revises so as to maximize the overall credence level of her set of beliefs.

In this section, I shall suggest a third alternative. Let me explain this alternative by invoking a familiar metaphor. Recall from chapter 1 that coherentists typically see the structure of justification as being something like a spider's web (as opposed to a building). It is commonly thought that some beliefs are more central to the web than others. For instance, my belief in the laws of logic is close to the center of the web whereas my belief that I was wearing a black t-shirt yesterday is close to the periphery of

the web. It is also typically thought that when one discovers conflicts between two beliefs one should revise so as to preserve the more central belief.

This point can be illustrated by an example.¹⁶ Suppose I believe that last week I weighed 200 pounds. Furthermore, I believe that I maintained a low calorie diet this week and have lost weight. In other words, I believe that I now weigh less than 200 pounds. I walk over to my trusty scale, step on it, and observe that my scale reads 205 pounds. I believe that my scale is extremely reliable and so infer that I now weigh 205 pounds. I am faced with an apparent conflict. I believe that I weigh less than 200 pounds and that I weigh 205 pounds. What to do?

One option is to revise my mathematical belief that 205 is greater than 200. A second option is to revise my beliefs about logic so that there is no conflict after all. There is yet a third option; my belief that my scale read 205 pounds was formed on the basis of sensory perception. Up until now, I believed that sensory perception was reliable. However, I could revise this belief.

Most of us agree that none of these options are the right way to resolve this conflict. Why is this? The explanation is that these beliefs are all fairly central to my web of belief. I am better off revising something at the periphery, perhaps my belief that I maintained my diet last week. In general, when faced with a conflict of beliefs one should revise the beliefs to the periphery of one's web.

What determines how central a belief is to one's web? It seems to me that the beliefs central to one's web are those that make significant contributions to the comprehensiveness and connectedness of one's overall set of beliefs. To see this imagine

¹⁶ This is a variation on an example from Thomas Nagel, "The Sleep of Reason," *The New Republic* (October 12, 1998), p. 35.

we removed our beliefs about mathematics or logic from our set of beliefs. Such a set would now be radically less connected and be able to pass judgment about much fewer propositions. It also seems to me that the beliefs that are close to the periphery of one's web are those that do not make much of a contribution to the comprehensiveness and connectedness of one's overall set of beliefs. My belief that "I wore a black t-shirt yesterday" could be rejected without radically decreasing the comprehensiveness and connectedness of my overall set of beliefs.

After reflecting on these and other examples, the following seems to me to be a good explication of the notion of centrality to one's web of beliefs: how central or peripheral a belief is to one's web of beliefs is determined by the contribution that belief makes to the comprehensiveness and connectedness of one's overall set of beliefs. The greater the contribution a belief makes to the comprehensiveness and connectedness of one's overall set of beliefs the more central that belief is to one's web of belief. Conversely, the lesser the contribution a belief makes to the comprehensiveness and connectedness of one's overall set of beliefs the closer to the periphery that belief is to one's web of belief.

My suggestion for how to handle the question of how one ought to revise her moral beliefs when she finds conflicts in the method of reflective equilibrium is the following: in the moral case, like the non-moral case, when we are faced with a conflict we ought to revise that belief that is less central to one's web of moral beliefs. Put differently, my suggestion is that when an agent comes across an inconsistency in her overall set of moral beliefs (i.e. her considered moral judgments, the moral principles that systematize them, and her background theories), she makes the revision which would

maximize the coherence of her overall set of moral beliefs, where coherence includes comprehensiveness and connectedness as well as consistency. On this account of belief revision, which belief one should revise in cases of conflict is determined by the *contribution* that belief makes to the overall comprehensiveness and connectedness of one's set of beliefs. For this reason, I will refer to this account as *the contributionist account of belief revision*.

I have quite a bit to say about this account, but let me begin by attempting to forestall a misunderstanding. The contributionist account does *not* hold that one should throw away their current set of moral beliefs and exchange it for the most coherent set of moral beliefs possible. If one were to do this, one might end up with a truly bizarre set of moral beliefs. Instead, the contributionist account holds that one starts with the moral beliefs they have and then searches for inconsistencies. One revises for the sake of coherence only when she comes across inconsistencies.

This means that, on the contributionist view, justification is not *solely* a function of coherence. A radical coherentist might simply equate justification with coherence. The contributionist account is a moderate coherentism, according to which coherence plays a crucial role in justification, however coherence by itself does not determine whether a moral belief is justified or not. The view still avoids foundationalism because it denies that any moral belief is justified independently of its relations to other beliefs (a point that I shall return to in section 4.10).

We now have three alternative accounts of how belief revision is supposed to work in reflective equilibrium. The reader may ask, why should they prefer the contributionist account to Rawls' or Brandt's? Recall our 205-pound man from earlier in

this section. We agreed that it would be a mistake for him to revise his logical or mathematical views to reconcile his beliefs about his weight. My worry is that both Rawls and Brandt's accounts allow one to pull off the moral analogue of this mistake. These accounts allow agents to revise beliefs at the center of their web of moral belief in order to preserve beliefs close the periphery. This is abundantly clear with Rawls' view, which seems to allow an agent to revise any way she sees fit. Brandt's suggestion is unhelpful since nothing prevents an agent from assigning a high credence level to a belief at the periphery of her web of moral belief. A defender of Rawls' or Brandt's view may deny that there is any mistake if one revises a belief at the center of her moral web in order to preserve something at the periphery. However, it is clearly a mistake in the non-moral case. If this is so, then, absent some argument, it is difficult to see why we should treat the moral case differently.

4.7 Time Slice Justification vs. Historical Justification

We need to modify this account to answer an important objection. Suppose that an agent face a conflict between two beliefs U and J. Suppose further that the contributionist account implies that U is unjustified and the agent ought to revise U and preserve J but the agent irrationally revises J to preserve U. Now he no longer believes J and, let us assume, there is no further belief that conflicts with U. Since U now coheres (or is at least not inconsistent) with his other beliefs, the theory of justification now implies that the agent may be justified in believing U but this result is absurd. The only reason the agent is justified in believing U is that she made an irrational revision to her set of beliefs. An irrational revision should not justify one in believing anything.¹⁷

¹⁷ I owe this point to Patricia Greenspan

To solve this problem, we need to insure that the beliefs that serve as inputs to the method of reflective equilibrium are not just the agent's *current* beliefs. We need to insure that those earlier beliefs of the agent that have not already been defeated (i.e. ruled unjustified by the procedure) serve as inputs to the method.

4.8 Moral Coherence

We can now see that coherence plays a crucial role in the theory of ethical justification. In section 4.6, I explicated the notion of coherence by importing an account of coherence that has been developed in the epistemology of perceptual beliefs. This is one case where results obtained in one area of philosophy (epistemology) can help to illuminate another area of philosophy (ethics). However, importing an idea from one field to another carries certain dangers. In particular, it carries the danger that one ignores the subtleties of the field receiving the import. For this reason, I want to spend some space discussing how the notion of coherence applies in the moral case. I will argue that there are some unique features of the moral case that deserve our attention.

Let us begin with consistency. In non-moral epistemology we are typically concerned with two kinds of consistency, logical and probabilistic. However, ethics is often, if not chiefly, concerned with a third kind of consistency, what I will call *practical consistency*. Like many other notions, this is best illustrated by example. Consider the following two claims

- 1) One ought to always be kind
- 2) One ought to always be just¹⁸

¹⁸ The examples and the term 'practical consistency' are taken from Roger Crisp, "Sidgwick and the Boundaries of Intuitionism" in Stratton-Lake (ed.), *Ethical Intuitionism*, pp. 56-75.

There is no logical inconsistency in holding both 1) and 2).¹⁹ However, there is a kind of practical inconsistency, in that there are certain cases in which one cannot consistently act in accordance with both 1) and 2).

I am inclined to say that ethical coherence involves not just logical and probabilistic consistency, but also practical consistency. However, this raises a further problem. Some philosophers believe that there are genuine moral dilemmas, cases where there are conflicting moral obligations such that no matter what one does, one acts wrongly.²⁰ To take an example from Sartre, we can imagine an able-bodied French man during World War II whose mother is extremely sick.²¹ It seems that he has both a duty to care for his sick mother and a duty to fight in the French resistance. A defender of moral dilemmas might claim that both of these duties are genuine all-things-considered duties even though our able-bodied French man cannot act in accordance with both of them. Accordingly, the defender of moral dilemmas urges, to have beliefs that are completely practically consistent is wrong. There are genuine practical inconsistencies and a failure to recognize them is a genuine failure.

¹⁹ It is true that one cannot act on both 1) and 2) but this is due to contingent features of our world and thus the inconsistency between 1) and 2) cannot just be a matter of logic. One might object here and insist that as a conceptual matter all kind actions are also just. This strikes me as implausible but, in any case, one who holds this view should just substitute another example. Perhaps ‘one should always keep the promises’ and ‘one should always save other lives when they can’ will conflict in a case where I have promised to meet someone at a certain time but I run across a person trapped in a burning vehicle.

²⁰ See e.g. Ruth Barcan Marcus, “Moral Dilemmas and Consistency,” *Journal of Philosophy* 77 (1980): 121-136 and Bernard Williams, “Ethical Consistency,” *Proceedings of the Aristotelian Society* supp. 39 (1965): 103-124.

²¹ Jean-Paul Sartre, “Existentialism is a Humanism,” in Walter Kaufmann (ed.), *Existentialism from Dostoevsky to Sartre* (New York: Meridian, 1957), pp. 287-311.

There are of course, philosophers who deny that there are moral dilemmas.²² Perhaps, the easy answer is to side with them and move on. However, that would be a mistake here since whether or not we should believe in moral dilemmas is something that should be decided within our theory of justification and not something that should be presupposed by it. For this reason, let me say a bit about how the method of reflective equilibrium and the contributivist account of belief revision should deal with the issue of moral dilemmas.

It will be helpful if we have in mind an actual argument for moral dilemmas. Let us take a look at one influential argument, which I will call the argument from guilt.²³ The argument notes that whatever the agent in Sartre's case does, it will be appropriate for him to feel guilt. If he joined the French Resistance, it would be appropriate for him to feel guilt for abandoning his sick mother and if he cared for his sick mother it would be appropriate for him to feel guilt for failing to aid the French Resistance. From this case, the argument derives the following existential claim:

G1) There are some cases where no matter what the agent does it is appropriate to feel guilt

The argument goes on to claim

G2) It is only appropriate to feel guilt if one acts wrongly

Therefore,

²² See e.g. Immanuel Kant, *The Metaphysics of Morals* (1787), pp. 224-225 and David Brink, "Moral Conflict and Its Structure," *Philosophical Review* 103 (1994): 215-247.

²³ This is a version of the argument offered by both Marcus, "Moral Dilemmas and Consistency," and Williams, "Ethical Consistency." Since, we will not be evaluating the soundness of this argument, I ignore some details and complications for expositional simplicity.

G3) There are some cases where no matter what the agent does, she acts wrongly. That is, there are genuine moral dilemmas.

In the method of reflective equilibrium G1) and G2) serve as background theories. Accepting these background theories would compel us to accept a third background theory, G3).

Suppose that we accept G1) through G3) and hold that Sartre's case is a genuine moral dilemma on the ground that it is appropriate for Sartre's Frenchman to feel guilt no matter what he does. We would be committed, on grounds of logical consistency to hold both:

- 1) It would be wrong for Sartre's Frenchman to abandon his mother
- 2) It would be wrong for Sartre's Frenchman to stay with his mother.

Denying 1) or 2) would be logically inconsistent with our background theories (G1-G3). However, accepting both 1) and 2) would be practically inconsistent.

Denying the claim that practical consistency is a part of overall coherence is much too costly. Such a denial would leave us with no grounds to say that one ought to revise either their judgment that 'one ought always to be just' or their judgment that 'one ought always to be kind.' What we have is a case where practical consistency conflicts with logical consistency. Of course, we noted earlier that there can be conflicts between the elements of coherence. Just as there may be cases where comprehensiveness and connectedness conflict, there can be cases where logical consistency and practical consistency conflict. We could resolve this conflict by rejecting the relevant background theories (G1-G3). However, if these background theories are sufficiently well supported by the other things we believe, it may be more coherent overall to accept the practical

inconsistency. This is consistent with claiming that other things being equal, the more practically consistent a set of beliefs is the more coherent it is. Here, we have stumbled on a case where other things are not equal and coherence may demand that we hold on to the sufficiently well grounded background theories and tolerate the practical inconsistency.

Whether or not it will make sense to tolerate the practical inconsistency will depend on how well grounded the background theories are. Opponents of moral dilemmas hold that these background theories are not sufficiently well grounded and should be rejected. For all I have said here, they may, in the end, be correct. My aim here is not to settle the debate between friends and opponents of moral dilemmas but to explain how a particular theory of justification need not beg the question against one side or the other. It need not beg the question because coherence itself, in this case coherence with well-supported background theories, may require that we tolerate a practical inconsistency. I conclude that we can accept that practical consistency is a genuine element of coherence and a desirable feature of a moral system without pre-supposing that there are no moral dilemmas.

This may, in some sense, constitute a qualification in my overall account of the method of reflective equilibrium. Earlier I said that when one finds a conflict one must make that revision which maximizes coherence. I still hold that his claim is correct. However, we must recognize that in some cases tolerating an (practical) inconsistency may be the best way of maintaining overall coherence. In these cases, we ought to tolerate the inconsistency.

This concludes my discussion of applying consistency to ethical coherence. I take it that connectedness raises no special problems for moral beliefs. However, comprehensiveness does raise some issues. Earlier we said that the comprehensiveness of a set of beliefs is determined by its ability to pass judgment on a varied set of further beliefs. If a set of beliefs can pass judgment on a large variety of further beliefs, we say it is very comprehensive. Applying this to the moral case, we say that the comprehensiveness of a set of moral beliefs is determined by its ability to pass judgment on varied set of further moral beliefs including case-judgments, principles, and moral background theories.

This account raises some apparent difficulties. In discussing the role of comprehensiveness in ethical coherence, Walter Sinnott-Armstrong notes that, “we do not want a moral system to tell us whether to part or hair on the left or the right.”²⁴ This is surely correct, but we do want a moral system to tell us something about the case, namely that it would be morally permissible to part our hair on either side.

As this example shows us, it is important not to conflate passing judgment with telling us what to do. We do not want a moral system to tell us what to do in the hair-parting case. However, a moral system can pass judgment on a case without telling us what to do, as does the moral system that tells us that it is permissible to part our hair on either side. My point is that we can hold that comprehensiveness is a desirable element of a moral system without holding the absurd view that a moral system must tell us what to do in all cases.

²⁴ *Moral Skepticisms*, p. 226.

Holding that comprehensiveness is a desirable element of a moral system is also compatible with holding that there are moral dilemmas. Indeed holding that case is a moral dilemma is a way of passing judgment on a case.

I now want to consider a different kind of objection to comprehensiveness inspired by Bernard Gert. Gert holds that there are some moral questions with no unique correct answer.²⁵ We might call such cases ‘morally indeterminate.’ To get a handle on ‘moral indeterminacy’ it will be useful to have a little background in Gert’s moral theory.

Gert claims that an action is immoral if it violates a justified moral rule without adequate justification.²⁶ Justified moral rules, according to Gert, are those rules that all rational persons would endorse as part of a public system of conduct that applies to all rational beings.²⁷ Gert thinks, *pace* some Kantians, that there are some cases where impartial rational beings will not agree about whether or not there should be a moral rule forbidding the action.²⁸ He gives the example of abortion.²⁹ Perhaps impartial rational beings could not agree on whether or not there should be a moral rule forbidding abortion. If so, then whether or not abortion is immoral is morally indeterminate.

One might worry that holding that comprehensiveness is a desirable feature of a moral system begs the question against theories like Gert’s which hold that there is some moral indeterminacy. However, this is not so. A set of beliefs fails to pass judgment on

²⁵ *Morality: Its Nature and Justification* Revised edition (Oxford: Oxford University Press, 2006).

²⁶ *Ibid*, p. 111.

²⁷ *Ibid*, pp. 159-162.

²⁸ *Ibid*, pp. 150-152.

²⁹ *Ibid*, pp. 143-133

a further belief only if it has nothing to say about it, as we saw that set E had nothing to say about $\sim C2$ or $\sim C3$. To take a moral example, imagine the set of moral beliefs M, which only includes one moral belief:

M1) Stealing is wrong

M has nothing to say about whether or not abortion is immoral.

In contrast, Gert's view does have something to say about abortion; it says that abortion is morally indeterminate. Holding that the moral permissibility of abortion is indeterminate is not an instance of having nothing to say about abortion. Indeed, it is an instance of saying something about it. This becomes clear if we recall that Gert actually provides a means of determining whether or not a case is morally indeterminate. It is not that Gert's theory fails to pass judgment on the morality of abortion. On the contrary, Gert's theory does pass judgment on it; it holds that it is indeterminate. This shows us that claiming that comprehensiveness is a desirable feature of a moral theory does not beg the question against theories that allow for moral indeterminacy.

Gert raises the point of moral indeterminacy as part of a more general objection to the method of reflective equilibrium. Gert charges that the method of reflective equilibrium presupposes that there is a correct answer to every moral question.³⁰ Gert's charge is false. The claim that there is not a correct answer to every moral question could easily be incorporated into the method of reflective equilibrium. It would be a background theory. Suppose one accepted it as a background theory and accepted some moral principle (e.g. the principle of utility, one ought to do that action which maximizes overall utility) that gives an answer to every moral question. One would then have a

³⁰ Ibid, pp. 380-381. Allen Wood has also raised this objection. See his *Kantian Ethics*, pp. 48-49.

conflict and have to revise either the principle or the background theory. If the background theory is as well grounded as Gert claims it is, then one should reject the principle in favor of one that does not provide an answer to every moral question.

Of course, this is compatible with the utilitarian claim that Gert's background theory is not well grounded and should be rejected. My purpose is not to adjudicate the dispute between Gert and the utilitarian but only to show that Gert's claim that there is genuine moral indeterminacy can be accommodated within the method of reflective equilibrium. Therefore, Gert's claim that the method of reflective equilibrium presupposes that there is a unique answer to every moral question is false.

In this section I have discussed some complications that arise when we attempt to apply the explication of coherence discussed in section 4.5 to moral epistemology. I think that provided we make some additions and clarifications we can apply this account to moral epistemology. In particular, we can do so without presupposing that there are no moral dilemmas or that there is no moral indeterminacy. In section 4.5, I conceded that my account of coherence is incomplete. Perhaps my account of how coherence applies to moral epistemology is similarly incomplete. While I am open to that possibility, I hope that I have dissolved at least a few of the main worries.

4.9 The Dogmatist Revisited

This lengthy discussion about the nature of coherence has been somewhat of a digression. The main aim of this chapter was to solve the worry about intuitions discussed in section 4.1. Before I attempt that, however, I want to return to a more specific worry about the dogmatist discussed in section 4.3. Seeing how we can solve this worry will allow us to see how to solve the main worry about intuitions.

Let me begin by re-stating the problem of the dogmatist. The dogmatist believes some considered moral judgment C so strongly that she is willing to revise any of her beliefs to accommodate C. We saw that, as described in Chapter 3, the method of reflective equilibrium allows the dogmatist to be justified in believing C, no matter what relationships C bears to the dogmatist's evidence. This consequence is unacceptable.

In section 4.6, we saw, however, that the account of reflective equilibrium given in chapter 3 needs to be modified. In particular, it needs to be modified in the following way: When one finds a conflict in her overall set of beliefs she must make the revision that would maximize coherence in her overall set of moral beliefs.

I contend that this modification can solve the problem of the dogmatist. As before, the dogmatist begins by searching for inconsistencies in her set of beliefs. Suppose C is inconsistent with some further belief F. The dogmatist must now compare the inferential and evidential relationships that C and F have to her other beliefs as well as the contribution C and F make to the comprehensiveness of her set of beliefs. To revert to the metaphor of the spider's web, she attempts to discern whether C or F is closer to the periphery of her web of belief. If her set of beliefs would be more coherent if she revised C, then she must revise C regardless of her confidence in it.

The problem we noted earlier is that as the method of reflective equilibrium had been described in chapter 3, it allowed C to survive in reflective equilibrium no matter what evidence the dogmatist has for it. The modified reflective equilibrium avoids this result. With our modification in place, C can only survive reflective equilibrium if it is supported by the dogmatist's best evidence, that is, if it bears strong inferential and evidential relationships to a highly coherent (not merely consistent) set of beliefs.

Note that the modification in the method of reflective equilibrium that yields our proposed solution is not *ad hoc*. It is ultimately motivated by the kind of concerns for producing a highly coherent set of beliefs that motivate coherentist theories of justification generally.

4.10 The Epistemology of Moral Intuitions

I now turn, finally, to moral intuitions. I hope it is clear that on the account we have been developing, it is likely that many of our intuitive judgments will not be justified. Going through the reflective equilibrium process we will find conflicts between our considered moral judgments and the principles that best systematize them as well as our background theories. When faced with such conflict we will often be required to revise our considered moral judgments rather than the principles and background theories that conflict with them. The method of reflective equilibrium leaves open the possibility that none of our considered moral judgments are justified.

Given this possibility, it should be clear that the corrective reply is correct in asserting that we owe no story about why we should trust our current set of intuitions. However, the question remains, why should we trust our considered moral judgments in reflective equilibrium? With the modification of the method of reflective equilibrium suggested in section 4.6 we can see why we should trust these judgments. We should trust them because they form a part of a highly coherent set of judgments, a set of judgments that includes a well-grounded set of background theories and systematizing principles.

We can see exactly how this is supposed to work if we consider an argument from Stefan Sencerz. Sencerz introduces a distinction between derivative credibility and non-

derivative credibility. Moral intuitions have *non-derivative credibility* “if and only if we can ascribe to them a kind of initial credibility independent of their being derived from, or being dependent on some moral principles [and, I would add, background theories].”³¹ A moral intuition has *derivative credibility* if and only if it has credibility that is dependent on moral principles or background theories. It should be clear that my view is that moral intuitions have merely derivative credibility. Their credibility is dependent upon the inferential or evidential relationships they bear to other principles or background theories.

However, Sencerz goes on to argue that for a defender of the method of reflective equilibrium to claim that moral intuitions only have derivative credibility is viciously circular.³² This is because the credibility of our intuitions would be dependent on moral principles and background theories. However, these principles and background theories are justified by their relationship to our moral intuitions. This certainly does appear to be a vicious circle

I think Sencerz’s description of the view is misleading. First, Sencerz describes things as if what is justified is not particular beliefs but *sets* of beliefs (e.g. intuitions and principles). Second, Sencerz’s description has one of set of beliefs, intuitions, being justified in what seems to be a linear fashion by another set of beliefs, principles, which are in turn justified linearly by intuitions. This is surely viciously circular but it is not the picture that I, or other defenders of the method of reflective equilibrium, advocate. First, what we aim to justify is not sets of beliefs (e.g. considered moral judgments) but instead

³¹ “Moral Intuitions and Justification in Ethics,” p. 92n5.

³² Ibid, p. 79

particular considered moral judgments. Each considered moral judgment is justified by its complex inferential and evidential relationships to a large set of beliefs that not only includes other considered moral judgments but also moral principles and background theories. These particular moral principles and background theories are then justified by their complex inferential and evidential relationships to each other and our considered moral judgments.

There is still an element of circularity here. A particular considered moral judgment C is justified by its relationship to our other moral beliefs. Suppose these moral beliefs include a further belief D. D itself would be justified by its relationship to a large number of moral beliefs that includes C. This is a kind of circularity. However, it is a very minimal circularity. If one objects to this kind of circularity, then one objects not only to our account, but to every coherentist theory of justification, moral or non-moral, as well as some of the weaker versions of foundationalism. Of course, one might press on and say that every single coherentist theory must be rejected because it involves this kind of circularity. However, I will not take up that argument here because it seems to me that moral epistemology has nothing unique to contribute. That battle must be fought in more traditional epistemology.

This completes my response to the worry about intuitions. To sum up, I agree with the corrective reply in claiming that it is not our initial set of considered moral judgments that requires credibility, it is our set of moral judgments in reflective equilibrium. However, as the dogmatist brings out, the corrective reply is not in a position to explain the credibility of those considered moral judgments in reflective equilibrium. Fortunately, the method of reflective equilibrium can be modified in the

following way: it can be modified so that when an agent is faced with an inconsistency, she must make that revision which maximizes overall coherence. This modification is not *ad hoc* since it derives its motivation from the coherentist considerations that motivate the method of reflective equilibrium in the first place. With this modification in place, we can see why we should trust our considered moral judgments in reflective equilibrium. We should trust them because they cohere well with our best evidence. These judgments have strong evidential and inferential relationships to other beliefs in a highly coherent set of judgments, a set of judgments that includes a well-grounded set of background theories and systematizing principles. In making this modification we are not claiming that moral intuitions have non-derivative credibility, only derivative credibility. However, to make this claim is not viciously circular since it involves only the kind of weak circularity that we find in all coherentist accounts of epistemic justification.

4.11 The Moral Center

Now that we have an idea of the role of coherence in the theory of ethical justification, it will be useful to return to some points made in our earlier discussion of the metaphor of the web of belief. Since, the notion of centrality to the web of belief plays a crucial role in the theory of ethical justification, it is a natural question to ask, ‘which moral beliefs lie closest to the center of our web of belief?’

This question is ultimately an empirical question and answering it would require laying out one’s moral beliefs and carefully examining the inferential and evidential relationships between them. While I cannot undertake such a project here, I do want to make a few points to forestall certain kinds of misunderstanding.

Earlier I suggested that the sorts of non-moral beliefs that lie closest to the center of our web of belief are our beliefs about mathematics, logic, and the reliability of sense perception. I suggested that my beliefs about what I wore yesterday or ate last week lie close to the periphery of our web of belief. This may give the impression that what lies at the center of our web of belief are deep abstract principles and what lies closer to the periphery is our more ordinary and particular beliefs. This may in turn suggest that the moral beliefs closest to the center of our web of belief are deep principles in ethical theory, or perhaps even metaethics, and the moral beliefs that lie at the periphery of our web of beliefs are our judgments about particular cases, types of cases, institutions, or policies. This impression, though perhaps suggested by our earlier discussion, is mistaken and I want to correct it here.

Recall that how central or peripheral a belief is to one's web of belief is determined by the contribution that belief makes to the comprehensiveness and connectedness of one's overall set of beliefs. It is also important to keep in mind that just as moral principles will provide evidential support for particular judgments, particular judgments will provide evidential support for moral principles. It is often the case that we subscribe to a moral principle because of the results it entails in a relatively small number of cases. Given these two facts, it is not at all obvious that abstract moral principles will lie at the center of our moral web. It may be that certain more particular moral judgments, say that racial discrimination is unjust or torture for amusement is wrong are far closer to the center of our web of belief than much of the sorts of beliefs we have in the more abstract parts of normative ethics and metaethics. For instance, my belief about whether or not a certain objection to a specific version of rule

consequentialism succeeds will likely have far fewer evidential and inferential relationships to my other beliefs than my view that racial discrimination is unjust. Thus, it is likely that in this case my belief about abstract normative ethics will be closer to the periphery than my more particular belief about racial discrimination.

These points help to show how I answer an objection that has been pressed the contributionist account. It has been suggested that even if this account solved the problem of the dogmatist, as I presented it, we can imagine a theory dogmatist who holds on to some abstract moral belief in the face of widespread evidence against it.³³ One might worry that the contributionist account cannot show that a theory dogmatist might have to revise his view in the face of widespread evidence against it. This objection is plausible if we think that abstract moral principles lie at the center of my web of belief while more particular moral beliefs lie at the periphery. However, as I have argued above, this is not the case. The moral beliefs closest to the center of my web of belief are likely to include certain beliefs about particular cases, types of cases, policies, and institutions. If these judgments conflict with an abstract moral principle, we will have to revise the principle unless the principle itself can be shown to better cohere with our evidence.

This allows the contributionist account to accommodate an important point made by Rawls. Rawls argues that within the method of reflective equilibrium certain “convictions are provisional fixed points which we presume any conception of justice

³³ This objection was raised by John Maier in his comments on my presentation at the Rocky Mountain Ethics Congress in August, 2009. Discussion with Maier pushed me to make explicit the points made in this section.

must fit.”³⁴ In our theory we can say that these provisional fixed points are the moral beliefs that are closest to the center of one’s web of belief. Of course, as Rawls notes, “even the judgments that we take to be provisionally fixed are liable to revision.”³⁵ In extreme circumstances, even these judgments may ultimately have to be revised. However, in the absence of such extreme circumstances we can be reasonably certain that these judgments are justified and that other judgments will have to be revised to accommodate them. We get this result not by appealing to anything independent of coherence, but by appealing to coherence itself. These judgments have this special place in the method of reflective equilibrium because of the inferential and evidential relationships they bear to other beliefs. It is worth emphasizing that we can get the result that these judgments are provisional fixed points without making the dubious assumption that the contents of these judgments are self-evident moral truths.

A related, but distinct objection that has been pressed against the proposal here is that it stacks the deck against pluralistic moral theories.³⁶ Some moral theories, such as act-utilitarianism, are extremely unified in that they claim that only one moral principle that can tell us whether or not any action is right or wrong by itself. Other theories are more pluralistic. Rossian deontology, for example, proposes a plurality of principles and requires judgment to determine what to do in cases of conflict. Other theories such as Aristotle’s are even less unified than Ross’s. It has been suggested that the contributionist account’s emphasis on coherence wrongly stacks the deck in favor of

³⁴ Rawls, *A Theory of Justice*, p. 20.

³⁵ Ibid.

³⁶ This objection was also pressed by Maier in his comments and Patricia Greenspan in discussion.

unified moral theories. This is a serious objection because how unified a theory should be is something that needs to be settled within a theory of justification not presupposed by it.

In response I want to point out that pluralist moral theories are often motivated by background theories about the role or function of ethics. This is certainly the case with Aristotle's moral theory. If these background theories cohere well with our evidence, they may give us reason, within the method of reflective equilibrium, to reject unified moral theories. Other times, pluralist moral theories are motivated by their ability to capture judgments about particular cases or types of cases more naturally than unified moral theories. Again, supposing these more particular judgments cohere with our best evidence, they give us reason, within the method of reflective equilibrium, to reject unified moral theories that cannot capture them.

Of course, the act-utilitarian will respond by claiming that such background theories either fail to cohere with our evidence or fail to support pluralism. The act-utilitarian will also make these claims about the relevant case judgments. My purpose here is not to adjudicate this dispute but only to note that it takes place within the method of reflective equilibrium. The method itself does not assume that one side or the other is correct.

4.12 Ideal and Non-Ideal Moral Reasoning

This completes my presentation of the theory of ethical justification. As I stated at the beginning of chapter 3, my view is that a moral belief is justified if and only if it would be the output of a version of the method of wide reflective equilibrium. The version of the method of wide reflective equilibrium that I endorse differs from that of

Rawls, Daniels, and Brink in that it makes determinate the process of belief revision in reflective equilibrium. In particular, it holds that when one finds an inconsistency, one must revise so as to maximize coherence. My account also differs from earlier defenses of the method of reflective equilibrium in that I do not concede that the plausibility of the method of reflective equilibrium hinges on providing any general story about why our intuitions are reliable. The credibility of those judgments that are justified comes from the evidential and inferential relationships they have to our other beliefs in reflective equilibrium.

Even though my presentation of the theory is complete, our work is not done. There are a rather large number of objections that have been proposed against reflective equilibrium accounts and if these objections are successful, they would rebut not only Rawlsian reflective equilibrium but my own variant as well. For this reason, I will devote the next chapter to showing that these objections fail.

In the remainder of this chapter, I want to consider an objection to the view that I have presented. I deal with this objection here, rather than the next chapter, in order to indicate that I give this objection a sort of priority. This objection cuts to the heart of what it is that our theory of justification is supposed to achieve.

This objection begins by noting that way back in section 1.1, I stressed the importance of a theory of justification being useful to actual believers trying to decide what to believe. However, it is not at all clear that the account I have given is useful to actual believers. Following the reflective equilibrium procedure would require being able to have all your moral beliefs, as well as the inferential and evidential connections between them, in your mind, so to speak, at the same time. By the phrase ‘in your mind,’

I mean it would require you to be cognizant of all of your moral beliefs at the same time. However, no human being has the working memory that would be required for such a project. Furthermore, even if one did, my account requires a moral believer to maximize coherence. However, I myself have not been able to state fully what is involved in coherence. What chance do non-philosophers have? Finally, even if we could figure out exactly what coherence was, there is some dispute about whether or not the project of determining the coherence of a complex set of beliefs is computationally tractable.³⁷ Given all of this, how can a theory like mine be useful to actual human moral believers?

In some sense, this is a variation on an old complaint. No one could actually achieve reflective equilibrium. As Rawls puts it, “Reflective Equilibrium... is a point at infinity we can never reach, though we may get closer to it in the sense that through discussion, our ideals, principles and judgments seem more reasonable to us and we regard them as better founded than they were before.”³⁸

The fact that none of us can ever achieve reflective equilibrium shows us that none of us will ever be fully justified in all of our moral beliefs. This is, I grant, an unhappy state of affairs. It would be better if there were some sure fire way to arrive at fully justified moral beliefs. However, a theory of justification that had this consequence would not do justice to the phenomenology of moral reflection. All of us who have engaged in serious moral reflection can agree about at least one thing; it is difficult. A theory that held that there were some easy to follow decision procedure for arriving at

³⁷ Paul Thagard and Karsten Verbeurgt argue in “Coherence as Constraint Satisfaction,” *Cognitive Science* 22 (1998): 1-24, that coherence itself is not computationally tractable but that there are algorithms which are computationally tractable that approximate coherence. I think that his claim dovetails nicely with what I say below.

³⁸ “Reply to Habermas,” *Journal of Philosophy* 92 (1995), p. 142

justified moral beliefs would not do justice to the difficulty of moral reflection. I concede, then, that this implication of our theory of justification is unhappy but it is, unfortunately, accurate.

As Rawls indicates above, the fact that we can never attain reflective equilibrium does not show us that the method of reflective equilibrium is worthless. While we cannot attain reflective equilibrium, we can follow procedures that get us closer. Let me suggest one such procedure. During the course of an agent's life, she seeks out opportunities for moral reflection. When she reads a novel or hears a story on the news, she thinks carefully about what she takes to be the moral implications of the events described. If she is like me, she will fairly often discover inconsistencies in her moral beliefs. When she does, she thinks carefully about the implications of her inconsistent beliefs. In particular she pays close attention to the connections between her inconsistent beliefs and her other beliefs (obviously, she will not be able to discover all of these connections). She then revises the belief that seems to her, after careful reflection, to be closer the periphery of her web of belief.

Given the cognitive limitations of human beings, this procedure will not lead one to reflective equilibrium. However, I contend that following this procedure throughout one's life will get one closer to reflective equilibrium. While reflective equilibrium is not something that can be attained, it is something that we can strive for. In providing a target to strive for, our theory of justification becomes useful to actual human moral believers after all.

Chapter 5: In Defense of Reflective Equilibrium: A Response to Some Objections

5.1 Reflective Equilibrium and Its Critics

In the nearly forty years since the publication of *A Theory of Justice*, the method of reflective equilibrium has been subject to a wide range of objections. In my view, many of these objections are interesting and require the defender of the method to make explicit some important background assumptions and features of the procedure. However, I believe that all of these objections can be answered without modifying the method of reflective equilibrium.

A slight complication emerges because the version of the method that I defend is somewhat different from the version that others defend. The objections that I consider in this chapter were all formulated to rebut the version of the method that Rawls and Daniels have defended. Nonetheless, I take it that all of these objections turn on features of Rawlsian reflective equilibrium that are shared by my own version. Therefore, it is important part of this project to answer these objections.

5.2 Reflective Equilibrium and Moral Conservatism

Objection: The method of reflective equilibrium is unduly conservative in the sense that it merely serves to reaffirm the status quo. This is because it asks agents to test principles against their intuitions but merely doing this is unlikely to result in any major revisions to our current set of moral beliefs.¹

¹ This objection is raised by Hare in "Rawls Theory of Justice;" Singer in Sidgwick and Reflective Equilibrium," and Wood in *Kantian Ethics*, p. 65

Reply: This objection ignores the fact, emphasized in the last chapter, that the joint pressures of systematization and accommodating background theories can produce extremely radical changes in one's set of moral beliefs. Indeed it is consistent with the method of reflective equilibrium that an agent may have to revise every single one of her intuitive moral beliefs. Because the method allows for such radical revision, it is not committed to any kind of conservatism and it has the resources to allow for extreme changes to status quo morality.

This argument shows that the structure of the method allows for radical revision. Critics, however, may not be convinced until they see how, in practice, the method of reflective equilibrium allows for, or even requires, one to modify a deeply held moral belief. I recognize the force of this challenge and attempt to meet it in the next chapter.

5.3 Reflective Equilibrium and Moral Relativism

Objection: It seems that two agents who accepted very different sets of considered moral judgments may end up with very different beliefs in reflective equilibrium. This fact shows that the method of reflective equilibrium is objectionably relativist or subjectivist. It has the implication that one set of moral beliefs may be true for one person but not another.²

² This objection has been raised in Hare, "Rawls' Theory of Justice," p. 82; Singer, "Sidgwick and Reflective Equilibrium," pp. 30-31; and Wood, *Kantian Ethics*, p. 51. Singer uses the word 'valid' rather than 'true,' 'however, 'true' seems to me to be the better word here since philosophers primarily use 'valid' to refer to inferences, not propositions (it sounds odd to say 'abortion is wrong is invalid.'). 'True' may seem to have its own set of problems because some non-cognitivists have denied that moral statements can be true. However, most contemporary non-cognitivists have no problem with attributing truth to moral propositions.

Reply: It is worth noting that the question of how much moral disagreement there would be if everyone had completed the method of reflective equilibrium is an open one. Brink puts the point as follows:

Because the dialectical process of the coherence theory can produce fundamental changes in the beliefs with which people begin the process, the fact that people begin the process with significant differences is no evidence that their disagreement would persist to the end of the dialectical process. Because the dialectical process is one we can at best approximate, even the existence of reflective moral disagreement fails to show that moral disagreement is in principle inevitable... For as we have seen, the moral beliefs with which people might emerge from the dialectical process need not overlap at all with the beliefs with which they started the process. This makes it possible, at least in principle, for those who enter the process with completely different moral beliefs to emerge with the same beliefs.³

This point is worth emphasizing since many critics of the method of reflective equilibrium speak as if the output of the method is merely a function of the considered moral judgments that one begins with.⁴ Brink responds to these critics by noting, as we did in the last section, that the method may produce radical changes in one's moral beliefs. Thus, the fact that persons currently disagree about the truth of moral propositions does not show that they will continue to disagree after the process is complete. Of course, there is no guarantee that there will be substantial agreement at the end of the process either.

Let us assume that there will remain substantial disagreement at the end of the process. Does this mean that the method has relativist or subjectivist implications? One

³ Brink, *Moral Realism and the Foundations of Ethics*, pp. 142-143.

⁴ See for example Jonathan Weinberg, Shaun Nichols, and Stephen Stich, "Normativity and Epistemic Intuitions," *Philosophical Topics* 29 (2001): 429-460. I do not discuss their paper here since their target is the use of the method of reflective equilibrium to generate epistemic norms rather than moral norms. However, the objection they discuss in this paper is a species of the objection being discussed here. For this reason, I believe what I say here constitutes a response to their objection.

must keep in mind that relativism and subjectivism are theories of what make moral claims true. For example, cultural relativism holds that the truth of a moral claim is dependent on the practices, beliefs, or traditions of one's culture. Subjectivism, in one form, holds that the truth of a moral claim is dependent on the mental states of the person making the claim.

The method of reflective equilibrium is *not* an account of what makes a moral claim true, it is an account of what makes a moral belief epistemically justified. Thus, it is not committed to any particular theory of moral truth and *a fortiori*, it is not committed to relativism or subjectivism.

Nonetheless, it is worth noting that the method of reflective equilibrium does have the implication that a moral belief can be epistemically justified for one person but not another. This implication, however, is not problematic. As we noted when discussing the concept of justification in section 1.1, whether or not one is justified in having a belief is surely dependent on what evidence is available to her. This means we must be prepared to accept that if you and I have very different evidence I may be justified in believing something that you are not and vice-versa.

Perhaps this confusion has arisen because some of reflective equilibrium's critics think that justification implies truth. It does not. If it did, then the traditional analysis of knowledge as justified true belief would be redundant. We can be, and often are, justified in believing false things. Testimony provides a useful example. Most of us who are not skeptics believe that if I tell you P, you have every reason to believe that I have knowledge of P, and you have no reason to suspect that I am lying about P, then you are justified in believing P. However, none of us think that testimony in these circumstances

is infallible. For a concrete example, suppose you believe that I wore a green t-shirt yesterday on the basis of my testimony. Suppose further that I have a good memory, am quite honest, and have no reason to lie. It seems that if you are ever justified in believing something on the basis of testimony, then you are justified in this case. However, I might be misremembering. If I am, then you have formed a justified false belief.

5.4 Reflective Equilibrium and Moral Truth

Objection: I can see that the method of reflective equilibrium has some virtues. It obviously leads one to a coherent set of beliefs, and there is a sense of ‘rationality’ in which reflective equilibrium leads one to beliefs that are rational to hold. However, the primary aim of belief is not coherence or rationality, it is truth. Indeed we are concerned about the coherence or rationality of our beliefs only insofar as coherence and rationality lead to truth. One thing that has not been shown is that following the method of reflective equilibrium will lead one to moral truth.⁵

Reply: This is of course, a variation on a familiar objection to coherentist theories of non-moral justification. Since belief aims at truth, a theory of what we ought to believe should tell us how to increase the probability that our beliefs are true, but it is far from clear that coherentist theories do this.

One might raise this objection because they have a particular theory of moral truth in hand. For such a critic the objection amounts to “A moral belief is true if and only if it meets a set of conditions C. Why think that the method of reflective equilibrium tends to lead one to beliefs that meet conditions C?” If this is all that the objection amounts to, it

⁵ This objection is raised by Brandt, *A Theory of the Good and the Right*, p. 20.

can be answered using a familiar pattern of argument. Either the theory of moral truth coheres well with the other things we believe or it does not. If it does not, then it is not clear why should we believe the theory of moral truth in the first place. If the theory of moral truth does cohere well with the other things we believe, then the theory of moral truth itself can play a role in the method of reflective equilibrium. It will be a background theory. If the background theory is sufficiently well-grounded, we will have to revise our other moral beliefs to cohere with the theory of moral truth and we will end up with a set of moral judgments that is not just coherent but coherent with a plausible theory of moral truth.⁶

One may raise this objection without any particular theory of moral truth in mind. In this case, I am less sure how to answer the objection. However, it is not the fact that I am a coherentist that makes the objection in this form hard to answer. I cannot see how a foundationalist could hope to show that their theory gets one closer to moral truth without invoking a particular theory of moral truth.

Allow me to make a controversial supposition. The nature of moral truth is not immediately graspable in the way that the truth of propositions such as ‘ $7 + 5 = 12$ ’ are. How then, could we hope to find out what moral truth is like? My guess is that we will

⁶ To see how a moral realist might expand upon this claim, the reader should refer to Brink, *Moral Realism and the Foundations of Ethics*, pp. 125-133. There, Brink argues that second-order realist beliefs (e.g. “beliefs about our psychological makeup, our cognitive and perceptual equipment and their hookup to the world,” p. 127) are among the beliefs that one seeks to bring into coherence. This solution parallels Laurence Bonjour’s defense of a coherentist epistemology and a realist metaphysics in *The Structure of Empirical Knowledge*, Ch. 8. Both Bonjour and Brink go further and suggest that second-order beliefs are *necessary* for a set of beliefs to be justified. I have doubts about this further suggestion because it seems to imply that ordinary people who have no particular view about the metaphysics of moral (or non-moral in Bonjour’s case) properties cannot be justified in having moral beliefs.

have to think carefully about our substantive moral judgments as well as various background theories that we hold in metaphysics, the philosophy of language, and perhaps other areas of philosophy and even the sciences. We will then have to begin with the theory of moral truth that best fits with these judgments and background theories. Of course, we may eventually have to modify some of the judgments and background theories to accommodate our account of moral truth and we may have to modify our initial account of moral truth to fit further judgments and background theories. We will have to work back and forth until everything fits together in a coherent way. Of course, this procedure looks a lot like the method of reflective equilibrium. If my supposition is right, then the search for the nature of moral truth is not divorced from the general search for reflective equilibrium.

One may doubt that the only way to reach an account of moral truth is to start with our substantive moral judgments and background theories. But such a critic would have to offer an alternative. Short of the implausible claim that the nature of moral truth is revealed by a sort of Cartesian light of reason, I do not know how else to arrive at a theory of moral truth.

5.5 Reflective Equilibrium and Moral Skepticism

Objection: The theory of justification defended here begs the question against moral skepticism, the view that no moral belief is justified. The theory defended here claims that a moral belief is justified if and only if it would be the output of the method of reflective equilibrium. It follows, then, that those moral beliefs that would be the output

of the method are justified. This is an assertion of what moral skepticism denies, namely that some moral beliefs are justified, and is thus question begging.⁷

Reply: Since this objection rests on a confusion, it is necessary to sort matters out before responding directly to the objection. In pressing this objection, David Copp writes, “skepticism itself should be counted as a theory of moral justification.”⁸ There is a sense in which this is misleading.

A theory of justification will typically present us with a condition or set of conditions that a belief must meet in order to be justified. Such a set of conditions will have an *output*, a list of which, if any, beliefs meet these conditions and are thus, according to the theory, justified. We must be clear to distinguish between a theory and its output.

Moral skepticism, by itself, is a view about the output, about which beliefs are (not) justified, not a view about the conditions under which a belief is justified. This can be seen from the standard way in which skeptics argue.

- 1) In order for beliefs in domain D to be justified, these beliefs must meet condition N
- 2) Beliefs in domain D do not meet condition N
- 3) Therefore, beliefs in domain D are unjustified.⁹

⁷ Copp, “Considered Moral Judgments and Justification: Conservatism in Moral Theory.”

⁸ Ibid, p. 142.

⁹ Many familiar skeptical arguments fit into this schema. Consider Barry Stroud’s well known reconstruction of the Cartesian argument for skepticism (*The Significance of Philosophical Scepticism* (Oxford: Oxford University Press, 1984), Ch. 1. While Stroud originally puts the argument in terms of knowledge, I put it in terms of justified belief but this makes no difference for my point about the structure of the argument.)

It is important to see that skepticism is the conclusion of this argument, not a premise in the argument. Thus, moral skepticism is the view that moral beliefs are unjustified, external world skepticism is the view that beliefs about the external world are unjustified, skepticism about other minds is the view that beliefs about other minds are unjustified, etc. In principle, two moral skeptics could agree that no moral beliefs are justified but disagree about which conditions a moral belief would have to meet in order to be justified.

This is why it is misleading to say that skepticism is theory of justification.

Foundationalism is a theory of justification in that it lists a set of conditions that a belief must meet in order to be justified. It is a further question whether there are any moral beliefs that meet those conditions. Skepticism is a view about which beliefs are justified, not a view about the conditions under which a belief is justified. Thus, foundationalism need not beg the question against the skeptic.

-
- 1) In order for a belief B to be justified we must know the falsity of all the things that we know to be incompatible with the truth of B
 - 2) For any belief about the external world we do not know the falsity of all the things we know to be incompatible with the truth of our beliefs about the external world (This is because we do not know that we are not dreaming, deceived by an evil demon, etc.)
 - 3) Therefore, our beliefs about the external world are not justified

Peter Unger's argument from certainty (*Ignorance: A Case for Skepticism* (Oxford: Oxford University Press, 1975), Ch. 3) can also be understood this way. For a belief to be justified it must be certain. Beliefs about the external world are not certain. Therefore, they are not justified. Joyce's argument for moral skepticism can also fit into this schema. If we have evolved to form some beliefs regardless of whether they are true, then those beliefs are not justified. We have evolved to form moral beliefs regardless of whether they are true. Therefore, moral beliefs are not justified (Joyce, *The Myth of Morality*, Ch. 6).

Similarly, the theory of justification defended here is a theory of justification in that it gives a condition that a moral belief must meet in order to be justified: it must survive the reflective equilibrium process. It is a further question whether or not any moral beliefs survive this process. It is a possibility that, in working through the reflective equilibrium process, the agent will discover that, perhaps given certain background theories, it is best to reject all of her moral beliefs.¹⁰ Indeed, Joyce has taken exactly this position. He writes:

Pursuing a strategy of “wide reflective equilibrium” may lead to a wholesale exclusion of moral discourse just as smoothly as it may lead to a vindication. I have argued that once we see clearly the quasi-mystical commitments embodied by moral discourse, elimination is the more probable result.¹¹

Joyce is correct that wide reflective equilibrium “may lead to a wholesale exclusion of moral discourse.” To be clear, I strongly disagree with Joyce that all of moral discourse embodies “quasi-mystical commitments” and I think that vindication of at least some more beliefs is much more likely than wholesale elimination. However, these disputes are best left for another occasion. Here, I am content to point out that, *pace* Copp, the dispute between Joyce and I takes place within the method of reflective equilibrium. Because reflective equilibrium is a theory about the conditions under which a belief is justified and not a list of beliefs that are justified, it does not beg the question against skepticism.

¹⁰ This point is also made by T.M. Scanlon in “Rawls on Justification,” in Samuel Freeman (ed.), *The Cambridge Companion to Rawls* (Cambridge, UK: Cambridge University Press, 2003), pp. 139-167.

¹¹ Joyce, *The Myth of Morality*, p. 170.

5.6 Reflective Equilibrium and Deductive Moral Arguments

Objection: There is a tradition in moral philosophy, call it *deductivism*, that attempts to justify moral theories and particular moral judgments by deductive argument. Why should we prefer the method of reflective equilibrium to *deductivism*?

Reply: Examples of what I am calling deductivism can be found in the work of Brandt, who attempts to derive a version of rule-utilitarianism from a theory of rationality¹² and R.M. Hare, who attempts to derive a utilitarian moral theory from a semantic analysis of moral terms conjoined with facts about human nature.¹³ It is, I think, no coincidence that Hare and Brandt are two of the method of reflective equilibrium's most vocal critics. For they see the method as in competition with the deductivist view they favor.

Though these authors seem to find deductivism inconsistent with the method of reflective equilibrium, there is actually no inconsistency. These thinkers both take it that there are strong reasons to favor their theories of practical rationality and/or the semantics of moral discourse. If there are such strong reasons, then the method of reflective equilibrium can accommodate such theories as background theories. If these theories deductively entail a set of moral judgments, there will be strong reason to accept these moral judgments, even if they conflict with other things that the agent believes.

Thus, my response to the deductivist objection is simply that the deductivist method is not inconsistent with the method discussed here. Though deductivists often

¹² Brandt, *A Theory of the Good and the Right*

¹³ See Hare, *The Language of Morals* (Oxford: Oxford University Press, 1952); *Freedom and Reason* (Oxford: Oxford University Press, 1963), and *Moral Thinking: Its Levels, Method, and Point* (Oxford: Oxford University Press, 1981).

believe they must reject the method of reflective equilibrium to preserve their deductivism, they need not.

5.7 The Charge of Vacuity

I have argued that the method of reflective equilibrium is quite a bit more flexible than its critics take it to be. For instance, it is compatible with the view that there are irresolvable moral dilemmas, moral questions about which there is no right answer, and deductive arguments for moral propositions. This naturally gives rise to a quite different charge that the method is vacuous. For instance, Singer writes,

The price for avoiding the inbuilt conservatism of the narrow interpretation [of the method of reflective equilibrium], however, is that reflective equilibrium ceases to be a distinctive method of doing normative ethics. Where previously there was a contrast between the method of reflective equilibrium and “foundationalist” attempts to build an ethical system outward from some indubitable starting point, now foundationalism simply becomes the limiting case of a wide reflective equilibrium.¹⁴

In this passage, Singer is unfair even to the foundationalist position that he is sympathetic with. For, as we saw in Chapter 1, the foundationalist need not, and should not, claim her starting points are “indubitable” but only that they are *prima facie* justified independently of their evidential and inferential relationships to other propositions.

Even if we accept this more modest account of foundationalism, there are two crucial contrasts with the coherentist position discussed here. First, the coherentist denies that any belief is *prima facie* justified independently of its inferential and evidential relations to other propositions. Thus, even allegedly foundational propositions require inferential or evidential support from other beliefs in order to be justified.

A second contrast is that for a foundationalist, foundational beliefs are *required*

¹⁴ Peter Singer, “Ethics and Intuitions,” *Journal of Ethics* 9 (2005), pp. 331-352, p. 347.

for non-foundational beliefs to be justified, whereas a coherentist can deny this. This is a contrast where, I believe, coherentism emerges as the clearly superior position. Suppose we adopt Sidgwick's foundationalism according to which certain consequentialist propositions are foundational. According to moral foundationalism, non-foundational moral beliefs are only justified if they have some evidential or inferential relationship to foundational moral beliefs. In other words, non-foundational beliefs can only be justified on the basis of foundational beliefs. Suppose, plausibly, that an agent can only be justified in believing *p* on the basis of *q*, if the agent also believes *q*.¹⁵ Propositions that an agent does not believe, or that she actively disbelieves cannot render her justified in believing other propositions. It follows that for an agent to be justified in believing any moral proposition, the agent would have to believe Sidgwick's consequentialist axioms. In other words, deontologists could not be justified in believing that torturing babies solely for amusement is morally wrong. This result strikes me as quite absurd even if we were to accept Sidgwick's utilitarianism.

Non-utilitarian foundationalists are likely to reply that this is due to the inherent implausibility of Sidgwick's foundational propositions but the same problem can be raised for a Rossian deontology. Suppose that it is foundational that we have certain *prima facie* duties and that any non-foundational moral belief is justified by bearing an inferential or evidential relationship to these foundational beliefs. Suppose further, that beliefs that the agent does not hold or actively rejects cannot render an agent justified in holding further beliefs. As we saw in chapter 1, many utilitarians are committed to

¹⁵ This is the first clause of what Richard Fumerton calls "the principle of inferential justification" (*Metaepistemology and Skepticism* (Lanham, MD: Rowman & Littlefield, 1995), p. 36). As Fumerton notes, this principle is widely accepted by foundationalists.

denying that we have Rossian *prima facie* duties. Thus, it follows that such utilitarians have no justified moral beliefs. But again, this strikes me as absurd.

In any case, the main point of this section is not to object to foundationalism yet again but to re-emphasize the contrasts between the coherentist position advocated here and foundationalist positions. Even in the very wide method of reflective equilibrium that we have been considering, these contrasts remain, and thus the theory is not vacuous.

Chapter 6: Aid to the Worst Off: A Case Study

6.1 Reflective Equilibrium, Moral Conservatism, and Deeply Held Moral Intuitions, Again

In the last two chapters, I have argued against two related objections to the method of reflective equilibrium. (1) Critics have charged that the method does not have the resources to require, or even allow for, the revision of strongly held intuitive judgments and (2) it has been argued that the method is unduly conservative, in the sense that it would merely reinforce the moral status quo. In previous chapters, we have seen that there are reasons to doubt both of these claims but these reasons were based on abstract arguments from the structure of reflective equilibrium. A critic may, perhaps not unreasonably, retain her doubts in the face of these abstract arguments until she sees that, in practice, the method may require revisions to deeply held moral intuitions and/or status quo morality.

In order to meet this challenge, I turn now to Peter Singer's well-known paper, "Famine, Affluence, and Morality."¹ The conclusion of Singer's paper is starkly at odds with our moral intuitions and the moral status quo. Singer is not exaggerating when he claims that, if his argument is correct, "the whole way we look at moral issues – our moral conceptual scheme – needs to be altered, and with it, the way of life that has come to be taken for granted in our society (p. 146)." In this chapter, I argue that an application of the method of reflective equilibrium modeled on Singer's original argument yields Singer's conclusion. Given how radical Singer's conclusion is, this

¹ *Philosophy and Public Affairs* 1 (1972): 229-43. Reprinted in *Unsanctifying Human Life* ed. by Helga Kuhse (Oxford: Blackwell, 2003), pp. 145-56. Page references refer to the reprint. All unattributed references in the main text refer to this work.

result would no doubt be very surprising if we thought the method of reflective equilibrium merely served to reinforce the status quo.

Before we begin, a few qualifications about the scope of the argument are necessary. My primary aim is neither to directly argue that Singer's conclusion is correct nor even to argue that the method of reflective equilibrium compels one to accept Singer's conclusion. Instead, the main aim of this chapter is to show something about the structure of the method of reflective equilibrium, namely, that the method can require agents to revise deeply held intuitive judgments. One might retain her doubts about Singer's conclusion while being convinced of the main point about the structure of the method of reflective equilibrium.

Secondly, while I will be drawing from Singer's text frequently, my aim here is not quite Singer interpretation. That is, my aim is not to argue that Singer himself uses the method, though I believe his argument does resemble an application of the method in certain ways. Instead, I will construct a new argument that is modeled on Singer's original argument but is framed explicitly as an application of the method of reflective equilibrium.

6.2 Starting Points

One of the things that I emphasized in chapter 3 is that the method of reflective equilibrium begins with an agent's considered moral judgments as they are at the outset of inquiry. For this reason, it will be worthwhile to characterize the considered moral judgments that we start with before we encounter Singer's argument. Of course, I cannot guarantee that all readers have (or had) these considered moral judgments before encountering Singer's argument. However, I have informally surveyed students, friends,

and family members about their intuitions on these issues, and an overwhelming majority share the intuitions described below.

Consider the following case.

Pond: Albert is walking by a shallow pond where a young child is drowning.

Albert can save the child easily but, if he does, his fifty-dollar pair of shoes will be ruined. Albert decides not to go in and the child drowns.²

Almost everyone that I have put this case to insists that Albert's action is impermissible. Indeed, in his recent book, Singer notes that an event that was very similar to this fictional case actually occurred in Manchester.³ Two police community service officers (PCSOs) did not enter the water as a ten-year-old boy, Jordon Lyon, drowned in a pond. The officer's inaction was defended on the grounds that PCSOs, unlike regular police officers, are neither trained nor required by job description to rescue drowning victims. Even if attempting to save Lyon was not a duty the PCSOs had as a result of their occupation, those who were interviewed by the BBC seemed to think that the officers had a moral duty to go in anyway. For instance, Lyon's mother told the BBC, "I don't know why they didn't go in... if you're walking down the street and you see a child drowning you automatically go in that water."⁴

In presenting the case, I sometimes add the detail that Albert is generally a Good Samaritan and has, in the past, donated money to relief organizations and even saved

² This is, of course, a variation of Singer's famous pond case (p. 147).

³ Peter Singer, *The Life You Can Save* (New York: Random House, 2009), p. 4. The BBC's report of the event can be found at http://news.bbc.co.uk/2/hi/uk_news/england/manchester/7006412.stm.

⁴ Quoted at http://news.bbc.co.uk/2/hi/uk_news/england/manchester/7006412.stm

other children in similar situations. I have yet to find anyone whose evaluation of Albert's action is changed by this variation. I will refer to this modified pond case as the *Good Samaritan pond case*.

Now consider another case

Donation: Brad could donate money to a relief organization that would use the money to save an African child dying of a preventable disease. If Brad does not do so, the relief organization will be able to treat one less child. Instead of donating money, Brad decides to buy a \$50 pair of shoes. He does not need these shoes for his livelihood. He purchases them solely because they appeal to his sense of fashion.

Few people claim that what Brad did was wrong. Sometimes, rather than giving a case, I ask people for their general intuitions on donating to relief organizations. Most take the position that, while doing so is praiseworthy, no one is required to make a donation. This view is indeed widespread, as Singer notes.

The bodies which collect money are known as "charities." These organizations see themselves in this way – if you send a check, you will be thanked for your "generosity." Because giving money is regarded as an act of charity, it is not thought that there is anything wrong with not giving. The charitable man may be praised but the man who is not charitable is not condemned. People do not feel in any way ashamed or guilty about spending money on new clothes or a new car instead of giving to famine relief (pp. 149-150).

A sizable minority of those whom I ask insists that we do have some obligation to donate to aid organizations. However, when I begin asking what percentage of one's income is

appropriate, it is clear the amount that these persons have in mind is very small. Few will say that one is required to donate even one percent of their income, much less five or ten percent.

I will refer to our considered moral judgment about the donation case and our general considered moral judgments about donating to aid organizations collectively as our ‘considered moral judgments about aid.’ I will only distinguish between them when it matters for the argument.

6.3 Singer’s Argument as an Application of the Method of Reflective Equilibrium

After we have characterized our initial considered moral judgments, the next step is to find a principle that accommodates them. Singer proposes two principles both of which entail our judgment in the pond case.

Singer’s Strong Principle: “If it is in our power to prevent something bad from happening without thereby sacrificing anything of comparable moral importance, we ought, morally, to do it (p. 147).”

Singer’s Moderate Principle: “If it is in our power to prevent something very bad from happening, without thereby sacrificing anything morally significant, we ought, morally, to do it (p. 147).”

Singer prefers the strong principle but, in what follows, I will use the moderate principle unless I specify otherwise. This is because the moderate principle is less controversial and is still enough for Singer’s argument.

Singer’s principles clearly entail our considered moral judgment about the pond case. Albert may protest that his shoes are very nice, but we take it that they are not morally significant and thus he must sacrifice them in order to save the drowning child.

While these principles, therefore, have some intuitive plausibility, both of Singer's principles conflict with our intuitions about aid given three plausible assumptions.

- 1) "Suffering and death from lack of food, shelter, and medical care are bad (p. 146)."
- 2) Much of what we do with our money does not bring about anything morally significant
- 3) Our money could be used to prevent suffering and death from lack of food, shelter, and medical care.

I will follow Singer in saying little about 1). Virtually every moral theory is committed to 1) as is ordinary moral thought. 2) is difficult to deny as well. While we do spend money on necessities that have moral significance, it is implausible to think that most of the things we spend our money on (e.g. expensive food that brings little health benefit, a higher rent or mortgage to live in a nicer building in a nicer part of town, clothes for fashion rather than warmth, etc.) are morally significant. Of course, 2) leaves it open where this line should be drawn. This is the most precise we can be without getting into some difficult debates in moral theory. However, on any plausible way of drawing the distinction between the morally significant and insignificant, much of what we bring about with our money will fall on the insignificant side.

3) is widely accepted though perhaps not entirely uncontroversial. There are, of course, better and worse ways that aid organizations can use our donations and some well-meaning organizations have used donations inefficiently and some continue to do so. Some actions that seem like they would benefit the worst off actually do more harm

than good. For example, giving food directly to the poor often has the effect of driving down food prices in local markets and putting farmers out of business in regions where farming is one of the few ways in which persons can earn an income. For this reason, major aid organizations rarely give food directly to non-disabled persons except in emergency situations (e.g. natural disasters where the local food market has collapsed). Nonetheless, there is much that can be done with our money that would directly save lives. Many persons in Africa are dying of easily treatable diseases, some of which have been eradicated in the West. With contributions from private individuals, aid organizations are able to provide treatment. Recently, aid organizations have had much success preventing starvation by giving locals the resources needed to start their own business and provide for themselves.⁵

We can now see how this argument, modeled on Singer's own, has the structure of an application of the method of reflective equilibrium. We propose a principle to capture a considered moral judgment. Given some plausible assumptions, this principle conflicts with another set of considered moral judgments. In cases of conflict, the method tells us that we need to make revisions. Singer would urge us to abandon our considered moral judgments about aid in order to preserve a principle and our judgment about the pond case.

This is, I think, sufficient to establish that an application of the method of reflective equilibrium can imply that we should abandon a deeply held considered moral

⁵ For a philosophically motivated review of the empirical issues, see Garrett Cullity, *The Moral Demands of Affluence* (Oxford: Oxford University Press, 2004), Ch. 3 as well as the book's two appendices.

judgment. Even those not convinced that they should abandon their judgment can accept the point about the structure of the method.

Nonetheless, if Singer's conclusion is easily overturned one may retain their doubts that, in practice, the method can allow for substantial revision to our considered moral judgments or status quo morality. Thus, in what follows I will examine the most common responses to Singer's argument. Some of these responses, I will argue, fail on their own terms. Others may raise difficulties for the details of Singer's position but they do little or nothing to preserve our considered moral judgments as I characterized them above.

6.4 Kant on Imperfect Duties

An obvious response to what has been said so far is that we were too quick to accept Singer's principle. While Singer's principle does capture our judgment about the pond case, perhaps another principle can capture this judgment without sacrificing our considered moral judgments about aid.

While Singer's principles do not imply act-consequentialism, since neither principle requires us to maximize the good (though the strong principle requires us to minimize the bad), they do have a sort of consequentialist flavor to them. Perhaps the place to look for alternatives is in non-consequentialist theories.

Many familiar non-consequentialist restrictions on maximizing the good simply fail to apply here. Since both the pond case and the donation case involve allowing death rather than causing death, a moral distinction between doing and allowing will be of no help. For the same reason, the doctrine of double effect cannot be appealed to either. If

such well-known non-consequentialist restrictions are of no help, the next obvious place to look is the moral theory of Kant.

Unfortunately, the relevant passages from Kant are obscure and their proper interpretation is controversial.⁶ However, this need not detain us because all that matters for our purposes is whether or not we can glean a plausible account from Kant's texts. We can accept such an account for our philosophical purposes even if a careful reading of the text shows that Kant thought otherwise.

In the *Metaphysics of Morals*, Kant makes a distinction between perfect and imperfect duties. Perfect duties require an agent to or forbid an agent from performing actions that have certain properties. If I have a perfect duty to refrain from committing suicide, then I must never perform an action that has the property of being an instance of committing suicide. In contrast, imperfect duties require an agent to adopt certain ends rather than prescribing certain actions.⁷ A non-moral analogy may be helpful here. Suppose that I have adopted the end of becoming a successful professional philosopher. In order to fulfill this end, I must become familiar with the literature in my area, write philosophy papers, etc. However, suppose that on a particular occasion, when I could work toward becoming a professional philosopher, I decide to instead go have a drink with friends. This need not mean I have abandoned my end of becoming a professional

⁶ For three incompatible interpretations of the relevant passages see: Thomas Hill, *Dignity and Practical Reason in Kant's Moral Theory* (Ithaca, NY: Cornell University Press, 1992), Ch. 8; Marcia Baron, *Kantian Ethics Almost Without Apology* (Ithaca, NY: Cornell University Press, 1995); David Gauthier, *Kantian Consequentialism* (Oxford: Oxford University Press 1996), Ch. 6. My interpretation largely follows Hill's unless otherwise noted.

⁷ Kant says that imperfect duties "prescribe only maxims of actions rather than actions themselves (*Metaphysics of Morals*, p. 390)."

philosopher. So long as I regularly work towards my goal, it is compatible with my having this end that I sometimes forego opportunities to advance my end. Imperfect duties are duties to adopt an end but they leave it up to us how that end is to be promoted. For this reason, Kant says that an imperfect duty, “leaves a playroom (*latitudo*) for free choice in following (complying with) the law, that is, the law cannot specify precisely in what way one is to act and how much one is to do by the action.”⁸

Kant regards the duty of beneficence as an imperfect duty to adopt “the maxim of making other’s happiness one’s end.”⁹ If beneficence is regarded as an imperfect duty we must perform actions that promote the happiness of others. However, this is compatible with my deciding, on particular occasions, to forego performing such actions even when I could. This understanding of the duty of beneficence seems better able to accommodate our considered moral judgments about aid than Singer’s principles.

It is sometimes objected that this understanding of beneficence as an imperfect duty simply abandons our considered moral judgment about the Good Samaritan pond case.¹⁰ For if I regularly perform actions that promote the happiness of others, and beneficence is an imperfect duty, why can I not simply decide while walking past the pond that, on this occasion, I will forego saving the child?

This objection does not make use of the most plausible reading of what constitutes an imperfect duty. Consider the non-moral analogy again. While I may, consistently with my end of one day becoming a professional philosopher, forego

⁸ Ibid.

⁹ Ibid, p. 452.

¹⁰ Brad Hooker raises this objection in *Ideal Code, Real World* (Oxford: Oxford University Press, 2000), p. 161.

opportunities to promote this end, it does not follow that I am never rationally required to perform or refrain from specific actions. For instance, one might think that having drinks with friends rather than attending one's dissertation defense is inconsistent with having the end of becoming a professional philosopher. Similarly, while one who has adopted the end of promoting the happiness of others has some latitude in deciding how to do this, one who simply walked past a shallow pond where a child was drowning cannot seriously claim to have the end of promoting the happiness of others. Just as having the end of becoming a professional philosopher may require certain actions in certain circumstances, so does the end of promoting the happiness of others.

This response is plausible but it only gives rise to a further problem. We now have to explain why saving the drowning child in the pond case is required but sending in fifty dollars is not required in the donation case. What is it that makes these cases different?

The fact that this question comes up again shows that we are not far from where we started. What is interesting is that the Singer problem, the problem of accounting for our considered moral judgments about aid and our judgment about the pond case, comes up again within a Kantian deontology. Though Singer himself is a consequentialist, it is not only consequentialists that must grapple with this problem.¹¹

Nonetheless, we might think that we can make more progress on this problem within a Kantian framework. Consider the following passage from Kant.

¹¹ As we'll see below, the problem comes up within both a Kantian and Rossian deontology. Though I will not discuss it here, Scanlonian contractualism faces the same issue. See Elizabeth Ashford, "The Demandingness of Scanlon's Contractualism," *Ethics* 113 (2003): 272-203.

A wide duty is not to be taken as permission to make exceptions to the maxim of actions but only as permission to limit one's maxim of duty by another.¹²

This passage can be read as implying that one may forego promoting the happiness of others only if one is performing an action that satisfies another moral requirement.¹³ This would render Kant's understanding of the duty of beneficence roughly equivalent to that of Ross.¹⁴ On Ross's view we have a *prima facie* duty of beneficence. Recall from Chapter 1 that *prima facie* duties are duties that become all things considered duties unless they conflict with another duty. On Ross's view we have an all things considered duty to improve the condition of others unless doing so conflicts with another duty.

This Ross-Kant-inspired view certainly captures our judgment in the pond case. Albert is required to wade into the pond so long as he can do so without foregoing an opportunity to comply with another duty. However, this view is not consistent with our considered moral judgments about aid. Indeed, if foregoing an opportunity to fulfill another duty is an instance of "sacrificing something morally significant," then this view is consistent with Singer's moderate principle.

Earlier we saw that, in order to account for our judgment in the Good Samaritan pond case, we must concede that there are times when beneficence requires certain actions of us. However, once we make this concession, it becomes difficult to explain

¹² Kant, *Metaphysics of Morals*, p. 390.

¹³ Though this does not strike me as the best reading of this passage. For an argument against it see Hill, *Dignity and Practical Reason*, p. 151.

¹⁴ A point also made by Hill, *Dignity and Practical Reason*, p. 150. I say "roughly" because Kant understands the duty of beneficence as a duty to promote the happiness of another person while Ross understands it as a duty "make [the condition of others] better in respect of virtue, intelligence, or pleasure (*The Right and the Good*, p. 21)."

why beneficence does not require a certain action of us in the donation case. It may seem, then, that we can preserve our considered moral judgments about aid by giving up our judgment in the Good Samaritan pond case, instead holding that one who has already done significant good or made significant sacrifices for others may permissibly walk past the pond. While this certainly counts as an intuitive cost, it is a smaller intuitive cost than giving up our considered moral judgments about aid.

Accordingly, one might interpret an imperfect duty of beneficence as requiring one to put forth some effort to promote the happiness of others but hold that how much one must help is limited. Call this interpretation of the duty of beneficence the weak imperfect duty account. Since this account still requires us to help others sometimes, it will require one who has not saved lives in the past to wade into the pond. Thus, it need not give up our judgment in the standard pond case.

The weak imperfect duty account will not provide a complete vindication of our considered moral judgments about aid. Most of us are not faced with opportunities to save drowning children from ponds and will have to promote happiness in other ways. In so far as the duty requires saving lives on at least some of the occasions in which we are presented with opportunities to do so, it will require those of us who have not already rescued persons to donate sometimes.

Nonetheless, this view does not require the kind of radical giving that Singer advocates. While it does not vindicate our considered moral judgments about aid, it comes closer than Singer's view. It does, however, have intuitive costs that Singer's view does not have. Namely, it requires us to give up our judgment about the Good Samaritan pond case. In this way, I believe its implications are similar to those of Liam

Murphy's view, which I discuss in section 6.10. For our purposes we can leave open the question of whether or not this view has more overall plausibility than Singer's view and note that we have yet to find a principle that preserves our judgments in both the original and Good Samaritan pond cases as well as our considered moral judgments about aid.

6.5 Kamm on Distance

A promising strategy is to return to the cases that we began with. Perhaps if we can find a difference between the pond case and the donation case, we can use that difference as the basis for a principle that could explain our pond judgment without giving up our considered moral judgments about aid. Of course, there are many differences between the pond case and the donation case but the challenge is not just to find a difference. It is to find a difference that could underwrite a plausible moral principle. This means that the difference has to be one that is morally relevant. For instance, the fact that Albert's name starts with 'A' while Brad's starts with 'B' cannot help us here since that difference has no moral relevance.

One difference between the two cases is that the drowning child is very close to Albert whereas the African child is far away from Brad. In "Famine, Affluence, and Morality," Singer is dismissive of the idea that distance has the requisite moral relevance.

I do not think I need to say much in defense of the refusal to take proximity and distance into account. The fact that a person is physically near to us, so that we have personal contact with him, may make it more likely that we *shall* assist him, but this does not show that we *ought* to help him rather than another who happens to be far away. If we accept any principle of impartiality, universalizability, equality, or whatever, we cannot discriminate against someone merely because he

is far away from us (or we are far away from him)... There would seem, therefore, to be no justification for discriminating on geographical grounds (pp. 147-148).

Nonetheless, in recent work, F.M. Kamm has suggested that distance does in fact make a moral difference. It is important to note that Kamm's thesis is not just that distance is correlated with something else that has moral relevance but that distance *itself* makes a moral difference. There are, of course, cases where distance does not make a moral difference. Compare the following two cases

Near Costless: A child is drowning in a pond near Candace. Candace can save the child by pressing a button. She does nothing and the child drowns

Far Costless: A child is drowning in a pond in Manchester. David, who is in the United States, can press a button to save the child. He does not and the child drowns.

Intuitively, what Candace does is just as bad as what David does. However, Kamm notes that from the fact that distance does not make a difference in one case, it does not follow that distance never makes a difference in any case.¹⁵ Her view is that distance matters only when the costs of helping are significant.¹⁶ Nonetheless, Kamm recognizes that it is not immediately clear why the fact that the person in distress is far from the agent is not

¹⁵ F.M. Kamm, *Intricate Ethics* (Oxford: Oxford University Press, 2007), p. 348. Shelly Kagan makes the same point in "The Additive Fallacy," *Ethics* 90 (1988): 5-31.

¹⁶ This idea might be more made more precise in the following way. Suppose that we can rank the strength of an agent's duty to aid on a scale from 0 to 1, with 1 being very strong and 0 being no duty to aid at all. The formula for determining the strength of the agent's duty is $1 - (\text{cost})(\text{proximity to the agent})$. If the cost is 0, then regardless of distance the agent's duty to help will be 1. If the cost is very small then, distance only makes a small difference in the agent's duty to aid. However, where the cost is high, distance can make a large difference.

like the hair color of the person in distress, something we can dismiss as not morally relevant. Thus, she writes, “we cannot, I think, truly justify the moral relevance of distance in some contexts without a theory explaining why this factor should have relevance.”¹⁷

The theory that she thinks justifies the role distance plays in morality runs as follows. It is a hallmark of many non-consequentialist moral theories that there are agent-centered prerogatives that allow agents to favor their own projects rather than maximizing the good.¹⁸ Of course, prerogatives themselves do not give rise to any duties. They are, after all, prerogatives, not obligations. However, Kamm thinks that in acting on a prerogative an agent thereby generates “a duty to take care of the area near herself or her means.”¹⁹

The idea has some plausibility. Suppose that I am offered a position as the general manager of a chain of banks. If I accepted the position, I would be responsible for the overall good of the bank and I would have to treat each branch equally. But suppose I were to decline the position in favor of being the manager of a particular branch. I would then have responsibilities to take care of my specific branch even if, in a particular circumstance, I could do more good for a different branch.

The idea is that in acting on a prerogative I decline to maximize the overall good but instead focus my energies on projects that are of special concern to me. However, doing so generates a special duty to take care of the things in the area near me.

¹⁷Kamm, *Intricate Ethics*, p. 379.

¹⁸This idea was originally developed, albeit within a broadly consequentialist framework, by Samuel Scheffler in *The Rejection of Consequentialism* (Oxford: Oxford University Press, 1982).

¹⁹ Kamm, *Intricate Ethics*, p. 387.

Despite this argument's initial plausibility, I do not think that Kamm has established a connection between acting on agent-centered prerogatives and distance. Consider the bank case again. While choosing to be the manager of a particular branch does give rise to special duties, these duties need not covary with distance. Presumably, I would retain these duties to my specific branch even if, at the moment, I am physically closer to another branch.

Suppose I act on an agent-centered prerogative and decide to build a house in California. Plausibly this gives rise to special duties in the area of the house. Perhaps I now have a duty to participate in a neighborhood watch program or watch out for children playing in the area. But these special duties need not always cover the area around me. I may take a vacation to India. In such a case, it is hard to see how building a house in California gives rise to special duties to help the Indians in the area near me. Thus while it is plausible that acting on an agent-centered prerogative may give rise to special duties, it is not clear at all why these duties need to covary with distance. Thus, Kamm's argument fails to show that distance *itself* is morally relevant.

6.6 Slote on Empathy

Another difference between the pond case and the donation case is that, assuming Alfred and Brad are normal human beings, Alfred will feel a great deal of empathy for the drowning child while Brad will not feel much empathy for the African child. The psychological mechanisms at work here are not complex. Alfred can see the suffering of the drowning child while Brad has no such direct access to the suffering of the African child. Noting this difference between the cases, Michael Slote proposes that empathy can

form the basis of a plausible principle that can capture our judgment in the pond case and our considered moral judgments about aid.²⁰

Here is Slote's first statement of such a principle

Speaking very roughly, an ethics of caring holds that an act is morally (all) right if it doesn't exhibit a lack (or the opposite) of caring and wrong if it does.

(Brushing your teeth may not evince caring but the point is that it also doesn't evince, exhibit, or reflect a lack of caring concern about others.)²¹

This principle does capture our judgment about the pond case. Albert's walking past the pond surely exhibits a lack of caring about the child drowning in the pond. But what is not initially clear is why Brad's not donating \$50 to an aid organization does not show a lack of concern for the African child who will not receive medical treatment. Indeed, as Slote himself suggests, one plausible psychological explanation for why we do not donate to aid organizations is that we do not have the same sort of caring concern for children whose suffering has not been made vivid for us as we would for a child drowning right in front of us.²² If we did have the same sort of caring concern for those whose suffering was not vivid for us, we would help them as well. If this is the explanation for why we do not do more for the world's worst off, then we should conclude that Brad's action exhibits a lack of empathy as well.

²⁰ Michael Slote, "Famine, Affluence, and Empathy," in David Boonin and Graham Oddie (eds.), *What's Wrong? Applied Ethicists and Their Critics* (Oxford: Oxford University Press, 2005), pp. 548-556.

²¹ Ibid, p. 548

²² Ibid, p. 551.

Slote did say that he was “speaking very roughly” in his first statement of the principle. He then offers another formulation of the principle that is not equivalent to the initial formulation.

One can say that actions are wrong or right depending on whether or not they reflect or exhibit a deficiency of normally or fully empathic caring motivation.²³

Slote thinks that this principle can explain why, other things being equal, it would be morally worse to prefer a fetus or embryo to a fully developed human being, because such a preference runs counter to the flow of developed human empathy or caring motivation that is shaped by such empathy. And similar points, arguably could be made about our moral relations with the lower animals.²⁴

This principle differs from the earlier one in that it makes reference to “*normally* or fully empathic caring motivation” and the “flow of developed human empathy.” One could argue that preferring a fetus to a fully developed human does exhibit less caring concern for the fetus than the fully developed human but since such a difference of concern is not counter to normal fully developed empathic caring motivation, this difference of concern need not be impermissible according to Slote’s second principle. Thus, in light of what Slote says about preferring a fully developed human to a fetus, a natural interpretation of the second principle is that the wrongness of an action is determined by whether or not it runs counter to what a human being with normal, fully developed empathic motivation would do. Preferring a fetus to a fully developed human being is indeed not what a normal human with a developed sense of empathy would do. The same goes for Alfred’s

²³ Ibid, p. 550.

²⁴ Ibid.

actions in the pond case, and thus these actions would, according to Slote's principle, be wrong. Brad's action, in contrast, is not counter to normally developed human empathy. It is the sort of thing that normal human beings with normal capacities for empathy do regularly, buying consumer goods with money that could be used to save lives.

Slote's principle appears to be our holy grail; it captures our judgment about the pond case without sacrificing our considered moral judgments about aid. Nonetheless, I will argue, the principle is not plausible. There are many situations in which human beings with normally developed empathic caring motivation do things that are uncontroversially wrong. Consider the Milgram experiments, where participants were told by an actor posing as a scientist to shock what they believed was another participant in the experiment (actually a confederate who was not being shocked). As the experiment went on, participants delivered higher and higher voltage electric shocks as they heard what they believed was another participant howling in pain. 65% of the participants turned a dial that they believed would deliver a 450-volt shock to another human being.²⁵ Or consider the Stanford prison experiment in which participants were randomly chosen to be either guards or prisoners and placed in a mock prison. The guards quickly turned sadistic, refusing to empty the sanitation buckets of prisoners, removing their mattresses so the prisoners had to sleep on concrete, and forcing the prisoners to go without clothing.²⁶ While the Milgram and Stanford experiments were controlled experiments, they have important parallels to real life events. Milgram's study was intended to help us understand the behavior of Nazi guards at concentration camps

²⁵ Stanley Milgram, *Obedience to Authority: An Experimental View* (New York: Harper Collins, 1974).

²⁶ The details of the experiment can be found at <http://www.prisonexp.org>.

and there are close parallels between the Stanford experiment and the events at Abu Ghraib. Optimists might hope that the participants in the Milgram and Stanford studies, as well as their real-life analogues, were all somehow different from normal human beings but as the experiments have been replicated and as more historical examples can be adduced, a more plausible explanation is that normal people with normally developed capacities for empathy will do very bad things in certain situations. Thus, the rightness or wrongness of an action does not depend on whether or not the action reflects a deficiency of normally empathic caring motivation.

Earlier we noted that Slote's view is that whether or not an action is right or wrong depends on whether or not it reflects "a deficiency of normally or *fully* empathic caring motivation." Thus far, I have been interpreting Slote's use of 'fully' to mean fully developed. Fully developed is not the same as normal. A two-year-old child may have normal empathic caring motivation without having fully developed empathic caring motivation. Nonetheless, this is not the only interpretation one can give of the phrase 'fully empathic caring motivation.' 'Fully' can also mean 'completely' as opposed to 'deficiently' or 'partially.' One might argue that, while the participants in the experiments described above possessed normal empathic caring motivation, they were not fully empathic. Instead their empathy was rendered partial or deficient. Thus, the experiments above do not constitute counter-examples to Slote's principle. Instead, they show us that normal human beings with normal and fully developed empathic caring motivation sometimes fail to be fully empathic.

If we interpret Slote's use of 'fully' this way we can avoid the problem raised by the Milgram and Stanford experiments. However, this interpretation renders Slote's second principle vulnerable to the same objection we raised against his first principle. For it now open to us to argue that our attitudes toward the worst off represent a case where human beings with normal empathic caring motivation fail to be fully empathic. As we saw above it is a plausible claim about our psychology that the fact that the suffering of those far away is less salient for us weakens our empathy for those persons. Why should we not conclude from this that our empathic caring for those who are far away is only partial? It is, after all, weakened by the fact that the suffering of those far from us is not salient.

The mechanism that allowed Slote's principle to capture our judgment about the pond case without giving up our considered moral judgments about aid is that it made reference to normal or fully developed empathic caring motivation. Once we allow that even actions with normal or fully developed empathic caring motivation can be wrong, we open up the possibility that our actions toward the world's worst off, though normal, are wrong. Slote's principle faces a dilemma. If we interpret his use of 'fully' to mean 'fully developed' we can capture our judgment in the pond case without sacrificing our considered moral judgments about aid. However, we are then faced other counter-examples. If we interpreted 'fully' to mean 'complete,' then we can avoid counter-examples but it is difficult to see how such a view can vindicate our considered moral judgments about aid.

6.7 Giving up the Pond

I have not considered every principle that could potentially capture our judgment about the pond case without conflicting with our considered moral judgments about aid. Nonetheless, I have shown that several prominent attempts fail and we should now be more pessimistic about the possibility of finding such a principle.

Now is a good time to return to our re-interpretation of Singer's original argument as an application of the method of reflective equilibrium to see if there are other strategies for resisting. The argument begins by characterizing our initial considered moral judgments. We then attempt to find a principle (or set of principles) that systematizes those judgments. In the absence of such a principle we must give up one of the judgments. Singer would urge us to give up our considered moral judgments about aid. Perhaps this is the place to object. We can concede to Singer that there is no plausible principle that captures both judgments but hold that he has chosen the wrong judgments to reject. Instead, we should hold that rescuing the drowning child is not required but instead supererogatory.

There are two things to note about this strategy. First, even if it works, the main point of this chapter, that the method of reflective equilibrium can, in practice, compel us to reject a deeply held considered moral judgment or a part of status quo morality, is still vindicated. It just turns out that the judgment that we should reject is not the one that Singer urges us to.

However, I do not want to rest here because the contributionist account of belief revision defended in chapter four gives us a means to adjudicate the debate between one who thinks that we ought to reject our considered moral judgments about aid and one who thinks we ought to reject our judgment about the pond case. What we need to do is

look closely at the evidential and inferential relationships that our pond judgment and our considered moral judgments about aid have to the other things we believe and attempt to discover which set of judgments makes the greater overall contribution to the comprehensiveness and connectedness of our overall set of beliefs.

It is important to note that on the contributionist account of belief revision, which judgment we should give up is determined solely by which judgment makes a greater contribution to the overall connectedness and comprehensiveness of the agent's set of beliefs. The contributionist theory does not take into account the agent's preferences with regard to belief revision. Thus, even agents who would rather reject the pond judgment may not be justified in doing so if the pond judgment makes a large enough contribution to the comprehensiveness and connectedness of the agent's set of beliefs.

At first, it may seem surprising that our judgment about the pond case could have stronger evidential and inferential relationships to the other things we believe than our considered moral judgments about aid. After all, the pond judgment is a judgment about a single case.

It is important to emphasize, however, that there is nothing special about the pond case. We can generate an infinite number of cases just like it, in which we are face to face with someone who we could easily save at little cost to ourselves. Unless there is some other morally relevant factor present, we almost always judge that it would be wrong not to save the person. Motivating our judgment in the pond case is the common-sense idea that we have a duty of easy rescue, which is a genuine duty in the sense that it gives rise to a moral requirement. This is what gets the Singer problem off the ground. Intuitively, we have a duty of easy rescue but we live in a world where we are faced with

millions of cases of easy rescue. One can either reject the duty of easy rescue or accept that morality requires that we dedicate many of our resources to rescue. Preserving our considered moral judgments about aid requires that we reject a duty of easy rescue.

There may indeed be principled reasons for doing so. Jan Narveson suggests that we are only required to help others if we consented to help (by formal contract or perhaps just informal promise) or are responsible for their predicament.²⁷ This is a principle that would indeed justify rejecting the duty of easy rescue but it would also involve rejecting the idea that we have any duty of beneficence at all.²⁸

Perhaps, it is not so surprising that rejecting a duty of easy rescue involves rejecting the notion of a duty of beneficence.²⁹ It would be an odd position to hold that we have duties to help others but no duty to save their lives when we could do so easily.

²⁷ Jan Narveson, "We Don't Owe Them a Thing! A Tough-minded but Soft-hearted View of Aid to the Faraway Needy," *The Monist* 83 (2003): 419-33.

²⁸ Narveson distinguishes between two senses of 'duty'. In the strong sense, in which a duty to X implies that one is required to X, we have no duty to help those we have not contracted with or harmed. There is a weaker sense of 'duty' in which to call an action a duty is to say that the actions proceed from morally commendable motives. Narveson thinks that there is a duty in this weaker sense to help those whom we have not contracted with or harmed. See Narveson, "We Don't Owe them a Thing!" p. 419. In this chapter I exclusively use 'duty' in the strong sense. This seems to me to comport better with ordinary usage.

²⁹ It might be thought that the weak imperfect duty account rejects a duty of easy rescue without giving up wholesale the idea of a duty of beneficence. I think it is more plausible to see this view as preserving a kind of weak duty both of easy rescue and beneficence. On the weak imperfect duty account there are genuine duties of easy rescue and beneficence but these duties do not require one who has already done a certain amount of good for others to continue aiding or rescuing. As I argued above, this view preserves our judgment in the standard pond case, though it does give up our judgment in the Good Samaritan pond case. Thus, one may be able to give up our judgment in the Good Samaritan pond case while maintaining that we have a duty of beneficence but I am not aware of any position that would allow us to maintain a duty of beneficence and give up our judgment in the standard pond case.

The view that we have no duty of beneficence is consistent but it involves radical revision to our moral conceptual scheme, a revision that is perhaps more radical than the one Singer suggests.

This is an example of how we might use the notions of comprehensiveness and connectedness, in practice, to settle a moral controversy. We begin with a case judgment and when we find that the case judgment conflicts with other judgments, we consider the inferential and evidential relationships that between these judgments and our other beliefs. If we find that we cannot give up the case judgment without giving up other beliefs that play a central role in our web of belief, then there are strong grounds to hold onto the case judgment. Here we find that our judgment in the pond case is supported by the belief that we have a duty of easy rescue and a general duty of beneficence. Since these beliefs play a central role in our web of beliefs, there are strong grounds for thinking that they should not be given up.

To show definitively that they should not be given up we would have to show not only that our judgment in the pond case makes a large contribution to the comprehensiveness and connectedness of our overall set of beliefs but that it makes a larger contribution than our considered moral judgments about aid. While I cannot complete this project here, I do want to note that it is much harder to find considerations at the center of our web of belief that support our considered moral judgments about aid. Most of us do have the belief that complying with the demands of morality leaves room for an individual to pursue the projects that she identifies with.³⁰ I will return to this idea

³⁰ This needs to be qualified. Most of us think that morality could require an agent to fight in a war of self-defense even if this means sacrificing goods necessary for the pursuit of projects one identifies with or even perhaps even if doing so requires

when I discuss Richard Miller's view in section 6.9. To preview the results from that section, this idea may support the belief that the duty of beneficence demands less than Singer thinks it does but it does not vindicate our considered moral judgment that we are permitted to do little or nothing for the world's worst off.

I think, then, that there is a strong *prima facie* case for thinking that our judgment about the pond case makes a greater contribution to the connectedness and comprehensiveness of our overall set of beliefs than our considered moral judgments about aid. If this is true, then the theory of justification defended in chapter four requires that agents abandon their considered moral judgments about aid even if they would prefer not to.

6.8 Concession and Mitigation

Understanding Singer's argument as an application of the method of reflective equilibrium involves the following steps.

- 1) Characterizing our initial considered moral judgments
- 2) Finding a principle that systematizes these judgments
- 3) Observing that any plausible, well-motivated principle that systematizes our judgment about the pond case conflicts with our considered moral judgments about aid
- 4) Realizing that faced with this conflict we should give up our considered moral judgments about aid.

The most obvious places to resist are 3) and 4). However, I have argued in previous sections that resisting at either of these points is not easy.

sacrificing one's life. Most of us think that morality demands paying back one's debts even if doing so requires sacrificing goods necessary for the pursuit of projects one identifies with. There are many more examples of this sort.

Before concluding, let us consider a different type of strategy, which I will refer to as ‘concession and mitigation.’ This strategy involves conceding that we do, as Singer suggests, have a duty to make sacrifices for the world’s worst off. However, proponents of this strategy then argue that this duty is mitigated in certain ways and thus does not require as much of us as Singer thinks. I will note below, that all versions of this strategy force us to give up the idea that we are required to donate nothing or very little to aid organizations. Thus, they all involve rejecting our considered moral judgments about aid as I characterized them above.

6.9 The Limits of Beneficence

One version of the concession and mitigation strategy is to insist that while we do have a duty to donate to aid organizations, there are certain things that this duty cannot require us to give up. Thus, this strategy aims to put a limit on the requirement of beneficence. The arguments that proponents of this strategy use to justify the idea that beneficence has limits are complex. Rather than discussing them in detail here what I will do is simply note that no version of this strategy yet proposed vindicates our considered moral judgments about aid. Instead, they all require us to make significant sacrifices for the least well off, though the sacrifices these authors advocate are not as extreme as what Singer advocates.

An example of this strategy is found in Garrett Cullity’s *The Moral Demand of Affluence*. Cullity defends the following principle.

If it is absurd to deny that your pursuing or having x can ground requirements on others to help you, then your pursuing or having x violates no requirement of beneficence.³¹

Certain goods, including what Cullity calls life-enhancing goods, are important enough that it is absurd to deny that they ground requirements on others to help you. Life-enhancing goods include friendship and the pursuit of certain intrinsically worthwhile projects.³² If this is right, then it follows from Cullity's principle that one may pursue friendships and certain intrinsically worthwhile projects without violating any requirement of beneficence.

Nonetheless, Cullity's principle is unlikely to help us to preserve our considered moral judgments about aid. While it may be absurd to deny that others are required to sacrifice to save my life, it is not absurd to deny that others are required to sacrifice so that I may purchase a \$50 pair of shoes solely because I find them fashionable. As Cullity acknowledges, "the conclusion I defend does still demand more of us than many of us find comfortable."³³ For instance, Cullity concludes that, "buying expensive clothes or furniture, a new car (or often, any car at all), or books for a private library is usually morally wrong, as the world now stands."³⁴

³¹ Cullity, *The Moral Demands of Affluence*, p. 149.

³² Ibid, p. 129 and 161-5.

³³ Ibid, p. 3.

³⁴ Ibid, p. 183.

A more permissive version of the same strategy is pursued by Richard Miller.³⁵ Miller's view is that the duty of beneficence cannot deprive one of the resources necessary to pursue "a goal with which someone identifies."³⁶ These goals are the sort of thing that is a "constituent of someone's personality" or "part of her description of "the sort of person that I am."³⁷ Miller's view is more permissive than Cullity's because one who identifies with displaying a certain kind of aesthetic sense is permitted to "occasionally purchase a luxury or frill, namely, some stylish clothing rather than pursue a less expensive, plain alternative," provided that doing so is necessary to pursue her goal of displaying a certain aesthetic sense.³⁸

Miller's view provides a partial vindication of our judgment in the donation case. If Brad is the sort of person who identifies with the goal of displaying a certain fashion sense, then he may occasionally purchase a frill such as a pair of shoes that he does not require for his livelihood. This is only a partial vindication because, I take it, most people have the considered judgment that what Brad does is permissible even if displaying a certain aesthetic sense is not a goal with which Brad identifies. However, Miller's view only vindicates Brad's action if he has such a goal. Furthermore, most people have the considered judgment that Brad is not acting wrongly if he purchases more shoes than are *necessary* for pursuing his goal. However, Miller's view only allows

³⁵ Richard Miller, "Beneficence, Duty, and Distance," *Philosophy & Public Affairs* 32 (2004): 357-383.

³⁶ Ibid, pp. 360-1.

³⁷ Ibid.

³⁸ Ibid, p. 361.

one to make the sort of occasional purchases that are necessary for pursuing the goals that they identify with.

Moreover, Miller's view is unable to accommodate our general judgments that we are permitted to give little or nothing to relief organizations. Miller recognizes this, noting that his view,

still requires significant giving from most of the nonpoor. The underlying goals to which most of us who are not poor are securely attached leave room for this giving: we could pursue these goals enjoyably and well and fulfill our other responsibilities, while giving significant amounts to the needy.³⁹

We have seen then that two versions of this strategy fail to vindicate our considered moral judgments about aid. This is not surprising. Indeed to use this strategy to vindicate our judgment that one is required to donate little or nothing to aid organizations one would have to hold that beneficence can only demand little or nothing from us. But, aside from the difficulties of defending such a position, the intuitive cost is high. For there are many cases, including the pond case, where we intuitively believe that beneficence can require us to make sacrifices.

6.10 Fair Share Views

One reason why Singer's principle gives rise to such stringent demands in our world is that so few people comply with it. If everyone complied with the principle, it would demand much less of each individual. As Singer notes, to infer from this that we must only comply with the lesser demand is to fallaciously affirm the consequent (p.

³⁹ Ibid, p. 362.

148). However, there is something appealing about the idea that it is unfair to require us to pick up the slack from others failing to comply with their duty.

Liam Murphy has attempted to defend this appealing idea by introducing a compliance condition, which he spells out as follows: “very roughly, the idea is that the demands on a complying person should not exceed what they would be under full compliance with the principle.”⁴⁰

How much Murphy’s condition would limit the demands of beneficence depends on some complicated moral and empirical matters. On the empirical side, we would have to know what it would cost to prevent all the suffering and death from lack of food and medical care that can be prevented. On the moral side, we would have to know what would be a fair way to distribute this cost. Simply dividing it evenly would presumably result in a large burden to the least rich persons in industrialized countries while only requiring trivial sacrifices from millionaires. Plausibly, the fairest distribution of these burdens would be a progressive system. On his website, Singer proposes such a system asking between 1% (for those who make less than US \$105,00) and 33% of one’s income (for those making more than US \$10,700,00).⁴¹

If Singer’s numbers are a fair distribution of the cost of preventing suffering and death from lack of food and medical care, then Murphy’s view requires less from us than Singer’s original demand but more than most people intuitively think they are required to donate. In other words, Murphy’s view does not preserve our considered moral judgments about aid but comes much closer than Singer’s view.

⁴⁰ Liam Murphy, *Moral Demands in Nonideal Theory* (Oxford: Oxford University Press, 2000), p. 7.

⁴¹ <http://www.thelifeyoucansave.com/pledge>

Murphy's view, however, has other sorts of intuitive costs. Suppose one has already sacrificed as much as they would be required to under full compliance. Then, on Murphy's view, one is free to walk to past the pond without helping the drowning child. Indeed, a poor person who is already worse off than she would be under full compliance is not required to do anything at all, even wade into a pond to save a drowning child.

Murphy acknowledges these intuitive costs but insists that we must give up some of our intuitions about these rescue cases because "any limited principle of beneficence designed to match common sense about rescue cases will turn out to be *very* much more demanding than common sense when it comes to general beneficence."⁴² In other words, if we attempt to match common sense about pond cases, we will end up with a principle that is extremely demanding in the actual world. Murphy insists that his view has more *overall* plausibility than its rivals.⁴³ Since no principle can capture all of our pond judgments and all of our judgments about aid, we must make some revisions to these judgments and Murphy asks for less radical revisions than rival views.

For our purposes, we can leave open the question of whether or not Murphy's view is more plausible than all its rivals. Murphy's view does not capture our considered moral judgments about aid or all of our judgments about pond cases but it does ask for less radical revisions to these judgments than some rivals. Here we need only note that preserving our considered moral judgments about aid *and* rescue cases as they now stand is not an option under Murphy's view.

⁴² Murphy, *Moral Demands in Nonideal Theory*, p. 129. Emphasis in original.

⁴³ *Ibid*, pp. 133-4.

6.11 Conclusion

In this chapter I have aimed to rebut a criticism of the method of reflective equilibrium according to which the method cannot require us to revise our deeply held moral intuitions and serves merely to reinforce the moral status quo. I have argued that one argument with an extremely counter-intuitive conclusion can be interpreted as an application of the method of reflective equilibrium.

I have also considered some of the most well known responses to this argument. In many cases, I have not taken a stand on whether these responses succeed but instead I have noted that even if these responses do succeed, they do not vindicate our considered moral judgments at the outset of inquiry. Those rival principles to Singer's that are well motivated and plausible fail to vindicate our considered moral judgments about aid. One who responds to Singer's argument by denying the pond judgment *ipso facto* rejects a deeply held considered moral judgment. Mitigating strategies may lessen our obligations of beneficence but they do not fully vindicate our considered moral judgments about aid and in Murphy's case entail rejecting other judgments about variants of the pond case.

From these considerations, I draw the conclusion that it is plausible, perhaps even likely, that the method of reflective equilibrium will compel us to revise deeply held moral intuitions and status quo morality. Therefore, the objection that the method cannot require such revisions can be put to rest.

Conclusion

In order to be plausible, a theory of justification must meet its own epistemic standards. For the theory of justification defended here, this means that the theory must cohere with our most central judgments about the nature of moral inquiry. Throughout this dissertation, I have argued that the method of reflective equilibrium when combined with the contributionist account of belief revision scores quite well on this measure. It acknowledges a role for intuitions in moral inquiry but it also acknowledges that these judgments must be subject to intense critical scrutiny. It acknowledges the Quinean insight that, when deciding what to believe, we have no choice but to start with what we already believe while at the same time avoiding the vice of being unduly conservative. The theory is compatible with a range of opinions on controversial matters in moral theory. For instance, it is compatible with both the view that there are and the view that there are not moral questions with no answer. It is also compatible with both the view that there are inescapable moral dilemmas and the view there are no such dilemmas.

If we acknowledge the virtues of the theory defended here, important questions remain within moral epistemology as well as potential applications to normative ethics and moral psychology. In this conclusion, I mention some of these questions as possible directions for future research

Moral Epistemology

Throughout this dissertation, I have treated moral knowledge as a special case of propositional knowledge. To know the wrongness of murder is to know the proposition that murder is wrong. There is, however, a tradition in moral philosophy that sees moral knowledge not merely as a kind of propositional knowledge but as a kind of know-how.

This distinction between propositional knowledge, or knowledge that, and knowledge how can be illustrated by example. An expert bike rider knows, in some sense of ‘know,’ how much to slow down when a tight turn is coming up but she may not be able to articulate this knowledge. In other words, while she may know precisely how much to slow down, she may not know the proposition, ‘a bike rider should slow down x amount when taking a tight turn.’ Her knowledge is best described as a sort of knowledge how rather than knowledge that.

It is not implausible to think of moral knowledge on the same model. A virtuous person may know, in some sense of ‘know,’ when (if ever) it is appropriate to lie to a friend to spare them pain. But she may not be able to articulate precisely how much pain would have to be spared in order to make telling the lie permissible. She may have a kind of knowledge how rather than knowledge that.

This raises a series of important questions for future research. First, is it true that moral knowledge is better thought of a sort of know how rather than knowledge that? Can the theory of justification explicated here accommodate that insight? If so, how? If not, how might the theory be revised?

A separate set of questions in moral epistemology concerns the role of testimony in moral judgment. In other areas of human thought, testimony plays a crucial role in deciding what to believe. Is there a comparable role for testimony in moral judgment? Can the theory explicated here accommodate a role for testimony? These questions become especially interesting when we receive testimony that contradicts our intuitions. How to weigh this sort of testimony against our considered moral judgments?

Finally, a third set of questions in moral epistemology concerns the plausibility of moral skepticism. In chapter 5, I argued that the theory of justification defended here leaves open the possibility that moral skepticism is true. Future research might discuss the extent to which skeptical challenges emerge within reflective equilibrium and whether or not these challenges can be defeated.

Moral Psychology

As we saw in chapter 2, questions remain about how our intuitions are produced. The answers to these questions have a direct bearing on how intuitions ought to be used in the method of reflective equilibrium. If we find that a set of intuitions is produced by an untrustworthy cognitive process, this may serve as a background theory in the method of reflective equilibrium and this background theory may, in turn, count against the intuition when it conflicts with other judgments. We have seen that there are cases in which our intuitions are the product of prospect theory biases and ordering effects. Which other intuitions might be the product of untrustworthy cognitive processes? This question raises further questions. Which cognitive processes count as untrustworthy? Recently, several thinkers have proposed that certain intuitions are the emotional responses and that emotional responses are untrustworthy.⁴⁴ Is such skepticism about the value of emotions warranted? To date, I am aware of no plausible argument that suggests an affirmative answer to this question.

⁴⁴ Greene, “The Secret Joke of Kant’s Soul;” Singer, “Ethics and Intuitions;” and Walter-Sinnott Armstrong, Liane Young, & Fiery Cushman, “Moral Intuitions as Hurestics,” (forthcoming). For a response to this sort of argument see Berker, “The Normative Insignificance of Neuroscience.”

Normative Ethics

Perhaps most importantly, the theory of justification defended here gives us a framework for thinking about what morality requires of us. There are debates in normative and practical ethics that persist after years. One hypothesis is that the participants in these debates rely on conflicting intuitions. Supposing this hypothesis is correct, the theory of justification here suggests that conflicting intuitions need not lead to argumentative deadlock. The contributionist account of belief revisions suggests a way of deciding between conflicting intuitions. I am cautiously optimistic that this account could break up some of the deadlock we find in normative and practical ethics and perhaps provide us with new ways to resolve old problems.

Bibliography

- Appiah, K. *Experiments in Ethics* (Cambridge, MA: Harvard University Press, 2008).
- Aristotle, *Posterior Analytics*.
- Asch, S.E. "The Doctrine of Suggestion, Prestige, and Imitation in Psychology," *Psychological Review* 55 (1948): 250-76.
- Ashford, E. "The Demandingness of Scanlon's Contractualism," *Ethics* 113 (2003): 272-203.
- Audi, R. *The Structure of Justification* (Cambridge, UK: Cambridge University Press, 1993).
- _____, "Intuitionism, Pluralism, and the Foundations of Ethics," in Walter Sinnott-Armstrong and Mark Timmons (eds.), *Moral Knowledge*, pp. 101-36.
- _____, "Moderate Intuitionism and the Epistemology of Moral Judgment," *Ethical Theory and Moral Practice* 1 (1998): 15-44.
- _____, *Epistemology: A Contemporary Introduction* 2nd ed. (London: Routledge, 1998).
- _____, *The Right in the Good: A Theory of Intuition and Intrinsic Value* (Princeton, NJ: Princeton University Press, 2004).
- Aune, B. *Knowledge, Mind, and Nature* (New York: Random House, 1967).
- Baron, M. *Kantian Ethics Almost Without Apology* (Ithaca, NY: Cornell University Press, 1995).
- Berker, S. "The Normative Insignificance of Neuroscience," *Philosophy and Public Affairs* 37 (2009): 293-329.
- BonJour, L. *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1985).
- Brandt, R.B. *A Theory of the Good and the Right* (Oxford: Oxford University Press, 1979).

- Brink, D. *Moral Realism and The Foundations of Ethics*, (Cambridge, UK: Cambridge University Press, 1989).
- _____. "Moral Conflict and Its Structure," *Philosophical Review* 103 (1994): 215-247.
- Cacioppo, J.T. and Petty, R.E. "The Need for Cognition," *Journal of Personality and Social Psychology* 42 (1982): 116-131.
- Christensen, D. "Epistemology of Disagreement: The Good News," *Philosophical Review* 116: 187-217.
- Copp, D. "Considered Moral Judgments and Moral Justification: Conservatism in Moral Theory," in David Copp and Michael Zimmerman (eds.), *Morality, Reason, and Truth* (Totowa, NJ: Rowman and Allanheld, 1984).
- _____, *Morality in a Natural World* (Cambridge, UK: Cambridge University Press, 2007).
- Crisp, R. "Sidgwick and the Boundaries of Intuitionism" in Stratton-Lake (ed.), *Ethical Intuitionism*, pp. 56-75.
- Cullity, G. *The Moral Demands of Affluence* (Oxford: Oxford University Press, 2004).
- Cummins, R. "Reflections on Reflective Equilibrium," in Michael DePaul and William Ramsey (eds.), *Rethinking Intuitions* (Oxford: Rowman & Littlefield, 1998), pp. 113-28.
- Cummiskey, D. *Kantian Consequentialism* (Oxford: Oxford University Press 1996).
- Dancy, J. *Moral Reasons* (Oxford: Blackwell, 1993).
- Daniels, N. "Wide Reflective Equilibrium and Theory Acceptance in Ethics," *Journal of Philosophy* 76 (1979): 256-282. Reprinted in *Justice and Justification*, pp. 21-46.
- _____, *Justice and Justification* (Cambridge, UK: Cambridge University Press, 1996).

- Cuneo, T. "Saying What we Mean: An Argument against Expressivism" in Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics* vol. 1 (Oxford: Oxford University Press, 2006), pp. 60-65.
- DePaul, M. *Balance and Refinement: Beyond Coherence Methods of Moral Inquiry* (London: Routledge, 1993).
- Descartes, R. *Rules for the Direction of our Native Intelligence*
_____, *Meditations on First Philosophy*.
_____, *The Philosophical Writings of Descartes* (trans. by John Cottingham, Robert Stoothoff, and Dugald Murdoch) (Cambridge, UK: Cambridge University Press, 1985).
- Drier, J. "Humean Doubts about the Practical Justification of Morality," in Cullity and Gaut (eds.), *Ethics and Practical Reason*.
- Evans, J. "Matching Bias in Conditional Reasoning: Do We Understand it After 25 Years?" *Thinking and Reasoning* 4 (1998): 45-82.
- Feldman, R. "Epistemological Puzzles about Disagreement," in Stephen Hetherington (ed.), *Epistemology Futures* (Oxford: Oxford University Press, 2006), pp. 216-236.
- Foot, P. "The Problem of Abortion and the Doctrine of Double Effect," *Oxford Review* 5 (1967): 5-15.
- Fumerton, R. *Metaepistemology and Skepticism* (Lanham, MD: Rowman & Littlefield, 1995).
- Gauthier, D. *Morals by Agreement* (Oxford: Clarendon Press, 1986).
- Gendler, T. "Philosophical Thought Experiments, Intuitions, and Cognitive Equilibrium," *Midwest Studies in Philosophy* 31 (2007): 68-89.
- Gert, B. *Morality: Its Nature and Justification* Revised edition (Oxford: Oxford University Press, 2006).

- Gigerenzer, G. *Gut Feelings: The Intelligence of the Unconscious* (New York: Penguin, 2007).
- Gilovich, T. Griffin, D. and Kahneman, D. (eds.) *Heuristics and Biases: The Psychology of Intuitive Judgment* (Cambridge, UK: Cambridge University Press, 2002).
- Gould, S.J. *Bully for Brontosaurus: Reflections on Natural History* (New York: Norton, 1991).
- Greene, J. "The Secret Joke of Kant's Soul," in Walter Sinnott-Armstrong (ed.), *Moral Psychology Volume 3: The Neuroscience of Morality* (Cambridge, MA: MIT Press, 2008).
- _____ and Haidt, J. "How (and Where) Does Moral Judgment Work?" *Trends in Cognitive Sciences* 6 (2002): 517-33.
- _____, Sommerville, R. Nystrom, L.E. Darley, J.M., and Cohen, J.D. "An fMRI Investigation of Emotional Engagement in Moral Judgment." *Science* 293 (2001): 2105-2108.
- Haack, S. *Evidence and Enquiry: Towards Reconstruction in Epistemology* (Oxford: Blackwell, 1993).
- Haidt, J. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment" *Psychological Review* 108 (2001): 814-34.
- _____ and Baron, J. "Social Roles and the Moral Judgment of Acts and Omissions," *European Journal of Social Psychology* 26 (1996): 201-218.
- _____, Koller, S. & Dias, M. "Affect, Culture and Morality, or Is It Wrong to Eat Your Dog?," *Journal of Personality and Social Psychology* 65 (1993): 613-28
- Hare, R.M. *The Language of Morals* (Oxford: Oxford University Press, 1952).
- _____, *Freedom and Reason* (Oxford: Oxford University Press, 1963).
- _____, "The Argument from Received Opinion" in *Essays on Philosophical Method* (Berkeley: University of California Press, 1972), pp. 117-135.

- _____, "Rawls Theory of Justice" *Philosophical Quarterly* 23 (1973): 144-55. Reprinted in Norman Daniels (ed.), *Reading Rawls: Critical Studies of a Theory of Justice* (New York: Free Press, 1975), pp. 81-107.
- _____, *Moral Thinking: Its Levels, Method and Point* (Oxford: Oxford University Press, 1981).
- Harman, G. *Thought* (Princeton, NJ: Princeton University Press, 1973).
- _____, *The Nature of Morality* (Oxford: Oxford University Press, 1977).
- Huemer, M. *Ethical Intuitionism* (New York: Palgrave, 2005).
- Herman, B. *The Practice of Moral Judgment*, (Cambridge, MA: Harvard University Press, 1993).
- Hill, T. *Dignity and Practical Reason in Kant's Moral Theory* (Ithaca, NY: Cornell University Press, 1992).
- Hooker, B. *Ideal Code, Real World* (Oxford: Oxford University Press, 2000).
- Horowitz, T. "Philosophical Intuitions and Psychological Theory," *Ethics* 108 (1998): 367-385.
- Hume, D. *A Treatise of Human Nature*.
- Jamieson, D. "Method and Moral Theory" in Peter Singer (ed.), *A Companion to Ethics* (Oxford: Blackwell, 1991).
- Joyce, R. *The Myth of Morality* (Cambridge, UK: Cambridge University Press, 2001).
- Kagan, S. "The Additive Fallacy," *Ethics* 90 (1988): 5-31.
- _____, "Thinking About Cases," *Social Philosophy & Policy* 18 (2001): 44-63.
- Kahneman, D. "Maps of Bounded Rationality: A Perspective on Intuitive Judgment and Choice," in Tore Frangsmyr (ed.) *Nobel Prizes 2002* (Almqvist & Wiksell International, 2003), pp. 449-489.

- Kahneman, D. Slovic, P. and Tversky, A. (eds.) *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge, UK: Cambridge University Press, 1982).
- Kahneman, D. and Tversky, A. "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* 47 (1979): 263-91.
- _____, "Choices, Values, and Frames," *American Psychologist* 39 (1984): 341-350.
- ____ (eds.), *Choices, Values, and Frames* (Cambridge, UK: Cambridge University Press, 2000).
- Kant, I. *Groundwork of the Metaphysics of Morals*.
- _____, *The Metaphysics of Morals*.
- Kamm, F.M. "Moral Intuitions, Cognitive Psychology, and the Harming-Versus-Not-Aiding Distinction," *Ethics* 108 (1998): 463-488.
- _____, *Morality, Mortality* Vol. 1 (Oxford: Oxford University Press, 1998).
- _____, *Intricate Ethics* (Oxford: Oxford University Press, 2007).
- Kelly, T. "The Epistemic Significance of Disagreement," in Tamar Gendler and John Hawthorne (eds.), *Oxford Studies in Epistemology* (Oxford: Oxford University Press, 2006), pp. 167-196.
- Klein, P. "Foundationalism and the Infinite Regress of Reasons," *Philosophy and Phenomenological Research* 58 (1998): 919-26.
- Kohlberg, L. *The Philosophy of Moral Development* Vol. 1 (San Francisco: Harper & Row, 1981).
- Kornblith, H. (ed.), *Epistemology: Internalism and Externalism* (Oxford: Blackwell, 2001).
- Kuhberger, A. "The Influence of Framing on Risky Decisions," *Organizational Behavior and Human Decision Processes* 75 (1998): 23-55.

- LeBoeuf, R. and Shafir, E. "Deep Thoughts and Shallow Frames: On the Susceptibility to Framing Effects," *Journal of Behavioral Decision Making* 16 (2003): 77-92.
- Lehrer, K. *Theory of Knowledge* (Boulder, CO: Westview Press, 2000).
- Levin, I.P. Gaeth, G.J., Schreiber, J. and Lauriola, M. "A New Look at Framing Effects: Distribution of Effect Sizes, Individual Differences, and Independence of Types of Effects," *Organizational Behavior and Human Decision Processes* 88 (2002): 411-29.
- MacIntyre, A.C. "Hume on 'Is' and 'Ought,'" *Philosophical Review* 68 (1959): 451-68.
- Mackie, J.L. *Ethics: Inventing Right and Wrong* (New York: Penguin, 1977).
- Marcus, R.B. "Moral Dilemmas and Consistency," *Journal of Philosophy* 77 (1980): 121-136.
- McGrath, S. "Moral Disagreement and Moral Expertise" in Russ Shafer-Landau (ed.) *Oxford Studies in Metaethics* Vol. 3 (Oxford: Oxford University Press, 2008), pp. 87-108.
- McGrew, T., "A Defense of Classical Foundationalism" in Louis Pojman (ed.), *Theory of Knowledge: Classical and Contemporary Readings* (Belmont, CA: Wadsworth, 1998).
- McMahan, J. "Moral Intuition," in Hugh LaFollette (ed.), *Blackwell Guide to Ethical Theory* (Oxford: Blackwell, 2000), pp. 92-101.
- Milgram, S. *Obedience to Authority: An Experimental View* (New York: Harper Collins, 1974).
- Miller, P.M. and Fagley, N.S. "The effects of framing, problem variations, and providing rationale on choices," *Personality and Social Psychology Bulletin* 17 (1991): 517-22.
- Miller, R. "Beneficence, Duty, and Distance," *Philosophy & Public Affairs* 32 (2004): 357-383.
- Moore, G.E. *Principia Ethica* (Cambridge, UK: Cambridge University Press, 1903).
- Morris, C. "A Contractarian Account of Moral Justification" in Sinnott- Armstrong and Timmons (eds.), *Moral Knowledge?*, pp. 214-242.
- Murphy, L. *Moral Demands in Nonideal Theory* (Oxford: Oxford University Press, 2000).

- Nagel, T. *Mortal Questions* (Cambridge, UK: Cambridge University Press, 1979).
- _____, "The Sleep of Reason," *The New Republic* (October 12, 1998).
- Narveson, J. "We Don't Owe Them a Thing! A Tough-minded but Soft-hearted View of Aid to the Faraway Needy," *The Monist* 83 (2003): 419-33.
- Newman, L. "Descartes' Epistemology", *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/descartes-epistemology/>.
- Nussbaum, M. *Love's Knowledge* (Oxford: Oxford University Press, 1990).
- Petrinovich, L. and O'Neill, P. "Influence of Wording and Framing Effects on Moral Intuitions," *Ethology and Sociobiology* 17 (1996): 145-171.
- Piaget, J. *The Moral Judgment of the Child* (New York: Free Press, 1932).
- Pohl, R. (ed.) *Cognitive Illusions* (Sussex, NJ: Psychology Press, 2004).
- Prichard, H.A. "Does Moral Philosophy Rest on a Mistake?" *Mind* 21 (1912).
- Pryor, J. "The Skeptic and the Dogmatist," *Nous* 34 (2000): 517-49.
- Quine, W.V.O. *Word and Object* (Cambridge, MA: MIT Press, 1960).
- Rawls, J. "Outline of a Decision Procedure for Ethics" *Philosophical Review* 60 (1951): 177-97.
Reprinted in *Collected Papers* (ed. by Samuel Freeman) (Cambridge, MA: Harvard University Press, 1999), pp. 1-19.
- _____, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971).
- _____, "The Independence of Moral Theory," *Proceedings and Addresses of the American Philosophical Association* 48 (1974): 5-22. Reprinted in *Collected Papers* ed. Samuel Freeman, (Cambridge, MA: Harvard University Press, 1999), pp. 286-302.
- _____, "Reply to Habermas," *Journal of Philosophy* 92 (1995).

- Richardson, H. *Practical Reasoning about Final Ends* (Cambridge, UK: Cambridge University Press, 1994).
- Ross, W.D. *The Right and the Good* (Oxford: Oxford University Press, 1930).
- Sartre, J.P. "Existentialism is a Humanism," in Walter Kaufmann (ed.), *Existentialism from Dostoevsky to Sartre* (New York: Meridian, 1957), pp. 287-311.
- Sayre-McCord, G. "Coherentist Epistemology and Moral Theory," in Sinnott-Armstrong and Timmons (eds.) *Moral Knowledge*, pp. 137-189.
- Scanlon, T.M. "Rawls on Justification," in Samuel Freeman (ed.), *The Cambridge Companion to Rawls* (Cambridge, UK: Cambridge University Press, 2003), pp. 139-167.
- Scheffler, I. "Justification and Commitment," *Journal of Philosophy* 51 (1954): 180-190.
- Scheffler, S. *The Rejection of Consequentialism* (Oxford: Oxford University Press, 1982).
- Sellars, W.F. *Science, Perception, and Reality* (London: Routledge & Kegan Paul, 1963).
- Sencerz, S. "Moral Intuitions and Justification in Ethics," *Philosophical Studies* 50 (1986): 77-95
- Shafer-Landau, R. *Moral Realism: A Defense* (Oxford: Oxford University Press, 2003).
- Sidgwick, H. *The Methods of Ethics* 7th Ed. (London: MacMillan, 1907).
- Sieck, W. & Yates, J.F. "Exposition Effects on Decision Making: Choice and Confidence in Choice," *Organizational Behavior and Human Decision Processes* 70 (1997): 207-19.
- Singer, P. "Famine, Affluence, and Morality," *Philosophy and Public Affairs* 1 (1972): 229-43. Reprinted in *Unsanctifying Human Life* ed. by Helga Kuhse (Oxford: Blackwell, 2003), pp. 145-56.
- _____, "Sigwick and Reflective Equilibrium" *Monist* 58 (1975): 490-517. Reprinted in *Unsanctifying Human Life*, pp. 27-50.
- _____, "A Response" in Dale Jamieson (ed.), *Singer and his Critics* (Oxford: Blackwell, 1999).

- _____, *Unsanctifying Human Life* (ed. Helga Kuhse) (Oxford: Blackwell, 2002).
- _____, "Ethics and Intuitions," *Journal of Ethics* 9 (2005): 331-352.
- _____, *The Life You Can Save* (New York: Random House, 2009).
- Sinnott-Armstrong, W. "Moral Relativity and Intuitionism" *Philosophical Issues* 12 (2002): 305-82.
- _____, *Moral Skepticisms* (Oxford: Oxford University Press, 2006).
- _____, "Framing Moral Intuitions" in Walter Sinnott-Armstrong (ed.) *Moral Psychology Volume 2* (Cambridge, MA: MIT Press, 2008), pp. 47-76.
- _____ and Timmons, M. (eds.), *Moral Knowledge? New Readings in Moral Epistemology* (Oxford: Oxford University Press, 1996).
- Slote, M. "Famine, Affluence, and Empathy," in David Boonin and Graham Oddie (eds.), *What's Wrong? Applied Ethicists and Their Critics* (Oxford: Oxford University Press, 2005), pp. 548-556.
- Smith, S.M. & Levin, I.P. "Need for cognition and Choice Framing Effects," *Journal of Behavioral Decision Making* 9 (1996): 283-90.
- Stratton-Lake, P. (ed.), *Ethical Intuitionism: Re-evaluations* (Oxford: Oxford University Press, 2002).
- _____, "Introduction," in Stratton-Lake (ed.), *Ethical Intuitionism*.
- Stroud, B. *The Significance of Philosophical Skepticism* (Oxford: Oxford University Press, 1984).
- Sunstein, C. "Moral Heuristics," *Behavioral and Brain Sciences* 28 (2005): 531-573).

- Swain, S. Alexander, J. and Weinberg, J. "The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp," *Philosophy and Phenomenological Research* 56 (2008): 138-155.
- Thagard, P. and Verbeurgt, K. "Coherence as Constraint Satisfaction," *Cognitive Science* 22 (1998): 1-24.
- Thomson, J.J. "Killing, Letting Die, and the Trolley Problem," *The Monist* 59 (1976): 204-17.
- Timmons, M. "Foundationalism and the Structure of Ethical Justification," *Ethics* 97 (1986): 596-609.
- Tolhurst, W. "On the Epistemic Value of Moral Experience" *Southern Journal of Philosophy* 29 supp. (1990): 67-87.
- _____, "Seemings," *American Philosophical Quarterly* 35 (1998): 293-302.
- Tversky, A. and Kahneman, D. "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment," *Psychological Review* 90 (1983) : 293-315.
- Uhlmann, E.L. Piazzarro, D.A. Tannenbaum, D. and Ditto, P.H. "The Motivated Use of Moral Principles," (m.s.).
- Unger, P. *Ignorance: A Case for Skepticism* (Oxford: Oxford University Press, 1975).
- _____, *Living High and Letting Die: Our Illusion of Innocence* (Oxford: Oxford University Press, 1996).
- Wason, P.C. "Reasoning," in B.M. Foss, *New Horizons in Psychology* (Harmondsworth, UK: Penguin, 1966).
- Weinberg, J. Nichols, S. and Stich, S. "Normativity and Epistemic Intuitions," *Philosophical Topics* 29 (2001): 429-460.

Williams, B. "Ethical Consistency," *Proceedings of the Aristotelian Society* supp. 39 (1965): 103-124.

Wright, J. "On Intuitional Stability: the Clear, the Strong, and the Paradigmatic," *Cognition* 115 (2010): 491-503

Wood, A. *Kantian Ethics* (Cambridge, UK: Cambridge University Press, 2008).