MORAL INTUITIONS, MORAL PERCEPTIONS, AND MORAL DILEMMAS

By

David Killoren

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
(Philosophy)

at the

University of Wisconsin-Madison

2012

Date of final oral examination: 5/17/2012

The dissertation is approved by the following members of the Final Oral Committee:
Russ Shafer-Landau, Professor, Philosophy
Daniel Hausman, Professor, Philosophy
Robert Streiffer, Associate Professor, Medical History & Bioethics
John Bengson, Assistant Professor, Philosophy
James Paul Kelleher, Assistant Professor, Medical History & Bioethics

ACKNOWLEDGMENTS

I cannot list all of the people who provided me with various forms of invaluable support throughout my graduate studies. I am particularly—and immensely, and indescribably—grateful to all of the following people: Russ Shafer-Landau, Dan Hausman, Rob Streiffer, Carolina Sartorio, Thomas Carson, Jonathan Lang, Kate Myrna, Will Hunt, Don Killoren, Mary Killoren, and Bryan Killoren. For valuable comments and conversations that were specifically related to the papers contained in this dissertation, I am grateful to the individuals just mentioned, as well as Jeff Behrends, Juan Comesana, Stewart Eskew, Casey Helgeson, Justin Horn, Holly Kantin, Matt Kopec, Joshua Knobe, Christian Koons, Hallie Liberto, Clayton Littlejohn, Pete Nichols, Jeff Pretti, Leigh Vicens, Peter Vranas, Naftali Weinberger, Bekka Williams, and two anonymous reviewers for the *Journal of Ethics & Social Philosophy*, among many others.

Introduction

This dissertation contains three papers. The first paper, "Moral Intuitions, Reliability, and Disagreement," defends intuitivism, the view that moral intuitions provide a minimally reliable guide to the moral truth. (I understand moral intuitions as seemings.) In essence, intuitivists believe that moral intuitions are correct more often than they are incorrect. I take it that those who rely on moral intuitions for the purpose of assessing moral theories and making moral decisions are committed to intuitivism. My aim in "Moral Intuitions, Reliability, and Disagreement" is to defend intuitivism against a range of arguments against it. In particular, I consider a kind of genealogical critique of intuitivism, according to which the psychological mechanisms that produce our moral intuitions make our intuitions unlikely to be reliable; and I consider an argument from disagreement, according to which the interpersonal and intrapersonal diversity of moral intuitions casts doubt on their reliability. I maintain that both of these lines of argument fail. If my arguments are successful, then they do not show that intuitivism is true, but they do show that at least two popular arguments against intuitivism are *not* successful.

In the second paper, "Non-Naturalist Moral Perception," I attempt to develop a model of moral intuition as a type of perception. I intend for this model to be consistent with metaethical non-naturalism, according to which moral facts are not natural facts. One important objection to the possibility of non-naturalist moral perception appeals to the causal constraint on perception, which says that, in order for S to perceive that P, there must be causal contact between S and P. If the causal constraint is correct, then non-naturalists must either say that there is causal interaction between non-natural facts and natural facts, or deny the possibility of non-naturalist moral perception. The first of these options is unattractive for various reasons (e.g., the first option seems incompatible with the appealing view that natural facts can be given a wholly natural explanation). Thus, it appears that non-naturalists ought to either deny the causal constraint on perception, or deny the possibility of non-naturalist moral perception. I believe that non-naturalists ought to attack the causal constraint on perception. I argue that the causal constraint ought to be replaced with an explanatory relatedness constraint, which requires that, in order for S to perceive that

P, S must bear an appropriate explanatory relationship with the fact that P. I argue that this constraint makes room for non-naturalist moral perception. In particular, I offer the Fork Model, which says that, when we intuit that P (where P is a moral proposition), this intuition is usually generated by our recognition of a set of natural facts, and these very natural facts make it the case that P is true. I argue, with reference to a range of examples of non-moral perception, that the Fork Model shows a way in which it is possible for moral intuition to count as a form of non-naturalist moral perception. If my argument succeeds, then I do not show that moral intuitions *really are* moral perceptions; I only show that moral intuitions *could be* moral perceptions, and that metaethical non-naturalism does not rule this out.

In the third paper, "Knots, Dilemmas, and Designers," I examine phenomena of moral conflict and their relation to the debate between metaethical constructivists and metaethical realists. I define a moral knot as a situation in which an agent cannot satisfy all of her absolute moral obligations (such that the agent cannot avoid violating at least one of her absolute moral obligations). I define a moral dilemma as a situation in which an agent has an absolute moral obligation to do X, and an absolute moral obligation to do Y, but the agent cannot do both X and Y. A moral dilemma is a type of moral knot: all dilemmas are knots, but not all knots are dilemmas. "Knots, Dilemmas, and Designers" contains one line of argument for the possibility of moral knots, and another line of argument for the possibility of moral dilemmas. The argument for the possibility of moral knots appeals to situations of infinite choice: situations in which (a) an agent has infinitely many options, and (b) for each option open to the agent, there is a better option that is also open to the agent. I argue that at least some situations of infinite choice count as moral knots. The argument for the possibility of moral dilemmas appeals to the famous problem raised by trolley cases. I suggest that one solution to the trolley problem (which has been mostly overlooked in the literature) is to diagnose (some) trolley cases as moral dilemmas. I suggest that trolley cases provide the basis for an argument in favor of moral dilemmas. (I will say a bit more about this argument below.) This argument does not provide a decisive case in favor of the existence of moral dilemmas, but it does suggest that there is a *reasonable chance* that moral dilemmas exist. Given this, I maintain, we ought to investigate what would follow from the existence of moral dilemmas. I argue that if

moral dilemmas exist, then this would be problematic for otherwise viable versions of metaethical constructivism, but would *not* be problematic for metaethical realism. Thus, I argue, constructivists ought to oppose the existence of moral dilemmas, but realists ought to be open to the existence of moral dilemmas, and perhaps ought to even welcome the existence of moral dilemmas.

All of these papers contribute to a larger project, which involves the defense of a position that I call Strict Pluralism. In the remainder of this Introduction, I will describe Strict Pluralism and its rationale, and will explain how the three papers contained in this dissertation contribute to the defense of Strict Pluralism.

To begin, consider two trolley cases. In $Switch_{20}$, 20 people will be struck and killed by a runaway trolley—unless a certain agent throws a certain switch, in which case the trolley will be diverted to a side-track, where it will kill a single person. In $Push_{20}$, 20 people will be struck and killed by a runaway trolley—unless a certain agent pushes a certain football player from a bridge, in which case the football player will be struck and killed by the trolley, and the trolley will stop before it reaches the 20 people. Here is an inconsistent tetrad:

I_1: The agent in $Switch_{20}$ faces an absolute moral obligation to throw the switch.

I_2: The agent in $Push_{20}$ faces an absolute moral obligation *not* to push the football player.

Difference Requirement: I_1 and I_2 are both true only if there is a morally relevant difference between $Switch_{20}$ and $Push_{20}$.

No Difference: There is no morally relevant difference between $Switch_{20}$ and $Push_{20}$.

These four claims are inconsistent. Any three of these claims form the premises of an argument against the fourth.

Most discussions of the trolley problem presuppose that, whatever we do, we should not reject the Difference Requirement. Indeed, the option of rejecting the Difference Requirement is rarely even considered. Thus, it is generally assumed that we have to either reject I_1, or reject I_2, or reject No Difference. So, for instance, some non-consequentialists reject I_1; a wide range of consequentialists reject I_2; proponents of the doctrine of double effect tend to reject No Difference—etc.

The problem, however, is that I_1, I_2, and No Difference are all attractive, at least initially. I_1 and I_2 are intuitively plausible: I_1 and I_2 both *seem* to (many of us to) be true. Further, No Difference is attractive in light of the history of the trolley problem. The above trolley cases have been scrutinized by many philosophers during the course of the past four decades, and many hypotheses about the relevant difference between $Switch_{20}$ and $Push_{20}$ have been proposed. But no such hypothesis has won the day. For instance, among those who wish to reject No Difference, the most popular view is that the relevant difference between $Switch_{20}$ and $Push_{20}$ is furnished by the doctrine of double effect. In $Switch_{20}$, the person on the side-track is killed merely as a side effect of an action squarely aimed at saving the twenty—whereas, in $Push_{20}$, the football player must be killed intentionally, as a means in order to save the twenty. According to the doctrine of double effect, this difference is morally relevant. Unfortunately, however, the doctrine of double effect turns out to have a number of important difficulties. For instance, although the implications of the doctrine of double effect for $Switch_{20}$ and $Push_{20}$ are intuitively plausible, its implications for a number of other cases (e.g., the so-called "loop case") are not intuitively plausible. In light of such difficulties, the doctrine of double effect remains a minority view among moral philosophers. The doctrine of double effect is not unusual in this regard. Every hypothesis about the morally relevant difference between $Switch_{20}$ and $Push_{20}$ has failed to win the majority support. The fact that no one has uncovered an uncontroversially relevant difference between $Switch_{20}$ and $Push_{20}$ provides some amount of prima facie evidence for the view that no such difference exists.

Given that I_1, I_2, and No Difference each possess a certain measure of attractiveness, there would be considerable advantage in any theory that could allow us to affirm I_1, I_2, and No Difference. Such a

theory would have to allow us to reject the Difference Requirement. A position that I call the Parity Solution can allow us to do just that.

The Parity Solution affirms all of the following claims:

I_1: The agent in $Switch_{20}$ faces an absolute moral obligation to throw the switch.

S_1: The agent in $Switch_{20}$ faces an absolute moral obligation *not* to throw the switch.

I_2: The agent in $Push_{20}$ faces an absolute moral obligation *not* to push the football player.

S_2: The agent in $Push_{20}$ faces an absolute moral obligation to push the football player.

The Parity Solution implies that $Switch_{20}$ and $Push_{20}$ are moral dilemmas. That is, it implies that, in both $Switch_{20}$ and $Push_{20}$, the agent has two absolute moral obligations, but the agent cannot satisfy both of those obligations.

There are many things that might be wrong with the Parity Solution. However, the Parity Solution can claim at least one advantage: the Parity Solution allows us to affirm I_1, I_2, and No Difference, while rejecting the Difference Requirement. The Parity Solution explicitly affirms both I_1 and I_2. Further, since the Parity Solution posits that $Switch_{20}$ and $Push_{20}$ are morally parallel—i.e., in both $Switch_{20}$ and $Push_{20}$, the agent is morally obligated to kill the one in order to save the twenty, *and* morally obligated not to do so—the Parity Solution allows us to deny that there is any morally relevant difference between $Switch_{20}$ and $Push_{20}$. So the Parity Solution allows us to affirm No Difference.

As I argue in "Knots, Dilemmas, and Designers," the prospects for a successful defense of the Parity Solution depend on whether it is possible to develop a plausible moral theory that would explain why the Parity Solution is true (if it is true). I suggest that such a theory might include a consequentialist principle (C) to explain I_1 and S_2, and a non-consequentialist principle (NC) to explain I_2 and S_1. These two principles would come into conflict in cases such as $Push_{20}$ and $Switch_{20}$, which would explain why $Push_{20}$ and $Switch_{20}$ are moral dilemmas.

A theory counts as a version of Strict Pluralism if (a) the theory includes principles that come into conflict in particular cases, and (b) such cases count as moral dilemmas. The conjunction of C and NC is a version of Strict Pluralism—one of many possible versions. Note: Strict Pluralism is an alternative to Rossian Pluralism. Rossian Pluralism, like Strict Pluralism, includes multiple principles that may come into conflict in particular cases; but for the Rossian, such conflicts are always only conflicts between prima facie (pro tanto) moral duties, not absolute moral obligations. Thus, Rossian Pluralism, unlike Strict Pluralism, does not predict the existence of genuine moral dilemmas.

My overall aim, in present and future research, is to articulate, develop, and defend Strict Pluralism. My defense of Strict Pluralism will need to have two main components: (1) The first component will involve defense of three claims: I_1, I_2, and No Difference. These claims provide the premises of an argument for the Parity Solution, and the Parity Solution, as we have seen, provides a rationale for Strict Pluralism. The principal reason to accept I_1 and I_2 is that I_1 and I_2 are intuitively plausible. Thus, defense of I_1 and I_2 requires defense of the view that our moral intuitions provide reliable information about the moral truth. My arguments in "Moral Intuitions, Reliability, and Disagreement" and "Non-Naturalist Moral Perception" contribute to the defense of that view. Defense of No Difference will require, among other things, an attack on the main hypotheses that posit a morally relevant difference between cases like $Switch_{20}$ and $Push_{20}$. Seeds for such an attack are in "Knots, Dilemmas, and Designers." (2) The second component of my defense of Strict Pluralism will require me to develop a specific version of Strict Pluralism, identify initially cogent objections to that theory, and answer those objections. I offer one specific version of Strict Pluralism in "Knots, Dilemmas, and Designers," and I address a range of objections to that version.

Thus, the three papers contained in this dissertation represent the beginnings of a defense of Strict Pluralism, and provide a foundation for inquiry into Strict Pluralism's main advantages and disadvantages. However, more work needs to be done before we can know whether Strict Pluralism should be counted as a viable moral theory.

Moral Intuitions, Reliability, and Disagreement

Overview

There is an ancient, yet still lively, debate in moral epistemology about the epistemic significance of disagreement. One of the important questions in that debate is whether, and to what extent, the prevalence and persistence of disagreement between our moral intuitions causes problems for those who seek to rely on intuitions in order to make moral decisions, issue moral judgments, and craft moral theories. Meanwhile, in general epistemology, there is a relatively young, and very lively, debate about the epistemic significance of disagreement. A central question in that debate concerns peer disagreement: When I am confronted with an epistemic peer with whom I disagree, how should my confidence in my beliefs change (if at all)? The disagreement debate in moral epistemology has not been brought into much contact with the disagreement debate in general epistemology (though McGrath [2007] is an important exception). A purpose of this paper is to increase the area of contact between these two debates. In Section 1, I try to clarify the question I want to ask in this paper – this is the question whether we have any reasons to believe what I shall call "anti-intuitivism." In Section 2, I argue that anti-intuitivism cannot be supported solely by investigating the mechanisms that produce our intuitions. In Section 3, I discuss an anti-intuitivist argument from disagreement which relies on the so-called "Equal Weight View." In Section 4, I pause to clarify the notion of epistemic parity and to explain how it ought to be understood in the epistemology of moral intuition. In Section 5, I return to the anti-intuitivist argument from disagreement and explain how an apparently-vulnerable premise of that argument may be quite resilient. In Section 6, I introduce a novel objection against the Equal Weight View in order to show how I think we can successfully resist the anti-intuitivist argument from disagreement.

1. Introduction

Consider two scenarios:

(1) Jeb is deciding whether to do *A*. Jeb's only goal is to do the morally right thing. To decide whether or not to do *A*, Jeb first carefully considers the non-moral ("brute") facts about *A* and about the circumstance in which *A* would be performed. Then Jeb consults his intuition,[1] i.e., he asks himself whether, in light of the brute facts, *A* seems right or seems wrong. Jeb's intuition is that *A* is morally right. So Jeb does *A*.

(2) Matt is deciding whether to do *A*. Matt's only goal is to do the morally right thing. To decide whether or not to do *A*, Matt takes a coin out of his pocket. Matt decides that "heads" means *A* is right and "tails" means *A* is wrong. Matt flips the coin. Heads is the result, so Matt does *A*.

My question is this: Are there any good reasons to think Jeb's way of deciding to act is no better than Matt's way of deciding to act? Matt's goal is to act rightly, but he has chosen a very bad way to aim at that goal – in fact, he could do no worse unless he would aim at acting wrongly. Is there good reason to think Jeb is in the same boat?

Why this question? Well, an affirmative answer would be disastrous for those who rely on their intuitions for guidance in moral decision-making. This fact, together with the fact that so many of us –

[1] What is a moral intuition? Some (e.g. Sinnott-Armstrong 2008a, p. 47) have defined intuitions as immediate beliefs – immediate in that they are non-inferential. I agree that intuitions are immediate in this way, but I do not think the definition of intuitions as beliefs captures standard (philosophical) usage of the term "intuition." One can doubt one's intuitions (e.g. Peter Singer appears to have confessed to having intuitions that are inconsistent with his utilitarian beliefs). I think it is better to understand intuitions as *seemings* (so, on this point, I think I follow Bealer 1998 and Huemer 2006). Just as a twig in water can seem bent even if I do not believe it is bent, we can intuit propositions we do not believe. Despite this, of course, most people do believe most of their intuitions. I will use "intuitive judgment" to refer to moral beliefs formed on the basis of one's intuitions.

philosophers and ordinary people alike[2] – do rely on our intuitions, seems sufficient to make my question interesting. Of course, a negative answer to this question would do little to vindicate reliance on intuitions. To vindicate reliance on intuitions, we would need to show that there are good reasons to think Jeb's way of deciding *is* better than Matt's. But I will not attempt to determine whether such reasons exist; the question I have chosen to examine is already difficult enough (as we will soon see).

Another question worth asking is whether Jeb is, or can be, *justified in believing* that *A* is the right thing to do. This is a question moral epistemologists have traditionally asked. But this question sets the bar higher than I (or Jeb) would want it. If I have a coin that is slightly biased toward heads, so that it comes up heads 50.1% of the time (and I know this), I am probably not justified in believing that the next flip will come up heads, but I might well be justified in *acting as if* the next flip will come up heads. For example, I ought to bet on heads in such a case (if I am not risk-averse). In general, it seems that it is easier to acquire justification to act as if *p* than it is to acquire justification to believe *p*. And for those of us who, like Jeb, want to know how to aim to do the right thing, justification to act as if *A* is right is sufficient; we do not need justification to believe *A* is right (although we would like to have that too).

Of course, even if we were to establish that moral intuition is better (from the perspective of those aiming to do the right thing) than coin-tossing, this alone would not be enough to establish that one is justified in acting as if one's intuitions were correct. That is because there might be another source of information about morality that is even better than intuition. If so, then when information from that source conflicts with the results of intuition, one might not be justified in acting as if one's intuitions were true, even if intuiting (taken by itself) were better than coin-tossing. Still, the comparison of intuition

[2] In fact, even though ordinary people are aware that they rely heavily on intuitions, they still may radically underestimate the degree of that reliance. A growing body of psychological research suggests that, even when it seems to us that we reach a judgment through some form of principles-based reasoning, it is often the case that the judgment was formed well in advance of conscious reflection. (See, e.g., Haidt 2001.) Thus, intuition both seems to be a large part of everyday moral thinking *and* is probably an even larger part of everyday moral thinking than it seems to be.

with coin-tossing provides a baseline. I call the view that intuition is better than coin-tossing **intuitivism** and the view that intuition is no better than coin-tossing **anti-intuitivism**.[3]

I will sustain two assumptions throughout the course of this paper. The first assumption is that there is an intuition-independent moral reality: one's moral intuition that *p* does not by itself make *p* true; so any given intuition *could* be mistaken. (This is not equivalent with the somewhat more controversial claim that intuitions have *no role whatsoever* in making moral judgments true.) I think this assumption is highly plausible, but that is not the reason I make it. I make it because if it were false, intuitivism would certainly be true. So there is an interesting question about anti-intuitivism's prospects only if this assumption is made.

Secondly, I will assume that there is no source of information about morality other than intuition. This means that if intuiting is no better than coin-tossing, then (e.g.) there is no form of principles-based moral reasoning better than coin-tossing, either. I call this assumption **No Alternatives**. I assume No Alternatives simply because the stakes are highest, and therefore this paper's problem is most interesting, if it is true: Given No Alternatives, we would be completely in the dark about our moral obligations unless our intuitions have some degree of reliability. Of course, it might be that the assumption of No Alternatives will influence the outcome of my analysis. For example, it might be that, assuming No Alternatives, we can show that there are no reasons to believe anti-intuitivism, but without the assumption of No Alternatives, we could produce very good reasons to believe anti-intuitivism. This might be a reason not to assume No Alternatives. And in fact, I do think that reasons for anti-intuitivism might be obscured by the assumption of No Alternatives. That is because, if there is an alternative source of information about the moral truth, and if intuition routinely clashes with that alternative source, this might provide us with very good reason to think intuition is completely unreliable. But this is simply obvious. The really interesting question, I think, is whether we might have reason to think intuition is completely

[3] I borrow the name "intuitivism" from Robert Audi 2005, but my usage differs somewhat from his (I think what he calls "intuitivism" implies what I call "intuitivism" but not vice-versa). My main reason to use this name, rather than the more common name "intuitionism," is that I want to allow intuitivism to be free of associations with Intuitionism, the philosophical tradition given fullest expression in the philosophy of W.D. Ross, a tradition which typically goes far beyond the very minimal assertion I want to use "intuitivism" to pick out.

unreliable if we *don't* have anything other than intuition on which to rely for moral guidance. To examine that question, we should assume No Alternatives.

2. Why anti-intuitivists should look beyond the mechanisms that produce our intuitions

One source of evidence that Jeb and Matt are in the same boat might be found in the persistence and prevalence of disagreement between intuitions.[4] We may distinguish two kinds of disagreement. First, there can be *interpersonal* disagreement, as when Ellen intuits that slavery is wrong but Ellen's uncle intuits that slavery is permissible. Second, there can be *intrapersonal* disagreement, as when, on one occasion, Ellen intuits that slavery is wrong, while on another occasion, Ellen intuits that slavery is permissible. It has seemed to some that, if either or both of these kinds of disagreement are widespread, then intuition is obviously unreliable. Beginning in the next section, I will discuss in detail the nature of the problems posed for intuitivism by these forms of disagreement. But according to an alternative line of argument, we can show intuitions to be unreliable by investigating the psychological mechanisms that give rise to them. That is the line of argument I will examine in the present section. I will first say something about a particular instance of this line of argument, which proceeds from research conducted by Joshua Greene. Then I will say something more general about the prospects for this line of argument.

Joshua Greene [2001] has used fMRI brain-imaging techniques to investigate the way in which we make intuitive judgments about trolley cases introduced into the philosophical literature by Philippa Foot [1967] and Judith Thomson [1976]. As is well known, most people intuit that inaction is morally required in the so-called Push Case, but inaction is not morally required in the so-called Standard Case.[5]

[4] If intuitions are seemings rather than beliefs (see fn. 1), one might ask how disagreement between intuitions is possible. I will say that one person's intuition disagrees with another person's intuition exactly if things cannot (possibly) be the way they seem to be to the first person and (at the same time) be the way they seem to be to the second person. So, e.g., if I intuit that *A* is (all-things-considered) morally wrong, and you intuit that *A* is (all-things-considered) morally permissible, then presumably we disagree in the intended sense (assuming *A* cannot be all-things-considered both wrong and permissible).

[5] In the Standard Case, a runaway trolley is hurtling down a track. The observer must choose whether to hit a switch or to do nothing. If she does nothing, then the trolley will continue along its current path and will kill five

Many philosophers have believed that our intuitions about these cases are obviously correct; the puzzle is that there is no obvious principled explanation why they are (both) correct. And so philosophers have spent decades looking for morally-relevant differences between the two cases. However, Greene's research suggests that people's emotional reactions to the Push Case are different, and more intense, than their emotional reactions to the Standard Case, and Greene suggests that this difference in emotional response may explain why we typically have different intuitive judgments about these cases. Commenting on Greene's research, Peter Singer [2005] writes:

> If…Greene is right to suggest that our intuitive responses are due to differences in the emotional pull of situations that involve bringing about someone's death in a close-up, personal way, and bringing about the same person's death in a way that is at a distance, and less personal, why should we believe that there is anything that justifies these responses? If Greene's initial results are confirmed by subsequent research, we may ultimately conclude that he has not only explained, but explained away the philosophical puzzle [posed by the trolley cases]. (p. 347)

Singer thinks that Greene has pointed the way to a debunking explanation of our intuitions about these cases. It is easy to imagine that this debunking project might be extended to our intuitions about many other cases. Perhaps it will eventually come out that all of our moral intuitions are explained by "emotional pull." If that happens, then it may seem that our intuitions would have shown themselves to be a quite poor guide to moral truth. We may then need to conclude that moral beliefs cannot be justified by intuitions. And – although Singer does not go quite this far – we might even need to conclude that our intuitions have no better chance of getting moral matters right than simply tossing a coin.

---

people who are tied to the track. If the observer hits the switch, then the trolley will go down a side track, avoiding the five, but killing one person who is tied to the side track. In the Push Case, a runaway trolley is hurtling down a track. The observer must choose whether to push a very large man from a bridge or do nothing. If she does nothing, then the trolley will continue along its current path and will kill five people. If the observer pushes the very large man from the bridge, then he will land in front of the trolley, stopping its motion, so the five will be saved. But the very large man will die in the process.

Singer is not the only one making this kind of an argument. Greene's work has inspired similar arguments,[6] and Greene's is not the only work in moral psychology that has been marshaled to support a similar conclusion.[7] But it is not entirely clear how a Greene-style debunking explanation successfully debunks. I see two different arguments that may be extracted from the paragraph excerpted above. First, Singer might intend to argue as follows: Clearly, the fact that we respond emotionally to a given action (e.g. pushing a person from a bridge in order to save five people) does not make that action right or wrong; therefore, if our intuitions are generated by our emotions, then they provide no good evidence as to what is right or wrong. But we should not attribute this argument to Singer. That is because this argument is horrible: the premise obviously provides no support for the conclusion. It is as if one were to say that a thermometer can provide no evidence of the temperature because the thermometer has no capacity to make it hot or cold.

A second line of argument might go as follows: We know that in many contexts, our emotional responses provide no evidence about right or wrong. For example, many people respond with disgust, even horror, at the sight of an open-heart surgery in progress; but that is no evidence that open-heart surgery is wrong. Therefore, if our intuitions are generated by our emotions, then they provide no good evidence as to what is right or wrong. I am not sure whether this is the sort of argument Singer intends; in any case, this argument appears to be stronger than the previous one. But it is still not very compelling. Here is why. When I observe an open-heart surgery in progress, I am fairly disgusted, but I experience no intuition that the surgeon is doing anything wrong. (That is why the example of open-heart surgery might seem useful for a defense of skepticism about intuitions generated by negative emotions.) By contrast, when I consider the act of pushing a man from a bridge to save five others, I respond with a negative emotion *and* I experience the intuition that the act is wrong. So, if we assume that Greene is right, then

[6] People like Singer, including Greene himself (in, e.g., his 2008), seem to think that Greene's empirical research in the psychology of intuitions will produce some compelling evidence for some form of consequentialism. Broadly speaking, their idea seems to be either that consequentialism is motivated by something other than intuitions, or that consequentialism is motivated by intuitions for which no debunking explanation is forthcoming, or perhaps a mix of these.

[7] Another line of argument that is in some ways similar has been offered by Sharon Street (e.g., in her 2006). I discuss Street's argument below.

one of these emotions has a certain effect (it produces an intuition), while the other emotion lacks that effect. That means there must be a psychological difference, of some kind, between our emotional response to open-heart surgery and our emotional response to pushing a man from a bridge to save five others. Perhaps that difference – whatever it is – is epistemically important. Perhaps emotions of the kind we experience when we observe open-heart surgery provide no evidence about the moral truth, but emotions of the kind we experience when we imagine pushing someone from a bridge provide significant evidence about the moral truth. Until that possibility is ruled out, I do not see how the argument presently under consideration can go through. And obviously, one cannot rule out this possibility simply by pointing out that in *some* cases, such as the open-heart surgery case, our emotional response provides no evidence about the moral truth.

There are probably better readings of the passage from Singer excerpted above, although I do not know what they might be. So, rather than dwell on that passage any longer, I want to try to draw a general lesson from Singer's argument. At present, moral psychology has not given us a clear picture of the mechanisms by which our intuitions are generated. Perhaps our intuitions are generated by a fairly simple mechanism involving emotions, as Greene's research suggests. Perhaps there is a different mechanism. Perhaps the real mechanism, once discovered, will seem even less reliable, as a way to know moral truth, than the emotion-driven process posited by Greene. However, I suspect that, whatever the mechanism turns out to be, it will be quite difficult to put together a compelling case for anti-intuitivism *simply* by inspecting that mechanism. This is because that mechanism must necessarily generate intuitions that we find to be intuitively plausible (since otherwise, it would not be the mechanism that generates our intuitions) and so it will seem to us that this mechanism is quite well-functioning indeed.

The fact that my own intuition-generating mechanism must produce intuitively plausible intuitions is not evidence that this mechanism reliably homes in on the moral truth. That fact merely precludes counterintuitive output of my intuition-generating mechanism, thus precluding one kind of evidence that my intuition-generating mechanism *does not* home in on the moral truth. And of course counterintuitive output of my intuition-generating mechanism is not the only possible kind of evidence

that could support an anti-intuitivist's contentions. For example, if it turns out that my intuitions are produced by a malevolent demon, this would be evidence that my intuitions are not trustworthy.[8] But a demon-controlled intuition-generating mechanism is obviously unreliable only because (we imagine) the demon knows the moral truth and wants to deceive us about it. Such a mechanism thus has a built-in truth-avoidance feature. It is much more difficult to imagine an intuition-generating mechanism that seems untrustworthy, solely in virtue of its psychological description, *and* (a) is describable in purely naturalistic terms, (b) does not involve the influence of an ill-intending intelligence in the functioning of human psychology, and (c) produces only intuitively plausible intuitions. Yet it is very likely that our intuition-generating mechanism, once fully revealed by psychological research, will be shown to have characteristics (a)-(c). Given this it seems unlikely that a compelling case for anti-intuitivism will emerge solely from psychological research into the mechanisms that produce our moral intuitions.

Now I want to show that will be able to say something similar even if we broaden our focus, and examine not just our intuition-generating mechanisms but also the processes by which our intuition-generating mechanisms are generated. I will show this by (very briefly) examining the case of biological evolutionary explanations for our intuitions. It looks as though some (but not all[9]) parts of our intuition-generating mechanisms are the products of evolution through natural selection. Sharon Street argues that

[8] This sentence seems true, and I am willing to grant it for the sake of argument. But I believe it stands a very good chance of being false. I do believe that malevolent demons are very unreliable guides to the moral truth. But why do I believe this? I think it is because (a) there is a long list of moral judgments that I presently believe to be correct (e.g. I believe eating ice cream is usually morally permissible; burning cats for fun is usually morally wrong; etc.), and (b) I think the odds are very low that the output of a demon-influenced intuition-generating mechanism would match this list. Yet (b) would surely be false if my moral beliefs are shaped by intuitions produced by a malevolent demon. Thus, the revelation that my intuitions are produced by a malevolent demon might simply provide me with reason to revise my opinion about the reliability of malevolent demons (rather than provide me with reason to revise my opinion about the reliability of my intuitions).

[9] This is evident from the fact that moral intuitions about many issues (including infanticide, slavery, gender roles, etc.) vary radically from culture to culture. Some of these differences may be attributable to differences in non-moral beliefs, but not all. For example, Brandt [1954] describes Hopi children who torture their pet birds (c.f. Doris and Plakias [2008]). This behavior seems morally wrong to many of us, but apparently did not seem wrong to the Hopi children or to their elders, even though Brandt was unable to explain this moral disagreement by appeal to any disagreement about the brute facts (e.g. the Hopi seemed fully aware of the birds' capacity to feel pain). So there appear to be cases of cross-cultural *fundamental* disagreement between moral intuitions. This kind of variation cannot be explained by natural selection simply because it is very unlikely that it can be explained by genetic differences between members of different cultures. Below, in Section 6, I will say a bit more about the epistemic significance of the role of culture in the shaping of our intuitions.

this natural history dims the prospects for moral knowledge, at least if realism about value is true (and, it bears emphasizing, I have already committed myself to realism).[10] According to Street, if we think that realism is true, and we think that moral knowledge is possible, then we must endorse the "tracking account," according to which "making certain evaluative judgments rather than others promoted reproductive success [and therefore were favored by natural selection] *because these judgments were true*." Yet Street argues that the tracking account is not plausible. She thinks we should believe instead that making certain evaluative judgments rather than others promoted reproductive success, regardless of whether or not they are true, "because they forged adaptive links between our ancestors' circumstances and their responses to those circumstances, getting them to act, feel, and believe in ways that turned out to be reproductively advantageous." (This Street calls the "adaptive link account.") So, Street concludes, there must be something wrong with a picture in which realism is true, moral knowledge is possible, and the evolutionary explanation of our moral intuitions is correct. (Street also thinks there is something wrong with a picture in which moral realism is true and moral knowledge is not possible. So she thinks that, given the evolutionary explanation of our moral intuitions, we should not be moral realists.)

Street's issue is different from mine. Street wants to show (in the course of an argument against realism) that there cannot be moral knowledge unless realism is false. But we can readily grant this view without settling the question about intuitivism, the view that intuiting is better than coin-tossing. This is simply because even if moral knowledge is impossible, intuiting might still be better than coin-tossing (just as I may be unable to know in advance who will win a very close election contest, even though a guess based on the latest poll numbers is better than a guess based on a coin toss). So I do not need to determine whether Street's argument succeeds; the crucial question, for the purposes of this paper, is

[10] Street defines realism as the view that "there are at least some evaluative facts or truths that hold independently of all our evaluative attitudes." Whether or not Street's definition of realism is accurate or charitable is open to debate – a debate I will not enter here. For the limited purpose of evaluating the Street-inspired argument at issue here, I will regard realism as the view that that there is an intuition-independent moral reality: one's moral intuition that *p* does not by itself make *p* true, so any given intuition could be mistaken. This is a view I have assumed above; I will not evaluate the view here.

whether Street-style evolutionary considerations provide any reasons to believe intuiting is no better than coin-tossing.

Let us simply grant Street's view that the adaptive link account is true and the tracking account is false. The tracking account would rule out anti-intuitivism, but the adaptive link account does not rule out intuitivism. For example, the adaptive link account is consistent with the view that intuitions and moral facts have a common explanans. The presence of such a common explanans could conceivably ensure that our intuitions reflect the moral facts more often than not. So the adaptive link account does not entail anti-intuitivism. Of course, the adaptive link account could provide a reason against believing intuitivism without entailing that intuitivism is false. But it is unclear what that reason might be. Natural selection, in Street's view, is responsible for our intuitions; therefore, natural selection has in our case produced a mechanism whose output is uniformly intuitively plausible, and thus immune from criticism by counterintuitive counterexample. I grant that there is nothing in what we know about natural selection that should make us think it a reliable way to produce creatures whose moral intuitions tend to be true. But this is just a lack of a reason to think intuitivism is true; it is not a reason to think intuitivism is false. Street's inspection of the evolutionary origins of our intuitions thus seems to be like Greene's inspection of the emotional infrastructure undergirding our intuitions: both lines of investigation can deprive the intuitivist of ammunition, but will probably not provide the anti-intuitivist with ammunition.[11]

But none of this is to say that ammunition for the anti-intuitivist cannot be found. The phenomenon of moral disagreement might be fertile territory for the anti-intuitivist. An anti-intuitivist argument from disagreement would not need to posit an alternative to intuitions as a source of information about morality; its evidence for anti-intuitivism would consist entirely in observations about the frequency of conflict between moral intuitions. Such an argument also would not rely on inspection of the mechanisms by which our intuitions are produced. I think these are advantages of this line of argument. In the next section I will explain how such an argument might look.

[11] Below, in Section 6, I will argue that the situation is even worse than this for the anti-intuitivist: I will argue that evolutionary explanations for our intuitions might even provide an effective response to an anti-intuitivist argument from disagreement, thus depriving the anti-intuitivist as well as the intuitivist of some ammunition.

3. An anti-intuitivist argument from disagreement

Suppose I consider some particular case in which slavery is practiced – say, a case typical of slavery in the antebellum American South, which I will hereafter call "the Typical Case" – and I intuit that slavery is wrong in this case, whereas you intuit that slavery is permissible in this case. Then one of us is mistaken.[12] Once I discover the disagreement between us, it seems I have only three options: I can believe my intuition is correct; or I can believe your intuition is correct; or I can suspend judgment.[13] I do not think it is mistaken to say that I have only these three options, but I do think that saying this obscures a spectrum of possibilities. For instance, suppose that, prior to discovery of our disagreement, I believed with very high confidence that slavery is wrong. After discovering our disagreement, I might continue to believe this with the same high level of confidence; or I might continue to believe this with some significantly lower level of confidence. These seem like importantly different options, even though they are both ways of believing my intuition is correct. More generally: Suppose levels of confidence in a proposition *p* can be measured on an inclusive scale from 0 to 1, where 0 represents certainty that *p* is false, 1 represents certainty that *p* is true, and .5 represents suspension of judgment about whether *p*, i.e.

[12] This is so assuming (as I will assume here) that my intuition is that slavery is *all things considered* wrong, and your intuition is that slavery is *all things considered* permissible, and also assuming that slavery cannot be both all things considered wrong *and* all things considered permissible.

[13] There might also be a fourth option: I might conclude that there is no fact of the matter about whether slavery is permissible. My own view is that taking this option would be every bit as troubling as (C), the conclusion of the anti-intuitivist argument from disagreement laid out below. So I believe that the intuitivist should want to avoid this option just as much as she wants to avoid (C). But I will not try to support this view here.

agnosticism about whether *p*.[14] Then we should say that, in the moment after I discover my disagreement with you, I have as many doxastic options as there are points in the interval from 0 to 1.[15]

If we are justified in issuing moral judgments on the basis of our intuitions, as intuitivists believe, then probably different intuitions justify different levels of confidence. I intuit that torturing innocent children for fun is morally wrong. I also intuit that pushing a man from a bridge to save five others from an oncoming train is morally wrong. But the first intuition seems to me to be somehow *phenomenally stronger*,[16] and thus to justify much higher confidence, than the second intuition. Suppose that the degree of confidence justified by an intuition is determined solely by the phenomenal strength of that intuition (so that stronger intuitions justify higher levels of confidence). Call this the Strength-Confidence View. Those who accept this view (as well as others) may talk about my justified level of confidence, and distinguish this from my actual level of confidence. For example, it might be that, in light of my intuition that slavery is wrong, I have actual confidence 1 that slavery is wrong, whereas the phenomenal strength of my intuition only justifies confidence .9999 that slavery is wrong.

Intuitivism, I have said, is the view that issuing moral judgments on the basis of one's intuitions is better than issuing moral judgments on the basis of the outcome of a coin toss. The above points suggest a way to explain what is meant by "better" in that formulation. In the absence of any reason to believe that a given action *A* is or is not wrong, I should be an agnostic about whether *A* is wrong. A coin toss provides no reason to depart from this agnosticism. By contrast, the intuitivist should say, intuiting

---

[14] Some people – perhaps including Juan Comesana (personal correspondence) – would say that .5 should not be taken to represent suspension of judgment. They would say that each point along the spectrum from 0 to 1 represents a different judgment, so suspension of judgment is not found anywhere on the spectrum from 0 to 1. I do not need to dispute this point.

[15] There is some debate among epistemologists about whether all belief-states fit on a continuous scale of the sort I am discussing here. The remarks I have made thus far do not imply a stance in that debate. I have not ruled out the possibility that one can believe *p* without having a precise level of confidence in *p*; and I have not ruled out the possibility that believing *p* is something more (or something other) than having sufficiently-high confidence that *p*.

[16] A clear characterization of what I mean by "phenomenal strength" may prove elusive, but here are a couple of examples to illustrate the notion. "2+2=4" seems true to me. "Modal realism is false" also seems true to me. But the first seeming is somehow more vivid and more compelling; it has more of what I am calling "phenomenal strength." Similarly: Diet Coke seems like regular Coke to me. (I am not always able to tell the difference between Diet Coke and regular Coke, although I am able to tell the difference more often than not.) Regular Coke also seems like regular Coke to me. But I think the first seeming is typically less compelling than the second seeming, which is phenomenally stronger.

does provide some reason to depart from agnosticism. That is: For the intuitivist, if I intuit that *A* is wrong, then I am justified in having some level of confidence greater than .5 that *A* is wrong. This characterization of intuitivism leaves open the question *how much* of a departure from agnosticism my intuitions justify; different intuitivists will give different answers to that question. For example, intuitivists who accept the Strength-Confidence View will say that the extent of departure from agnosticism justified by an intuition is determined by the phenomenal strength of the intuition. But all intuitivists should agree (at a minimum) that some of my intuitions are *epistemically significant*, that is, some of my intuitions sometimes justify a departure from agnosticism.

Thus we have the first premise of the anti-intuitivist argument from disagreement:

(1) If intuitivism is true, then some of my intuitions are epistemically significant.

According to the next premise in the argument, I should attribute equal epistemic significance to each intuition of which I am aware, and whether that intuition is mine or someone else's, except in cases where there is some reason to discount or disregard a given intuition.[17] That is:

(2) Your intuitions have the same epistemic significance as mine unless there is some reason to attribute less significance to your intuitions than to mine.

The question raised by premise (2) is: What constitutes a reason to attribute less significance to your intuitions than to mine? A natural suggestion is to answer this question in terms of *epistemic parity*. The idea would be that, if my intuition conflicts with yours, e.g. if I intuit that slavery is wrong while you intuit that slavery is permissible, then I may not discount or disregard your intuition if you are my

[17] This idea is plausible because it is consistent with *any* view about what does or does not count as a reason to discount someone's intuition. For example, it is consistent with the view that one should discount one intuition, with respect to another intuition, when the first intuition is phenomenally weaker than the second; it is also consistent with the denial of this view.

epistemic peer; I may discount or disregard your intuition only if you are my epistemic inferior. (Something similar would be true about my past selves: I would be permitted to discount or disregard my past self's intuition only if I were epistemically superior to my past self.) If this is right, then we should agree to the following view:

> (3) **Equal Weight View**: There is no reason to attribute less epistemic significance to your intuition than to mine if I correctly consider you to be my epistemic peer or epistemic superior.[18]

To evaluate the Equal Weight View, we will need a definition for epistemic parity. What does it mean for one person to consider another person to be an epistemic peer? If I disagree with you about a given question, I consider you to be my epistemic peer if I consider you to be *as good as I am* at evaluating whatever sort of question is at issue – or (equivalently) if I consider your *epistemic credentials* to be as good as mine with regard to whatever sort of question is at issue. For example, I might consider you to be as good as I am at arithmetic, i.e. my arithmetical epistemic peer. Then (according to the Equal Weight View in its standard formulation) if we disagree about a given sum, I should not attribute more epistemic significance to your result than to mine. Rather, in the event of such a disagreement, I must (pending further evidence) be an agnostic about whether my result or your result is correct.

The example gives rough sense of what epistemic parity is, but this will not be sufficient for evaluating the Equal Weight View as it appears in premise (3). We may have a fairly good intuitive grasp of what it means for one person to be as good as another person at *arithmetic*, but it is not at all obvious

---

[18] For comparison, two alternative versions of the Equal Weight View deserve mention here:

> Version 1: There is no reason to attribute less epistemic significance to your intuition than to mine if I consider you to be my epistemic peer or epistemic superior.
>
> Version 2: There is no reason to attribute less epistemic significance to your intuition than to mine if you *really are* my epistemic peer or epistemic superior.

The view in (3) is entailed by Version 1 and by Version 2. Thus the view in (3) should be acceptable regardless of whether one accepts Version 1 or Version 2. In the next section I will discuss the difference between Version 1 and Version 2.

what it would mean for one person to be "as good as" another at *moral intuition*. Thus, before we can even understand the claim made by the Equal Weight View, let alone evaluate it, we will need to do some work to show how epistemic parity ought to be understood in the context of moral intuition. That is what I will do in the next section (Section 4).

Now for a brief interlude: I want to show how a skeptical argument offered by Sarah McGrath (2007) can be improved. This exercise will help motivate and explain the next premise in the anti-intuitivist argument from disagreement. McGrath works with a conception of epistemic parity developed by Adam Elga (2007). For McGrath (as for Elga), "[y]ou consider someone your epistemic peer with respect to a given question just in case: in advance of either of you reasoning about the issue, you would have predicted that the person in question was just as likely as you to arrive at the correct answer." And for Elga (as for McGrath), the Equal Weight View is the view that "one should give the same weight to one's own assessments as one gives to the assessments of those one counts as one's epistemic peers."[19] However, unlike Elga, McGrath thinks the Equal Weight View has skeptical consequences when applied in the moral domain. To show this, McGrath considers the example of Ann, a conservative Republican who thinks abortion is morally abhorrent in most circumstances, and Beth, a liberal Democrat who thinks abortion is morally permissible in most circumstances. Suppose Ann and Beth meet. After this encounter, could either Ann or Beth hold on to their opinions about abortion without falling afoul of the Equal Weight View? McGrath thinks not; she writes:

> [W]e would expect Ann and Beth to agree about the answers to any number of moral questions. We would expect them to agree, for example, that slavery is morally abhorrent, that it is wrong to cause others pain for the sake of one's own amusement, that lying is *prima facie* wrong, and about countless other issues. … In short, Ann and Beth's disagreements about abortion…,

[19] This is different in several respects from the formulation of the Equal Weight View I give above; one important difference is that Elga's definition refers to "assessments," whereas my definition is tailored specially for the epistemology of *moral intuitions*, which are seemings (not assessments). See Section 4 for discussion relevant to this difference.

> although substantial, almost surely take place against a relatively wide background of shared moral beliefs. … [T]he relatively wide background of agreement seems to tell in favor of Ann's taking it that Beth is more or less equally likely to get the hard questions right. (p. 105-6)

If Ann should think that she and Beth are about equally likely to get the "hard questions" right, then (given Elga's definition of epistemic parity) Ann should regard Beth as her epistemic peer. But then the Equal Weight View would require Ann to attribute the same significance to her own opinion as she does to Beth's opinion. In that case, it is hard to see how Ann could be permitted to hold on to her pro-choice convictions after she learns of Beth's pro-life convictions. Rather, it seems Ann should become an agnostic about the ethics of abortion.

I think McGrath is right that the Equal Weight View has skeptical implications, but the example of Ann and Beth does not by itself show this. According to the Equal Weight View, Ann ought to attribute the same significance to *each* of the opinions of *all* the people she considers to be peers. So Ann might not be required to give up her pro-choice convictions after learning of Beth's opinions even if Ann considers Beth to be her epistemic peer. Ann would be required to give up her pro-choice convictions only if she finds that, among *all* of her peers, pro-life convictions are as common as pro-choice convictions.[20] And perhaps Beth is not the only person whom Ann considers to be a peer.

[20] This may be counterintuitive. Suppose that, in one scenario, Beth is the only peer Ann knows, while in another scenario, Ann knows ten peers: Beth and nine others who have pro-choice opinions like Ann. If I am correct, then (according to the Equal Weight View) Ann should be moved to agnosticism by Beth's pro-life opinion in the first scenario, whereas in the second scenario Ann should retain her pro-choice opinion. But this makes it look as if Beth's opinion becomes (relatively) less significant as the size of Ann's acquaintances increases – a strange result. Even so, I think it is going to be very difficult to avoid this result without giving up the Equal Weight View. Suppose that Ann is moved to agnosticism in the second scenario. Then Ann (tacitly) considers Beth's pro-life opinion to be ten times as significant as any of her peers' pro-choice opinions. This is because, in the second scenario, there are ten peers (including Ann herself) who have pro-choice opinions – so if Ann becomes an agnostic about whether abortion is wrong, she treats these ten pro-choice opinions as if they are precisely counterbalanced by Beth's lone pro-life opinion. It would be contrary to the Equal Weight View to treat Beth's opinion as if it is ten times as significant as any of the other peers' opinions. By contrast, in the first scenario, there are only two peers: Ann and Beth. So if Ann is moved to agnosticism in that case, she treats all known peers' opinions as if they are equally significant, just as the Equal Weight View prescribes.

Similar points will hold when it comes to disagreement among moral intuitions about slavery. To show that I ought to give up my anti-slavery intuitions, it would not be enough to show that I have one epistemic peer with pro-slavery intuitions. Instead, what is needed is to show something like the following:

> (4) Among those who are either my epistemic peers or my epistemic superiors, a majority intuit that slavery is permissible in the Typical Case.

To motivate (4), one might begin with a seemingly-plausible historical claim: Pro-slavery intuitions have been very common throughout most of human history, and therefore anti-slavery intuitions represent a minority of all intuitions concerning slavery. I will discuss this historical claim below, in Section 5; it is not obviously true. But *if* it were true, then to deny (4) we would have to claim that the majority of my epistemic peers and superiors are concentrated among a minority of all the individuals who have ever lived. This apparently provides a prima facie presumption in favor of (4) – an obligation not to deny (4) unless there is a compelling objection to (4).

The foregoing is obviously not a complete defense of (4). I offer these points here only to give an idea why someone might believe (4). In Section 5, I will discuss in depth the possibilities for defending and attacking (4).

We have thus far assembled the following premises: (1) If intuitivism is true, then some of my intuitions are epistemically significant. (2) Your intuitions have the same epistemic significance as mine unless there is some reason to attribute less significance to your intuitions than to mine. (3) Equal Weight View: There is no reason to attribute less epistemic significance to your intuition than to mine if I correctly consider you to be my epistemic peer or epistemic superior. (4) Among those who are either my epistemic peers or my epistemic superiors, a majority intuit that slavery is permissible in the Typical Case. These premises strongly support the following conclusion:

> (C) If intuitivism is true, then I should not believe that slavery is wrong, and should not be an agnostic about whether slavery is wrong in the Typical Case; rather, I should have some confidence greater than .5 that slavery is permissible in the Typical Case.[21]

I think it is fairly clear that (C) makes intuitivism unacceptable. I am highly confident that slavery in the Typical Case is wrong – and highly confident that this confidence is justified. If intuitivism implies I must doubt whether slavery is wrong, then I must abandon intuitivism. So the intuitivist should want to deny one of premises (1)-(4). (1) and (2) seem obviously true, and I think should not be disputed. So the intuitivist ought to concentrate her attack on either (3) or (4). In Section 5, I will consider what can be said against (4); in Section 6, I will attack (3). But before we get to all that, it will be useful to address a few technicalities having to do with the concept of epistemic parity.

4. Epistemic parity in the epistemology of moral intuition

In trying to use the notion of epistemic parity in the context of this paper, I face two problems. The first problem is that the *standard* notion of epistemic parity is used in conventional epistemology and is therefore typically understood in terms of belief (or other doxastic states). But moral intuitions are not beliefs (or other doxastic states); they are, I have claimed, seemings (see fn. 1). Therefore, since moral intuitions are my focus in the present context, I will not be able to rely on an unmodified form of the standard notion of epistemic parity; I will need to develop an alternative notion. Ordinarily, the way to develop such an alternative notion would be to begin with the standard notion as a model and to adjust it

[21] It is not actually the case that (C) follows from (1)-(4). What follows from (1)-(4) is that, if intuitivism is true, then: If (a) I intuit that slavery is wrong, (b) my intuition that slavery is wrong (taken by itself, absent any other evidence) justifies a departure from agnosticism, (c) I can know that premise (4) is true, and (d) I have no other evidence (beyond my own intuitions and my knowledge of others' intuitions) about the morality of slavery, *then* I should have some confidence greater than .5 that slavery is permissible. But (a) is certainly true, (b) seems eminently plausible from any viable intuitivist perspective, (c) is probably true if (4) can be defended, and (d) is guaranteed by the assumption of No Alternatives from Section 1. Thus, if (1)-(4) are not contested, then (C) will prove to be very difficult to avoid.

in order to make it suit its new context. But this will not work, because there is in fact no such thing as *the* notion of epistemic parity. Rather, there are several different definitions in use by different epistemologists. Worse, these different definitions correspond with slightly different ways of understanding the Equal Weight View. This diversity of available models is the second problem I face.

I cannot here resolve existing disagreements about how we should understand epistemic parity, so I will have to remain neutral between several major approaches. In this section, I will discuss three different definitions of epistemic parity and will explain how we can use each of these three definitions as models for developing alternative definitions of epistemic parity for use in the epistemology of moral intuition. Then, in the next section (Section 5), I will use these results to explore the possibilities for a case against premise (4) of the anti-intuitivist argument from disagreement.

Adam Elga (whose view I discussed briefly in the previous section) takes what I shall call *the probabilistic approach* to epistemic parity. That approach is represented in the following definition:

> **Definition 1** S_1 considers S_2 to be an epistemic peer, with respect to a question Q, exactly if, prior to knowing S_2's opinion about Q, S_1 considers S_2 to be as likely as S_1 to answer questions like Q correctly.[22]

An important feature of Definition 1 is that it says only what it is to *consider* someone to be my peer; it does not give conditions for someone to *actually be* my peer. Thus Definition 1 leaves open how to determine whether I *correctly* consider you to be my epistemic peer. In fact, some philosophers who take the probabilistic approach maintain that there is not much at stake in the question whether you actually

[22] To fully understand this definition (and the alternative definitions considered below), it would be necessary to have criteria by which to determine whether a given question is "like" another. Spelling out such criteria is a lengthy project I will not be able to undertake here. Fortunately, it will often be clear whether one question is relevantly like another: for instance, any given arithmetic question is usually relevantly like any other arithmetic question of similar difficulty; and any given arithmetic question is usually not relevantly like the question whether Napoleon died in France or the question whether the man who just passed us on the street is Tom Cruise. But there will be hard cases, and a complete elaboration of the concept of epistemic parity would need to provide a way to decide such cases.

are my peer. These philosophers may say that there is simply no fact of the matter about whether you actually are my peer; or they may say that you actually are my peer just in case I consider you to be my peer.[23] In their view, if I find that my opinion conflicts with your opinion, I can determine how (if at all) I ought to modify my opinion by determining whether I consider you to be my peer; I do not need to determine whether you actually are my peer.

An alternative view of epistemic parity is represented in what I will call the *epistemic-virtues approach*, which is expressed by Thomas Kelly (2006):

> [T]he class of epistemic peers with respect to a given question are equals, not only with respect to their possession of the sort of general epistemic virtues enumerated by Gutting [which are: 'intelligence, perspicacity, honest, thoroughness, and other relevant epistemic virtues'], but also with respect to their exposure to evidence and arguments which bear on the question at issue. (p. 3)

Despite the specificity of the virtues listed here, those who accept a Kelly-style conception of epistemic parity will accept that the requirements for epistemic parity differ from one domain of inquiry to another. For example, if you and I disagree about a difficult question in mathematics, the fact that I am less intelligent than you might prevent me from being your epistemic peer. But if you and I disagree about whether the man who just passed us on the street is Tom Cruise,[24] the fact that I am less intelligent than you is unlikely to make much difference; it is much more important to know whether my eyesight is as good as yours, whether I have seen as many Tom Cruise movies as you, whether I am as good at recognizing faces as you are, etc. To put the point another way: intelligence is an important epistemic virtue in mathematics, whereas good eyesight is not; good eyesight is an important epistemic virtue in the recognition of faces, whereas intelligence is not. Presumably something similar will be true about

[23] I take Juan Comesana (personal correspondence) to be inclined to endorse the second of these possibilities.
[24] This example comes from Sinnott-Armstrong 2002.

epistemic vices: bad eyesight, or stupidity, will be vices in some contexts, not others. Thus, in the present view, to determine whether you are my epistemic peer, I must determine the epistemic virtues and vices relevant to whatever is the question at issue; and I must evaluate my mix of vices and virtues and compare it with your mix of vices and virtues.

Here is a definition of epistemic parity to represent the epistemic-virtues approach:

> **Definition 2** S_1 and S_2 are epistemic peers, with respect to a question Q, exactly if S_1's mix of epistemic virtues and vices is just as good (for the purpose of answering questions like Q) as S_2's mix of epistemic virtues and vices.

Those who take the epistemic-virtues approach will need to explain how to compare different mixes of vices and virtues. This question is going to prove difficult.[25] But according to one plausible criterion, I am more virtuous (and less vicious) than you (with respect to a given question Q) exactly if my belief-forming process, when used to form beliefs about questions like Q, is more *reliable* than yours.[26] If we accept this criterion, we may be able to dispense with the language of virtues and vices altogether and adopt what I will call the *reliability approach* to epistemic parity, in which we may use a definition of epistemic parity like the following:

> **Definition 3** S_1 and S_2 are epistemic peers, with respect to a question Q, exactly if S_1's belief-forming process, when used to form beliefs about questions like Q, is just as reliable as S_2's belief-forming process.

---

[25] To illustrate one source of difficulty, consider an example modified from Christensen [2007]. Suppose I believe that you are just as intelligent, thoughtful, and free from bias as I am. We both compute 376x215. My result is 80840. Your result is 2. It seems clear that, in this case, I should not attribute any epistemic significance to your opinion – but it is not at all clear whether I have found reason to think you lack some relevant vice. Those who accept Definition 2 will need to show how your result has revealed some epistemic vice that I did not previously know you possessed, or else they will have to abandon the Equal Weight View.

[26] Roughly: One belief-forming process is as reliable as another, with respect to questions like Q, if the first process produces true beliefs about questions like Q as often as the second process does. Thus reliability is a certain sort of measure of accuracy.

Thus far I have laid out three different conceptions of epistemic parity. I think each is initially plausible. These definitions are not obviously inconsistent with one another, although I suspect that Definition 3 is inconsistent with Definition 1.[27] Now I want to note that these three definitions correspond with different ways of spelling out the Equal Weight View. On the one hand, there is an *internalist* version of the Equal Weight View, which says that if I *consider you to be* my epistemic peer, then I ought to attribute the same epistemic significance to your opinion as to mine. On the other hand, there is an *externalist* version of the Equal Weight View, which says that if you *really are* my epistemic peer, then I ought to attribute the same epistemic significance to your opinion as to mine. A definition like Definition 1 (the probabilistic approach) may be useful in articulating the internalist version; a definition like Definition 2 (the epistemic-virtues approach) or Definition 3 (the reliability approach) may be useful in articulating the externalist version. It is not easy to decide whether the externalist version or the internalist version of the Equal Weight View is better; both have apparent disadvantages.[28] But note that these two versions of the Equal Weight View have conflicting implications primarily in cases where one person *mistakenly* judges another person to be her epistemic peer. Therefore it is mainly when dealing with such cases that we will need to pay special attention to the difference between the internalist and externalist versions. (Of course, as we have seen, it is at least arguable that such cases are not even possible.)

We can use each of the definitions I have reviewed as a model for developing a conception of epistemic parity for use in the context of moral intuition. Following Definition 3, which defines epistemic parity in terms of the reliability of *belief*-forming processes, we may wish to understand

---

[27] Definition 3 would be consistent with Definition 1 if it were the case that: S_1 considers S_2 to be as likely as S_1 to correctly answer questions like Q *exactly if* S_1 judges that S_2's belief-forming process is just as reliable as S_2's belief-forming process when applied to questions like Q. Unfortunately, this biconditional is probably not strictly true (for reasons I will not go into here), although it might be approximately true.

[28] The internalist version implies that, if I consider you to be my epistemic peer for very bad reasons, I still ought to attribute equal significance to your opinion as to mine. The externalist version implies that, if in fact you are my epistemic peer, but I have no way of knowing this, I still ought to attribute equal significance to your opinion as to mine. Obviously, neither of these counterintuitive implications alone provides decisive refutation of either the internalist or the externalist versions.

epistemic parity, in the context of moral intuition, in terms of the reliability of *intuition*-forming processes. In that case, we may use the following definition of epistemic parity:

> **Definition 3.I** S_1 and S_2 are epistemic peers, with respect to a case *C*, exactly if S_1's intuition-forming process, when used to form intuitions about cases like *C*, is just as reliable as S_2's intuition-forming process.[29]

Similarly, if different epistemic virtues and vices are relevant in different kinds of truth-seeking endeavors, then perhaps too there are virtues and vices relevant in the domain of moral intuition. For example, it seems not unreasonable to think that, if you do not have enough intelligence to understand the details of a given case, then you cannot be my peer (provided my intelligence *is* sufficient to understand the details of the case). (I will discuss some other potentially intuition-relevant virtues in the next section.) So, in the domain of moral intuition, the possession of a minimum level of intelligence looks like an epistemic virtue. Such considerations inspire the following definition:

> **Definition 2.I** S_1 and S_2 are epistemic peers, with respect to a case *C*, exactly if S_1's mix of epistemic virtues and vices is just as good as S_2's mix of epistemic virtues and vices.

It is also possible to take a probabilistic approach to understanding epistemic parity in the domain of moral intuition. In this approach we would define consideration of someone to be a peer as follows:

> **Definition 1.I** S_1 considers S_2 to be an epistemic peer, with respect to a case *C*, exactly if, prior to knowing S_2's intuition about *C*, S_1 considers S_2 to be as likely as S_1 to have correct intuitions about *C*.

[29] Roughly: One intuition-forming process is as reliable as another, with respect to cases like *C*, if the first process produces true true intuitions as often as the second process does. Thus, as with belief-forming processes, reliability here can be considered to be a measure of accuracy.

I do not know which of these definitions we ought to adopt, so in what follows I will strive to remain neutral between them. In the next section I will discuss some arguments for and against premise (4) of the anti-intuitivist argument from disagreement.

5. How many of my epistemic peers have pro-slavery intuitions?

Here again is premise (4):

(4) Among those who are either my epistemic peers or my epistemic superiors, a majority intuit that slavery is permissible in the Typical Case.

Here is a quick-and-dirty argument in favor of (4). Most people, during most of human history, have had pro-slavery intuitions; in fact, the intuition that slavery is wrong in cases like the Typical Case is a relatively recent phenomenon. Thus we can assume that the sheer number of people who intuit (or when alive did intuit) that slavery is permissible in the Typical Case is significantly larger than the number of people who intuit that slavery is wrong in the Typical Case. Given this, the only way to resist (4) is to argue that the *percentage* of people with anti-slavery intuitions who are my peers (or superiors) is much higher than the percentage of people with pro-slavery intuitions who are my peers (or superiors). That is: In order to resist (4), one must argue that the epistemic credentials of people with anti-slavery intuitions tend to be significantly better, on average, than the epistemic credentials of people with pro-slavery intuitions. But it turns out not to be possible to make that argument. No matter which definition of epistemic parity we prefer, people with anti-slavery intuitions turn out to be no more likely (on average) to be my epistemic inferiors than people with pro-slavery intuitions. Thus, we cannot find a good argument against (4). But in the absence of a good argument against (4), we should accept (4). As we have just seen, in order to reject (4) we have to claim that people in a certain group (i.e. the pro-slavery

group) tend to have poorer epistemic credentials than people in another group (i.e. the anti-slavery group). The burden of proof is on anyone who makes a claim of this form, because in the absence of reasons to think otherwise, we should assume that the members of any two groups have equally-good epistemic credentials; otherwise, we would be guilty of a certain kind of (epistemic) prejudice.

There are three main points at which the preceding quick-and-dirty argument for (4) can be criticized. First, there is the argument's claim that, in the scope of human history, the number of people who have or have had pro-slavery intuitions is greater than the number of people who have or have had anti-slavery intuitions. It is hard to see how one could verify this claim; in any event, this claim is not obviously true, and there are reasons to think it is false (as I will show below). Another point for criticism is the argument's claim that the burden of proof is on those who wish to deny (4). Indeed, there is some question whether it ever makes sense to say that one side or the other in a given dispute has the "burden of proof."[30] But set these two points to one side, for the moment. A third point for criticism is the argument's claim that, no matter which definition of epistemic parity we prefer, people with anti-slavery intuitions turn out to be no more likely (on average) to be my epistemic inferiors than people with pro-slavery intuitions. I want to examine the case against this claim in some depth, with reference to each of the three definitions of epistemic parity discussed in the previous section.

Suppose we accept Definition 2.I, the definition of epistemic parity in terms of epistemic virtues and vices. Can we show that people with pro-slavery intuitions tend to lack certain key virtues possessed by those with anti-slavery intuitions?

Degree of consistency between one's intuitions might be thought to be an epistemic virtue.[31] For example, it might be that there is greater consistency between my anti-slavery intuition and my other intuitions than there typically is between the pro-slavery intuitions of my pro-slavery opponents and their other intuitions. If this were true, it would go a long way toward impugning the epistemic credentials of those with pro-slavery intuitions. But what would it mean for one intuition to be inconsistent with

[30] A valuable discussion of this question occurred on the blog *PEA Soup*, in the comments to a post by Michael Cholbi: http://peasoup.typepad.com/peasoup/2009/07/shifting-burdens-of-proof-in-practical-ethics.html

[31] The following discussion has been significantly influenced by insights provided by an anonymous reviewer.

another intuition? Certainly, if I intuit that *A* is wrong, and at the same time I intuit that *A* is permissible, then my intuitions are inconsistent. But this will not help much, simply because this sort of intuitional inconsistency is probably quite rare – even among those who intuit that slavery is permissible in the Typical Case.

Another possibility is that inconsistency arises when one is unable to identify relevant differences to explain divergent intuitions. For example, imagine a slave-owner who intuits that it would be morally wrong for him to enslave his own mother but also intuits that it is morally permissible for him to enslave some particular woman – a woman who happens to be one of his slaves – even though he is unable to identify a relevant difference between his mother and the slave. I suspect that most slave-owners would have found themselves in this kind of position. Does this mean most slave-owners exhibited the vice of inconsistency? Unfortunately, no. Such a slave-owner's position is fully consistent. To see this, recall the two trolley cases discussed above: Standard Case and Push Case. Many people intuit that inaction is required in Push Case but not required in Standard Case, despite being unable to find a relevant difference between the two cases, even after expending great effort looking for such a difference. It seems clear that such people's intuitions are not inconsistent.[32] (Perhaps these people are criticizable on other grounds, but it is not obvious what those grounds might be.)

A third possibility is that inconsistency arises when one has divergent intuitions about cases one believes to be similar in all relevant respects. For example, suppose the slave-owner intuits that enslaving his mother is wrong, and intuits that enslaving the slave is permissible, while at the same time believing

[32] It might be claimed that this analogy fails because, even if some people cannot quite explain why they think Standard Case and Push Case differ morally, most people are able to offer (or at least gesture in the direction of) a few *potentially* relevant differences between Standard Case and Push Case. But first of all, this sets a very low bar; I suspect that most slave-owners would be able to offer *potentially* relevant differences that *might* explain their divergent intuitions about the morality of slavery. Secondly, it seems to me that consistency simply does not require us to clear even this very low bar. Suppose Leroy has been thinking about Standard Case and Push Case for some time. Leroy has the usual intuitions about these cases. When Leroy first learned about these cases, he had a few half-formed ideas about why these cases differ morally, but he has found that none of his initial ideas stand up to scrutiny (so he has abandoned them) and he has not come up with any new ideas. Leroy suspects that, if he were to continue thinking, he might find a relevant difference between Standard Case and Push Case – but now Leroy has to leave for work, and he will not be able to think about these cases again anytime soon. Suppose that Leroy continues, after this point, to believe that inaction is required in Push Case but not required in Standard Case. Does this make Leroy inconsistent? Clearly not, I think.

that there is no relevant difference between enslaving his mother and enslaving the slave. I grant that this would be a genuine form of inconsistency and a serious epistemic vice, but I wonder how many people with pro-slavery intuitions are really guilty of this form of inconsistency. My suspicion is that not very many are.[33] Thus I think it will be difficult to show that people with pro-slavery intuitions tend in large numbers to exhibit vicious inconsistency.

Perhaps *familiarity with relevant brute facts* is an epistemic virtue. For instance, suppose you believe that people of African descent have the intelligence and emotional capacities of chickens or some other farm animal, and (in large part because you hold this mistaken belief) you intuit that slavery in the Typical Case is permissible. Then it would seem very strange to think you are my epistemic peer (assuming I do not have similarly mistaken beliefs).

I grant that peers ought to have similar levels of familiarity with the relevant brute facts. But I am not sure that people with pro-slavery intuitions usually do worse, on this score, than people with anti-slavery intuitions. Consider that many of the people who have had pro-slavery intuitions lived in slave societies, so either owned slaves themselves, worked with slaves, or had seen and heard about slaves. I conjecture that wildly false beliefs about slaves would not survive very well in such an environment.[34] For example, in a slave society like the antebellum South, it would be well-known that slaves can do

[33] In claiming this, it might be thought that I am suggesting that a typical slave-owner would be a Hare-style fanatic, who would be willing to be enslaved if his own enslavement turns out to be required by whatever principle he believes justifies his enslavement of others. But I have not suggested this (and do not believe it). I suspect that a typical slave-owner would intuitively judge his own enslavement to be wrong. And I doubt that a typical slave-owner would be able to provide any defensible principle that would permit him to enslave others but would forbid others from enslaving him. But (for reasons I have already discussed) I do not see that this would make him inconsistent. Inconsistency *would* arise if a slave-owner endorses a principle that unavoidably implies either that his enslavement of others is wrong, or that others' enslavement of him would be permissible. However, I suspect that almost any principle endorsed by a slave-owner that *appears* to have such implications could be shown not have such implications. For example, Jefferson's phrase "All men are created equal," which is often taken to represent a moral principle of some kind, and which has been endorsed by many slave-owners, does not seem to me to have any interesting implications at all (unless supplemented by auxiliary principles). But I do not have the space to make the case for this view here.

[34] When I speak of a "wildly false" non-moral belief about slaves, I primarily have in mind a belief *B* such that, if *B* were true, then slavery would (by our lights) be morally defensible. For example, if it were the case that human slaves have the intelligence and emotional capacities of farm animals, then slavery might be defensible (although even this is not entirely clear). My conjecture is that this sort of belief would not be held by the vast majority of those who live in a slave society. See note 35.

things like marry and speak English; such facts would prove to anyone with common sense that slaves have high levels of intelligence and emotional capacities. Of course, there can be little doubt that people living in slave societies would have had many false beliefs. Even so, it seems probable that a person living in such a society would have been aware of most of the brute facts that we would offer as reasons why slavery is wrong. That is: I believe that I could explain why slavery is wrong, to someone from such a society, in terms of beliefs widely held in such a society.[35] Indeed, people in slave societies might have been privy to relevant facts unknown to us: consider that a slave-owner would have personally witnessed many of the cruelest effects of slavery, whereas I can only dimly and imperfectly imagine these facts. Thus, it is at least not obvious that people with anti-slavery intuitions would normally have better familiarity with the relevant brute facts than people with pro-slavery intuitions.

Let us now turn to reliability. Suppose it turns out that the intuitions of people in the anti-slavery camp are generally more reliable than the intuitions of people in the pro-slavery camp. This could provide the basis for a very strong case against (4). Indeed, this would provide for an *airtight* case against (4) if we were to accept the reliability approach to epistemic parity exemplified in Definition 3.I. Further, the case against (4) might also be strong if we were to accept the epistemic-virtues approach of Definition 2.1, because (as we saw in Section 3) the fact that one person's intuitions are more reliable than another's can, on some views, show that the first person is more epistemically virtuous (and less vicious) than the second person.

How could we support the view that the intuitions of people in the anti-slavery camp are generally more reliable than the intuitions of people in the pro-slavery camp? Here is an argument that seems too easy. The people in the pro-slavery camp intuit that slavery is permissible. But slavery is obviously wrong. Getting such questions wrong is strong evidence of unreliability. By contrast, the

[35] This is to say that I could simply grant all of a typical slave-owner's non-moral beliefs, including his mistaken beliefs, and still have sufficient reason to believe slavery is wrong. For example, it appears that in the antebellum South it was widely believed that people of African descent are innately childlike intellectually and emotionally. Even if I were to grant this mistaken belief, I would still have good reasons to consider the enslavement of people of African descent to be wrong. (These are the same reasons why I would say it is wrong to enslave anyone who is childlike – e.g. mentally disabled people, actual children, etc.)

intuitions of people in the anti-slavery camp are obviously correct (at least when it comes to the ethics of slavery). Thus we have evidence that the intuitions of those in the pro-slavery camp are less reliable than the intuitions of people in the anti-slavery camp.

This argument seems too easy because it looks like it must somehow beg the question. But question-begging is not always bad.[36] For instance, this argument begs the question against those who believe that slavery is permissible; but I would think we should have little interest in addressing such people, so it is not bad to beg the question against them. However, the argument might beg the question against the anti-intuitivist. After all, we are presently considering ways to attack premise (4) of the anti-intuitivist argument from disagreement. The conclusion of that argument is that intuitivists are committed to the obviously-wrong view that slavery is permissible. So perhaps it begs the question, in a bad way, for the intuitivist to attack (4) by means of an argument that relies on the obviously-correct view that slavery is wrong.

Whether this charge of question-begging (of the bad kind) can stick to the foregoing argument is unclear. But we might be able to circumvent that charge altogether. Suppose we were to survey the intuitions of people in the anti-slavery camp about a wide variety of different moral issues – issues that have little to do with slavery. And suppose we were to find that, in each case, their intuitions conflict with the intuitions of those of us in the anti-slavery camp. This result could produce evidence of unreliability that does not assume that slavery is wrong, and therefore would not even arguably beg the question against the anti-intuitivist. Moreover, this result could be relevant even if we were to eschew the reliability approach to epistemic parity and adopt the probabilistic approach as exemplified in Definition 1.I. The fact that members of a certain group tend to have intuitions that conflict with mine might justify me in judging that the intuitions of members of this group are less likely to be correct than mine.

But I think this strategy will not work. In general, if the intuitions of people in one group routinely conflict with the intuitions of people in another group, such conflicts might be evidence that the

[36] Shafer-Landau 2009 includes valuable insights about when question-begging is and is not a mistake; some of what I say here follows him.

intuitions of members of *both* groups are unreliable, but it is difficult to see why such conflicts should be evidence that one group has less reliable intuitions than the other. Similarly, such conflicts might justify me in judging that neither group is likely to have correct intuitions very often, but it is difficult to see why such conflicts should be evidence that the intuitions of members of one group are less likely to be correct than the intuitions of members of the other group.[37]

I have just considered a few ways to argue that people with anti-slavery intuitions are less likely (on average) to be my epistemic peers (or superiors) than people with pro-slavery intuitions. I have certainly not considered *all* ways to make that argument, but I hope I have shown that there will be some difficulty in defending this position. Of course, even if that argument cannot succeed, there are other ways to undermine the quick-and-dirty argument for (4) that I gave at the beginning of this section. One could argue, for example, that pro-slavery intuitions are very unusual in history. Perhaps it is really quite rare for slavery to seem morally permissible to anyone; perhaps, in slave societies, slavery is practiced for its material and economic benefits *despite* the fact that it seems wrong to most people in such societies – or would seem wrong to them, if they would allow themselves to consider the moral question. Mundane demographic considerations are also relevant here. Due to technological advances in the 19th and 20th centuries, anti-slavery intuitions have become widespread during a period in which the human population has exploded. Thus, even if *everyone* born prior to (say) the year 1800 had had pro-slavery intuitions, it still might be that anti-slavery intuitions are more common among all the people who have ever lived. This would be enough to show that (4) is false even if people with anti-slavery intuitions are no more likely to be my epistemic peers (or superiors) than people with pro-slavery intuitions.

Thus, it is far from obvious that the quick-and-dirty argument for (4) succeeds. But it is also not obvious that it cannot succeed. Some intuitivists will be satisfied with this. They will say that, in the absence of a clear argument for (4), we should not accept it – and therefore should not accept the anti-

[37] Of course, things would be different if we had an extra-intuitional source of information about morality and could appeal to this source of information in the event of conflicts between intuitions. But the existence of such a source of information conflicts with No Alternatives. See the Introduction for my reasons to assume No Alternatives.

intuitivist argument from disagreement. But other intuitivists will be nervous. They will say that, in the absence of a decisive refutation of (4), we should worry that (4) might really be true.

Fortunately for the nervous intuitivist, one of the other premises in the anti-intuitivist argument from disagreement can be refuted – or so I will argue. The vulnerable premise is premise (3) – the Equal Weight View. In the next section I will show how this premise can be effectively undermined.

6. Against the Equal Weight View

The Equal Weight View is premise (3) of the argument considered in Section 3. Here it is again, for ease of reference:

> (3) **Equal Weight View**: There is no reason to attribute less epistemic significance to your intuition than to mine if I correctly consider you to be my epistemic peer or epistemic superior.

I will begin my case against this principle with an analogy. Suppose I am a watchmaker and I correctly consider myself to be very good at watchmaking. I have been recruited to participate in an experiment that will be conducted by Alex, a psychologist. Alex has given me a list containing tens of thousands of watchmakers, all of whom have also agreed to be in the experiment. I have read through the entire list (I am a very fast and thorough reader) and it turns out that I know all of the people on the list (I am a very well-connected watchmaker). I believe, correctly, that each watchmaker on the list is as good a watchmaker as I am (but no better). Alex tells me that our experiment will proceed as follows. On Tuesday, I and all the other watchmakers on the list will make separate watches in our separate workshops. Then, on Tuesday evening, Alex will visit each of us to collect the watches we have made. On Wednesday, Alex will return to my workshop and will present me with two boxes. In the first box will be my watch. In the second box will be a watch made by one of the other watchmakers. This is all that Alex tells me; he refuses to say more.

On Tuesday I dutifully make my watch. That evening Alex comes, as expected, and collects the watch I made. On Wednesday, Alex arrives to my workshop carrying two boxes. Alex says:

> Before I allow you to open these boxes, I will tell you how I decided what to put in them. The first box contains your watch, as planned. To select a watch for the second box, I used the following procedure. First, I arranged into a single row all the tens of thousands of watches made by the other watchmakers participating in this experiment. Then I made a guess about what time it was, and I set my wristwatch to that time. (I have previously shown experimentally that my guesses about the time are rarely exactly correct, although it is also rare that my guesses are more than a few minutes off.) Then I went down the row, one by one, looking for a watch consistent with my guess. I won't tell you how long it took me to find such a watch, but I did eventually find one. That is the watch I put into the second box.

After saying this, Alex gives me the two boxes. In the first box I find my watch. I am sure it is mine because it bears my engraved signature. It says the time is 11:56. In the second box I find a watch that says the time is 11:50. This watch bears the inscription *R. Smith*. I recognize this as Robert Smith's signature; I have high confidence in Smith's abilities and regard him as my equal when it comes to watchmaking. (In this respect Smith is no different from any of the other watchmakers who participated in the experiment.)

Alex now asks me to make a judgment about the time. Alex forbids me to consult a third watch. His forbiddance is unnecessary – I do not need any more evidence; I say confidently, without hesitation, that the current time is 11:56. Alex takes note of my answer and thanks me for my participation in the experiment. I ask him to stay for coffee, but he says he has no time. He has to visit each of the other watchmakers on the list; they are all going to be presented with the same problem.

Is my confidence that it is 11:56 justified? It seems so. Smith's watch says it is 11:50, and Smith is ex hypothesi an excellent watchmaker – every bit as good a watchmaker as I am. But Smith's watch

was selected only because it matches Alex's guess, and I should have no great confidence in Alex's guesswork. It might be argued that Smith's watch would not match Alex's guess, or would be very unlikely to match Alex's guess, unless Alex's guess were correct. This is, in one sense, true: In normal circumstances, I should think it very unlikely that Smith's watch would match Alex's guess unless Alex's guess were correct. But it does not follow from this that I should trust Smith's watch in the present unusual circumstance. After all, Smith's watch was one in a batch of tens of thousands. We might expect there to be quite a few bad watches in a batch of tens of thousands, even if each watch in the batch had been made by an excellent watchmaker. Thus, even if we assume that Alex's guess had been wrong, it would be no great surprise that Alex is able to find a watch made by an excellent watchmaker to match it. And it would be no great surprise if Smith's watch just happened to be that watch.

What is going on in this example? I have been presented with two watches. One is made by an excellent watchmaker (me); the other is made by an equally-excellent watchmaker (Smith). Aside from these two watches, I have no further evidence about the current time (e.g. I have not looked at a third watch to see whether it is consistent with my watch or with Smith's) and yet I am justified in placing much greater trust in my watch than in Smith's. I am so justified because of the way in which Smith's watch was selected by Alex. The selection of Smith's watch was, we might say, *rigged*: Smith's watch was selected from a large batch solely for its compliance with a guess in which I should have little trust. By contrast, the selection of my watch was not rigged; my watch was presented to me just because it is the watch I made, and the fact that a given watch is a watch I made is good reason for me to trust it, since I am ex hypothesi an excellent watchmaker.

It seems that similar points would hold in an example involving moral intuitions and epistemic peers. Imagine a crowd of ten thousand people, all of whom I correctly and justifiably believe to be my epistemic peers. If ten people are chosen randomly from this crowd, and all ten report pro-slavery intuitions, then perhaps this should lead me to seriously question my own anti-slavery intuitions. But suppose instead that ten people are selected from the crowd, not randomly, but by Alex, who has gone into the crowd searching specifically for ten people with pro-slavery intuitions. It would be strange to

think that, when Alex presents me with these ten people, I am given any good reason to doubt my anti-slavery intuition. This is because Alex would surely have been able to find ten people with pro-slavery intuitions as long as no more than 99.9% of the people in the crowd have anti-slavery intuitions. There is not much evidence that slavery is permissible in a demonstration that no more than 99.9% of my peers intuit that slavery is wrong. Consider also that, unless the crowd is very nearly uniform, Alex would have been able to find ten people with anti-slavery intuitions, if he wanted, just as well as he could find ten with pro-slavery intuitions. In that case, if I ought to attribute equal significance to the intuitions of anyone Alex selects from the crowd, then Alex would have control over how I ought to believe: Alex could freely choose whether I ought to relinquish my anti-slavery convictions or not simply by choosing whether or not to bring forth people with anti-slavery intuitions or people with pro-slavery intuitions. But this seems absurd.

In this example, a rigged selection process executed by Alex poisons any evidence that the pro-slavery intuitions of these ten people might otherwise have provided. Yet any evidence that might be provided by my own intuition is not poisoned in this way. It is not as if I am aware of my own intuition because Alex brought it to my attention; I am aware of my own intuition *automatically*, as it were, simply because it is *my own* intuition. This appears to be an epistemically-relevant difference between my encounter with my own intuition and my encounter with the intuitions of the ten people selected by Alex. In light of this difference, it seems that I could be entitled to attribute greater epistemic significance to my own intuition than to the intuitions of any of these ten people.

If this is right, then the Equal Weight View is false: there can, after all, be reason to attribute less epistemic significance to your intuition than to mine even if you are my epistemic peer. If my encounter with your intuition is rigged, as in the Alex example just considered, then it seems I should not attribute the same significance to your intuition as to mine; rather, I should attribute considerably less significance to your intuition than to mine.

The Equal Weight View can be modified to deal with this problem, as follows:

> **Better Equal Weight View**: There is no reason to attribute less epistemic significance to your intuition than to mine if (a) you are my epistemic peer, and (b) my encounter with your intuition is not rigged.[38]

Perhaps, as I have argued, the *mere* fact that an epistemic peer disagrees with me is not, by itself, epistemically significant. But it would seem significant if I am shown that a peer *could easily have different intuitions than mine*. If a peer could easily have different intuitions than mine, then perhaps I myself could easily have had different intuitions.[39] This would show that my intuition is *unsafe*.[40] If this is right, we may have a diagnosis of the second Alex case: the rigged presentation of a disagreeing peer is insignificant in that case because it provides no evidence as to whether a peer could easily have different intuitions than I have. But perhaps we should expect this problem would not arise in cases in which a peer is presented in a non-rigged fashion. Thus the Better Equal Weight View seems to stand a good chance of being correct.

I have no interesting objections to the Better Equal Weight View. And the Better Equal Weight View might still cause problems for intuitivism. Suppose that a historian shows that, in a random sample of all the people who have ever lived, a sizable majority would intuit that slavery is permissible in the Typical Case. It seems clear that, when I learn of this poll, my encounter with the intuitions of the people in the pollster's sample is not rigged. Then, unless we can refute premise (4) of the anti-intuitivist

---

[38] What do I mean by "rigged" here? I do not think I need to provide necessary and sufficient conditions. The Alex examples show that if I encounter a peer *simply because* her opinion is consistent with a guess made by a bad guesser, or simply because she is a peer who disagrees with me, then my encounter with her is rigged. Perhaps, then, we should say that a rigged encounter with a given intuition is an encounter that occurs because of the intuition's content. But I am not entirely satisfied with this definition, for reasons there is not space to discuss here. Alternatively, it might be suggested that a rigged encounter is any encounter that is not statistically random. But in a moment we will see reasons to think this is unsatisfactory.

[39] I am grateful to an anonymous reviewer for calling my attention to this point.

[40] *S*'s intuition that *p* is *safe* if and only if, "as a matter of fact, though perhaps not as a matter of strict necessity, *S* would not intuit that *p* without it being so that *p*." I take this concept from Ernest Sosa [1999], who defines safety in terms of *belief,* not intuition. The preceding quotation is Sosa's except for my substitution of "intuit" for "believe."

argument from disagreement, it seems the Better Equal Weight View will require me to abandon my anti-slavery convictions.

But the matter is complicated by the evolutionary history of our intuitions. I will defend two claims. First, I want to show that, given an evolutionary history of others' moral intuitions, my encounter with others' intuitions should *almost always* be considered rigged, even when they are randomly selected; this gives me a good reason to discount the significance of others' intuitions. Call this **the Other-Thesis.** Second, I want to show that an evolutionary history of *my own* moral intuitions would *not* provide reason to consider my encounter with *my own* intuitions to be rigged; thus, I am not given a reason to discount the significance of my own intuitions when I am given an evolutionary history of my own intuitions. Call this **the Self-Thesis.** If both of these theses were established, then it would follow that, given an evolutionary history of our moral intuitions, it is consistent with the Better Equal Weight View for an anti-slavery intuitivist to attribute high epistemic significance to her own anti-slavery intuitions while discounting, or even disregarding entirely, the pro-slavery intuitions of her pro-slavery peers.

Evolution is a multiple-generation selection process, and in a moment I will consider a process of this type. But I will begin my argument for the Other-Thesis and for the Self-Thesis by considering a single-generation selection process. Suppose (contrary to fact) that a single gene causes anti-slavery intuitions and a single (different) gene causes pro-slavery intuitions. Imagine two populations, Tribe A and Tribe B, living far apart from one another in separate areas. Due to a strange disease that takes hold in Tribe A's area, everyone with the pro-slavery gene suddenly dies; those with the anti-slavery gene are spared. At around the same time, a similar disease afflicts Tribe B's area, but it has a reverse effect: everyone in Tribe B with the anti-slavery gene suddenly dies; those with the pro-slavery gene are spared.

Suppose Lisa is a member of Tribe A, unaware of the existence of Tribe B. Lisa has anti-slavery intuitions, so she survives the epidemic. Suppose Lisa does not know even roughly how many members of her tribe were eliminated by the disease (because Lisa is reclusive and spends all her time in her hut). But Lisa does know that an epidemic killed all and only those with pro-slavery intuitions – and that it would have killed her, if she had not had anti-slavery intuitions. Should having gone through this

selection process cause Lisa to revise downward any epistemic significance she previously attributed to her own intuitions? It seems not. The fact that those who disagreed with Lisa have been killed does not show Lisa's own intuitions to be less trustworthy; and the fact that she was spared death only because of the content of her intuitions is not a reason for her to distrust that content. These points, I want to emphasize, do *not* show that Lisa should attribute high (or any) epistemic significance to her own intuitions; rather, they show that the selection process to which Lisa has been subjected does not give Lisa a reason to attribute less significance to her intuitions than she would have done absent that selection process. This is so, I suggest, because in this case, after the disease has had its effect, Lisa's encounter with her own anti-slavery intuitions is *permitted* by a selection process – but it is not *rigged* by that selection process.

Now suppose that, shortly after the disease has had these effects on Lisa's tribe, Lisa takes a long walk and discovers the previously-unknown Tribe B. Lisa learns about the effect of the disease that afflicted Tribe B: anyone in Tribe B who had anti-slavery intuitions is now dead, and everyone still living has pro-slavery intuitions. (No one will tell her how many members of Tribe B were killed by the disease, because Tribe B has a rule against counting dead people.) In that case it seems Lisa should attribute little or no epistemic significance to the intuitions of the still-living members of this tribe. This is so even if we assume that Lisa considers Tribe B to be a more enlightened society than her own and therefore considers the members of Tribe B to be her epistemic superiors. Lisa has, in that case, found that 100% of a group of her superiors have pro-slavery intuitions. But Lisa can know that she would have found this even if only 1% of the members of this enlightened society had had pro-slavery intuitions at the moment before the disease took effect. Lisa finds pro-slavery intuitions prevalent in this society *just because* a disease ensured she could find nothing else. So, although Lisa considers these tribe-members to be her superiors, her encounter with their intuitions is rigged, not just permitted, by a selection process. Given this, it seems that Lisa should attribute little or no significance to their intuitions.

Thus, in the example of these single-generation selection processes, there appears to be a self-other epistemic asymmetry: a selection process permits, but does not rig, Lisa's encounter with her own

intuition; but a precisely similar selection process rigs Lisa's encounter with others' intuitions; thus Lisa is a provided reason to regard others' intuitions as epistemically insignificant without being provided reason to regard her own intuitions as epistemically insignificant. Next I will argue that we find a similar asymmetry when we consider multiple-generation selection processes. But first I want to briefly mention an objection to my argument about Lisa. Note that, if my argument goes through, then a similar argument would work equally well for a person from Tribe B. Imagine that Rick, a person from Tribe B with pro-slavery intuitions, meets Lisa. Suppose that Rick and Lisa both know about the selection processes to which each has been subjected, and Rick and Lisa agree on all the relevant facts about these selection processes. I am committed to say that, even after agreeing about all these facts, Rick could be justified in attributing significance to his own intuition but not Lisa's, while Lisa could be justified in attributing significance to her own intuition but not Rick's. Thus, after Rick and Lisa meet, Rick could be justified in continuing to believe that slavery in the Typical Case is permissible, while Lisa could be justified in continuing to believe that slavery in the Typical Case is wrong. But this seems odd.

Of course, there is nothing particularly odd about two people being justified in believing differently. I may be justified in believing the world is round even if a person who has been subjected to misleading evidence, e.g. someone raised in a family of Flat-Earthers, is justified in believing the world is flat. But there is something odd about a case in which two people can be justified in believing differently *after having come to an agreement about all the relevant facts*. Such a case would apparently be a counterexample to the Uniqueness Principle, which says that "given one's total evidence, there is a unique rational doxastic attitude that one can take to any proposition" (White 2005). Given the Uniqueness Principle, if two individuals have all the same evidence bearing on a given proposition, then they cannot rationally disagree about that proposition. It seems that, in the case just described, Rick and Lisa have all the same evidence – yet, I claim, they could still rationally disagree about whether slavery is wrong. This implication is counterintuitive, but it is defensible. Different parts of this shared body of evidence have been learned by Rick and Lisa in different ways. Rick and Lisa both know that Rick has a pro-slavery intuition, but Lisa learned this fact in a rigged fashion, whereas Rick did not. Likewise, Rick

and Lisa both know that Lisa has an anti-slavery intuition, but Rick learned this fact in a rigged fashion, whereas Lisa did not. I have argued that this is an epistemically-relevant difference. If that argument has been successful, then it explains why this case provides a counterexample to the Uniqueness Principle.

Let us now consider a multiple-generation selection process. Suppose that, after her visit to Tribe B, Lisa returns home to Tribe A. She has children, grows old, and dies. The disease afflicting Tribe A is never cured and never dies out. Thus, generations later, Lisa's descendents, along with all other members of Tribe A, have anti-slavery intuitions. Rob is one of Lisa's descendents. Suppose Rob knows that he comes from a long line of people with anti-slavery intuitions, and knows that his lineage has been able to continue because the disease that afflicts his tribe only kills those with pro-slavery intuitions. Should this lead Rob to consider his anti-slavery intuitions to be untrustworthy? It seems not. Perhaps it will be argued that the mere fact that Rob has inherited a genetic disposition toward anti-slavery intuitions should make Rob distrust those intuitions. But I have already responded to this view in Section 2, where I argue that it will be difficult to show that facts about one's intuition-generating mechanism provide reason to distrust one's intuitions. Perhaps it will be argued that Rob should distrust his intuitions, not just because those intuitions are due to an inherited disposition, but because this inherited disposition is the result of a selection process. But we already considered an argument like this – and it has not gotten any stronger in the generations that separate Rob and Lisa. The fact that, among Rob's tribe, pro-slavery lineages have been wiped out (or been prevented from existing in the first place) by a disease is not a reason for Rob to distrust the anti-slavery intuitions prevalent in his own lineage. Likewise, the fact that Rob's own lineage has been unaffected by the disease, and thus has been able to thrive, is not a reason for Rob to distrust the anti-slavery intuitions prevalent in his own lineage. Thus Rob's inheritance of anti-slavery intuitions is merely permitted, not rigged, by a selection process.

Suppose now that Rob takes a long walk and finds Tribe B, still in the same place it had been in Lisa's day. Rob finds that Tribe B continues to be afflicted by a disease that kills people with anti-slavery intuitions, so all the people of Tribe B continue to have pro-slavery intuitions. How much significance should Rob attribute to the intuitions of the people he encounters in Tribe B? Not much, I argue, for

reasons identical with those that applied generations ago: Rob finds pro-slavery intuitions prevalent in this tribe just because a disease ensured he could find nothing else; thus his encounter with the tribe-members' intuitions is rigged, not just permitted, by a selection process.

The moral of this story is as follows. The multiple-generation selection processes I have just considered are obviously far simpler than any intuition selection process that might occur in the actual world. Our moral intuitions regarding slavery are not caused by a single gene; rather, they are probably the result of a complex interplay between genetic, cultural, idiosyncratic, and other factors. Further, if (as seems probable) forces of biological and cultural evolution shape many of these factors, those forces are certainly more complicated than the disease described in this story. But our real-world predicament may be *epistemically* similar with Rob's predicament. If my own anti-slavery intuition is the result of a cultural or biological evolutionary selection process, and if others' pro-slavery intuitions are the result of different evolutionary selection processes, this may give me, like Rob, a reason to deny epistemic significance to others' pro-slavery intuitions without giving me a reason to deny epistemic significance to my own anti-slavery intuitions.[41] Then the Other-Thesis and the Self-Thesis would be true. And in that case, even if I consider *everyone* with pro-slavery intuitions to be my epistemic peers, and even if I believe that, in the scope of human history, pro-slavery intuitions vastly outnumber anti-slavery

[41] Here it is worth mentioning that some pro-slavery intuitions are probably the result of the same selection process that has issued my anti-slavery intuitions. For instance, suppose that cultures are subject to selection processes, and suppose that my anti-slavery intuitions are the result of my immersion in a particular culture. Some members of my culture have pro-slavery intuitions; e.g. there are Ku Klux Klan members, who are arguably part of my culture, yet who seem to have pro-slavery intuitions. If these Klan members are part of my culture, then my encounter with their intuitions is not analogous with Rob's encounter with the intuitions of Tribe B, so I have not provided any reason to think that my encounter with Klan members' intuitions is rigged. Thus it may be that I ought to attribute equal significance to Klan members' pro-slavery intuitions as to my own anti-slavery intuitions. I think this particular result would not be very worrying simply because Klan members are rare; I believe that the vast majority of the members of my own culture have anti-slavery intuitions. Therefore, even if it turns out that I ought to attribute as much significance to the intuitions of each member of my culture as I attribute to my own intuitions, I think I would still be justified, on intuitivist grounds, in maintaining an anti-slavery position. But this is just one example illustrating a more general concern. For all I have said here, it might turn out that the selection process that issues my anti-slavery intuitions (whatever that process turns out to be) usually issues pro-slavery intuitions. I do not have space here to determine whether this is so. The issue is partly empirical, but also partly conceptual (since it depends on how we ought to individuate selection processes). So I have to concede that, if it turns out that the selection process that issues my anti-slavery intuitions usually issues pro-slavery intuitions, then it is at least possible that a compelling anti-intuitivist argument is available.

intuitions, I could still be an intuitivist without being committed to abandon my anti-slavery convictions. The anti-intuitivist argument from disagreement would then fail.

Conclusion

The question with which I began is whether we can give good reasons to think that trusting one's intuitions is no better than tossing a coin. I do not think we have found such reasons. But that may be because I have looked in the wrong places, or because I have looked in the right places but have overlooked the reasons there. Of course, even if I were to have conclusively shown that there are no good reasons to think that trusting one's intuitions is no better than tossing a coin, that may seem to be a small accomplishment. After all, it might be that there are no good reasons to think that trusting one's intuitions *is* better than tossing a coin. In that case we would have no reasons for or against; and in general, when one has no reasons for or against believing something, one had better not believe it. Still – I believe that there are reasons to think that trusting intuitions is better than tossing a coin (although I have not discussed any of those reasons here); and so I will be pleased if it turns out there are no reasons against that view. It is also worth pointing out that, even if one were to conclusively show that trusting one's intuitions is better than tossing a coin, it still might be that trusting one's intuitions is only *very narrowly* better than tossing a coin. I worry about this. My view (which I have not defended here) is that No Alternatives is true; I think we have no alternative (other than agnosticism about morality) to trusting our intuitions. So I think we should hope that our intuitions are as trustworthy as they can possibly be. However, I suggest that, somewhat paradoxically, it is far better for our intuitions to be narrowly better than coin-tossing, than it is for our intuitions to be no better than coin tossing. If our intuitions are no better than coin-tossing, then we do not even know how to aim at doing the right thing. If our intuitions are narrowly better than coin-tossing, then our aim might not be very good – but at least we will know how to aim.

Non-Naturalist Moral Perception

Overview

According to the causal constraint on perception, in order for S to perceive that P, there must be causal contact between S and P. According to standard moral non-naturalism, moral facts are causally inert. So, it seems, moral non-naturalism is incompatible with the hypothesis that moral facts are perceivable. Justin McBrayer defends non-naturalist moral perception by attacking the causal constraint. As I understand him, he thinks that the causal constraint should be replaced by what I call the *reliability constraint*, which requires only that perception be generated by a reliable mechanism. In §1 and §2, I attack the reliability constraint. Then, in §3, I attack the causal constraint. Finally, in §4, I argue in favor of a different constraint on perception: the *explanatory relatedness* constraint. This constraint makes room for non-naturalist moral perception. So, like McBrayer, I defend non-naturalist moral perception, but I believe that my defense improves on his.

Introduction

A subject S experiences a *moral intuition* when P seems true to S, and P is a moral proposition. S experiences a *moral perception* when S perceives that P, and P is a moral fact (i.e., a true moral proposition). Perception is a type of seeming; thus, moral perception is one of many conceivable types of moral intuition.[42] Justin McBrayer has recently argued in defense of the possibility of non-naturalist moral perception.[43] I agree with McBrayer that non-naturalist moral perception is possible, but I think that his argument is vulnerable to an important objection. In this paper, I will explain what I think is wrong with McBrayer's argument, and then will propose an alternative line of argument for the same conclusion.

[42] There has been some resistance to this classification. Robert Audi, for instance, explicitly distinguishes "intuitive moral seemings" from "perceptual moral seemings" (See Audi 2010 p. 90.) But if we use "moral intuition" to refer to the broad category of moral seemings, as I intend to do here, then I see no good reason not to include moral perception as a type of moral intuition.

[43] McBrayer 2010, esp. section IV.

Moral non-naturalists deny that moral facts are natural facts,[44] which seems to commit them to the view that moral facts lie outside of the causal order. (There is no logical incoherence in the view that non-natural facts can be causally efficacious; but I will assume they cannot be, at least for the purposes of this paper.) This makes non-naturalist moral perception (i.e., a mode of moral perception that is consistent with moral non-naturalism) especially difficult to defend, since it is unclear how something that lies outside of the causal order could be perceivable. This worry is expressed in the following argument:

> (1) For S to perceive that P, S must be in appropriate causal contact with the fact that P (the causal constraint on perception).
>
> (2) Human subjects are never in appropriate causal contact with moral facts.
>
> (3) So, moral perception is impossible (for human subjects).[45]

McBrayer believes that non-naturalists ought to grant (2). So, he thinks that defenders of non-naturalist moral perception ought to deny (1), the causal constraint on perception. I agree. But I have qualms about his case against the causal constraint. In Section 1, I summarize McBrayer's argument against the causal constraint. In Section 2, I criticize his argument. In Section 3, I lay out an alternative argument against the causal constraint. In Section 4, I suggest an alternative to the causal constraint which allows for the possibility of non-naturalist moral perception. In the Conclusion, I discuss the significance of my results.

1. McBrayer's argument against the causal constraint

Perception has at least two essential features. First, perception is essentially veridical: I cannot perceive that P if P is false. (If P is false, yet I have a perceptual experience as if P is true, then my experience is an

[44] Needless to say, this is not a very informative definition. I will not attempt a better definition, since my purpose here does not require it. A clear example of the sort of non-naturalism that I have in mind is laid out and defended in Shafer-Landau 2005.

[45] This is a slight modification of the argument presented by McBrayer. McBrayer prefers to talk about perception of *properties*, whereas here I prefer to talk about perception of *facts*.

illusion or a *hallucination*—not a genuine perception.[46]) Second, perception is essentially immediate: I cannot perceive that P by *inferring that* P is true. (If I infer that P is true, I might say, "I see that P is true," but this is a non-literal usage of perceptual language.)

But a given experience can be both immediate and veridical without being a genuine perception. To show this, McBrayer considers the case in which a desert traveler hallucinates an oasis with his eyes closed—and by pure coincidence, there is a real oasis in front of him.[47] If the hallucinatory oasis happens to look just like the real oasis, then the traveler's experience is veridical as well as immediate, but it is does not count as a genuine perception.

The causal constraint, as expressed in premise (2), can explain why this is so. The fact that there is a real oasis in front of the desert traveler is causally unrelated to the traveler's experience—since the traveler would presumably have had the same hallucination even if there were no oasis in front of him. It is just a coincidence that the traveler's experience matches reality.

Thus, the causal constraint has an explanatory advantage: It explains why the desert traveler does not perceive the oasis. And by considering a range of other, similar cases, we might be able to assemble a fairly persuasive case in favor of the causal constraint. But this argument still leaves open the question of *why* we are inclined to accept the implications of the causal constraint in the desert traveler case and similar cases. Provided that the traveler's experience of the oasis is veridical, why should it matter whether his experience is causally related to the real oasis in front of him?

One possibility, which McBrayer favors, is that perceptual experiences are generally more *reliable* when they satisfy the causal constraint. The desert traveler's hallucination is veridical, but only by accident; the next time he experiences a hallucination, he probably will not be so lucky. By contrast, if the traveler had perceived the oasis in the usual way—by looking at it, absorbing light reflected from it to his retinas, and converting that light into electromagnetic impulses interpreted by his brain—then it would

[46] The view that all perception is veridical, and the attendant distinction of perception from illusion and hallucination, is not universally accepted. Some people hold that illusion and hallucination are just varieties of perception. I use "perception" in this paper as a success term, to refer to what these people would call "veridical perception."

[47] See McBrayer *Ibid.* p. 301. The example appears in Pears 1976.

have been no accident that his perceptual experience accurately represents the oasis as it really is. Ordinary perception of physical objects, such as oases, satisfies the causal constraint—and we value this because satisfaction of the causal constraint ensures the reliability of perception.

This consideration might seem to support the causal constraint by providing it with a clear rationale. However, McBrayer argues that this consideration actually undermines the causal constraint, since it turns out that perceptual experiences do not have to satisfy the causal constraint in order to be generated in a reliable way. So, if our principal aim in distinguishing perceptions from other veridical experiences is to distinguish reliably-generated experiences from unreliably-generated ones, then we ought to endorse a *reliability* constraint, and dispense with the causal constraint. A suitable reliability constraint might look like this:

> (1') For S to perceive that P, S's perceptual experience of P must be generated by a reliable mechanism. An experience-generating mechanism is reliable if experiences generated by it are *usually* veridical.

Thus, by explaining the appeal of the causal constraint in terms of reliability, we can motivate rejection of the causal constraint in favor of a reliability constraint such as (1').[48]

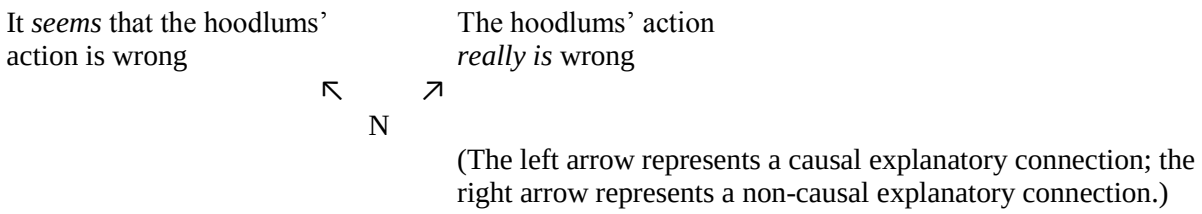
McBrayer's next move is to observe that the reliability constraint is fully compatible with the possibility of non-naturalist moral perception. To see why this is so, take a standard case: you see a gang of hoodlums pour gasoline on a cat and ignite it, and immediately it seems to you that what they are doing is wrong.[49] In order for the causal constraint to be satisfied in this case, there would need to be some sort of causal interaction between the fact that the hoodlums' action is wrong, on the one hand, and the fact

[48] Here I have presented McBrayer's view slightly differently from the way that he himself presents it. McBrayer suggests that we reject the causal constraint in favor of what he calls an "emended" causal constraint. The emended causal constraint, however, allows for the perception of causally impotent properties, such as moral properties as construed by moral non-naturalism, provided that the process by which they are generated is "non-accidental" or reliable. So, it seems to me, it is better not to call the new constraint a *causal* constraint at all—it is really a reliability constraint.

[49] Harman 1977 ch. 1.

that the action *seems* to be wrong, on the other. This is ruled out by (2), which, we are assuming, non-naturalists are committed to accept. Thus, if we accept the causal constraint, then non-naturalists have to deny that moral perception occurs in the cat-burning case.

Nevertheless, it is entirely possible for your moral intuition, in the case of the hoodlums, to be generated in a reliable way. Imagine the following. Suppose that the hoodlums' action has a number of morally relevant natural properties (e.g., their action causes the cat to be in pain; it is deliberate; it is pleasurable for the hoodlums—etc.). Call this set of natural properties N. Suppose that N has a dual explanatory role. First, N explains why the hoodlums' action *seems to be* morally wrong to the observer of their action. This explanatory connection is causal. Second, N explains why the hoodlums' action *really is* morally wrong. This explanatory connection is non-causal, and obtains in virtue of the supervenience of moral properties on natural properties. This scenario is represented in the following diagram:



In this case, the observer experiences the hoodlums' action as morally wrong; this experience is veridical, and its veridicality is no mere coincidence. Now suppose that moral intuition is usually generated by this type of process. That is: When we intuit that P, this intuition is usually generated by our recognition of a set of natural facts, and these very natural facts make it the case that P is true. I call this the "Fork Model" of moral intuition.[50] If the Fork Model were true, then our moral intuitions would be generated by a reliable process. Nevertheless, our moral intuitions would fail to satisfy the causal constraint. Moreover,

[50] McBrayer does not use the term "Fork Model," but my understanding of his view (as it is laid out in J. McBrayer *Ibid.* section IV) implies that he would endorse what I am calling the "Fork Model" as a viable model of non-naturalist moral perception.

the Fork Model is fully consistent with moral non-naturalism.[51] This shows that if we shift from the causal constraint to a reliability constraint, then we make room for the possibility of non-naturalist moral perception.

Let me sum up. Here, in a nutshell, is how I believe that McBrayer wants to respond to the causal objection to non-naturalist moral perception. In distinguishing between perceptions and other forms of seeming states, we aim to distinguish *reliably produced* seeming states from other kinds of seeming states. This makes the causal constraint seem initially attractive. Satisfaction of the causal constraint is one way to ensure reliability. But satisfaction of the causal constraint is not the *only* way to ensure reliability. Given this, we ought to reject the causal constraint, and replace it with a reliability constraint. But once we do this, we make room for non-naturalist moral perception. So, non-naturalist moral perception cannot be ruled out by the causal objection. Indeed, non-naturalist moral perception is consistent with the *underlying rationale* for the causal constraint.

2. Problems for McBrayer

We have seen that, according to McBrayer's response to the causal objection, we endorse the causal objection primarily because we think that genuine perception is generated by a reliable process. I am not convinced that he is right about this.

Consider the following case. You are travelling through Fake Barn Country, where most of the apparent barns are mere facades, but a few of the apparent barns are real barns. You are sleeping in the passenger seat of a car, while your friend drives. Then you wake up, glance out the window, and see what you believe to be a real barn. (You are unaware that you are in Fake Barn Country.) It turns out that you are correct. By chance, you have laid eyes on one of the few real barns in Fake Barn Country.

[51] The main commitment of the Fork Model, with regard to moral metaphysics, is that natural facts are *somehow* explanatorily responsible for moral facts. This is compatible with moral non-naturalism. For instance, it is compatible with standard non-naturalist accounts in which the moral supervenes on the natural. (By the same token, the Fork Model does not *require* the supervenience of the moral on the natural; it only requires an explanatory relation between the moral and the natural.)

According to a popular view, which I accept, you do not *know* that what you saw was a barn (i.e., a *real* barn, as opposed to a fake one). Nevertheless, when you look out the window and see the barn, you do *see* the barn. Further, if you were to look out the window, see the barn, and say, "I see that there is a (real) barn on the hillside," I think you would not be mistaken. Nevertheless, your perceptual experience of the barn is not generated by a reliable process: If you see an apparent barn in Fake Barn Country, you are probably looking at a fake barn, not a real one.[52]

This case suggests that reliability is not required in order for a perceptual experience to count as a perception. And this, in turn, tells against McBrayer's idea that we rely on the causal constraint simply in order to distinguish reliably generated perceptual experiences from other kinds of perceptual experiences.

Consider another case. Andy believes that everyone has a spirit animal. He believes this because, when he looks at any given person, it just *seems* to him that that person has a spirit animal. Andy experiences these seemings because he has a brain tumor. And, as it happens, Andy's seemings are veridical. Everyone actually does have a spirit animal. But no one (aside from Andy) believes this, because spirit animals are non-natural, and therefore invisible, odorless, intangible, and so on.[53]

Andy's seemings are generated by a reliable process: Whenever it seems to Andy that someone has a spirit animal, that person actually does have a spirit animal. Yet it seems quite clear that when Andy looks at a given person, Andy does not *perceive* that the person has a spirit animal. This judgment, taken by itself, is compatible with McBrayer's view that reliability is *necessary* for perception. However, given that Andy's seemings are reliable, we need an explanation of why his seemings do not count as perceptions. If the point of the causal constraint is simply to distinguish reliable experiences from

[52] Here it might be objected that the process by which your perception is generated *is* reliable. It is only unreliable in a highly unusual place like Fake Barn Country. The fact that a process can generate non-veridical experiences in highly unusual circumstances does not show that process to be unreliable *on the whole*. So let us consider a different case. Suppose you are in a world in which 60% of visible objects are fake, although somehow you have never discovered this. So, for instance, when you see an apparent chair, you usually take yourself to have laid eyes on a real chair, but this is actually true only about 40% of the time. In this world, the process by which your experiences are generated would be unreliable—not just in unusual circumstances, but on the whole. So, if the reliability constraint were correct, then genuine perception would never occur in such a world. Yet this implication seems false to me.

[53] This combines a pair of examples considered in Bedke 2009 pp. 197-199.

unreliable ones, as McBrayer contends, then it is puzzling that we do not count Andy's seemings as perceptions. The reliability constraint, taken alone, offers no answer to this puzzle.

By contrast, the causal constraint easily solves this puzzle. The proponent of the causal constraint can say: There is no causal connection between Andy's experiences and the spirit animals he thinks he perceives, so *of course* Andy does not perceive the spirit animals. Further, the causal constraint is consistent with the judgment that genuine perception of barns can occur in Fake Barn Country. The proponent of the causal constraint can say: There is a causal connection between the barn and the perceptual experience of the person looking at barn, so it is no surprise that the person looking at the barn genuinely perceives it. McBrayer's position, by contrast, is incompatible with the judgment that genuine perception of barns can occur in Fake Barn Country.

These points raise a basic question. If the veridicality and the immediacy of an experience are not sufficient to make that experience a genuine perception, then a further ingredient is required. McBrayer's position seems to be that reliability is the required further ingredient. But the Fake Barn Country Case and the case of Andy suggest that he is mistaken to think so. But then what *is* the further ingredient needed to make a veridical and immediate experience into a perception?

Perhaps this further ingredient is provided by the causal constraint after all. That is, perhaps a given veridical and immediate experience can count as a genuine perception only if it satisfies the causal constraint. If this were correct, then the prospects for non-naturalist moral perception would be quite dim. However, I think that I can provide a convincing case against the causal constraint without resorting to considerations of reliability. I will attempt to do so in the next section.

3. Counterexamples to the causal constraint

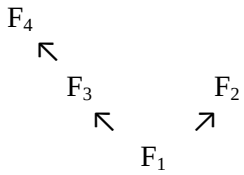
I am going to offer a couple of cases that I believe serve as counterexamples to the causal constraint. Then, in the next section, I will discuss their significance for the possibility of non-naturalist moral perception. (Both cases are purely fictional—so far as I know.)

> *First case*. There is a certain type of hawk whose prey is a species of bat. The bat flies so speedily that whenever the hawk looks at the bat in flight, the bat is already in a different location before the hawk's visual system can fully process the visual input from the bat. In the past, this used to prevent the hawk from catching the bat. But then the hawk's visual system evolved an adaptation to solve the problem. Now, whenever the hawk receives visual input of this type of bat, the hawk's brain automatically adjusts for the bat's motion. Thus, if the bat is at location X at time T, the hawk's visual system makes it *seem* to the hawk that the bat is at a different location Y. But the process of generating this seeming is not completed until T+2—and at T+2, the bat will usually have reached location Y. Consequently, when the hawk looks at the bat, the bat is usually precisely where it seems to the hawk to be. This adaptation allows the hawk to effectively hunt the bat.

In this case, there are a number of facts in play:

> F_1: At T, the bat is at location X.
>
> F_2: At T+2, the bat is at location Y.
>
> F_3: At T, the hawk receives visual input from the bat at location X.
>
> F_4: At T+2, it seems to the hawk that the bat is at location Y.

These facts are causally related in various ways. F_1 is causally related to F_2. (To causally explain why the bat is at Y at T+2, we would need to make reference to the fact that the bat was at X at T.) And F_1 is causally related to F_3. (To causally explain the hawk's visual input at T, we would need to make reference to the bat's location at T.) F_3, in turn, is causally related to F_4. (To causally explain the hawk's seeming state at T+2, we would need to make reference to the hawk's visual input at T.) The diagram below represents these connections (all arrows represent causal explanatory connections):



As the diagram shows, F_2 and F_4 have no direct causal interaction. That is: In order to causally explain why the bat *seems* to the hawk to be at Y at T+2, we need not (and should not) include mention of the fact that the bat *really is* at Y at T+2. Put differently, there is no causal contact between the hawk and the facts about the bat's actual location at time T+2. Given this, if the causal constraint were true, then we ought to deny that the hawk really perceives that the bat is at location Y at time T+2.

This implication, I think, is mistaken. It seems that the hawk has a genuinely *perceptual* system—the hawk is not simply subject to veridical hallucinations whenever it looks at bats in flight. Further, it seems, the outputs of this perceptual system are seeming states such as the one experienced by the hawk at T+2. So the case of the hawk appears to be a counterexample to the causal constraint.

Here is another case to consider:

> *Second case*. A projector projects an image onto the main screen of a movie theater. Simultaneously, the projector sends a signal to a monitor that is located in a small viewing room. The image displayed on the monitor in the viewing room is always the same as the image displayed on the main screen of the movie theater. You are the theater manager. You use the viewing room to check whether the correct movie is being shown to theater customers. You look at the monitor in the viewing room and see credits rolling.[54]

This scenario is illustrated in the diagram below (both arrows represent causal explanatory connections):

[54] I owe this example to conversation with Jonathan Lang.

It *seems* (when looking at the viewing room monitor) that credits are rolling on the main screen

Credits are rolling on the main screen

↖ ↗

The projector is functioning

As this diagram illustrates, there is no causal connection between your experience when you look at the viewing room monitor, and what is happening on the main screen.

Suppose that you have long been in the habit of looking at the viewing room monitor to check what is on the main screen. So, when you look at the monitor, you do not engage in any conscious inference; instead, it just immediately *seems* to you that whatever you see on the monitor is what is on the main screen. Then, when you look at the monitor, you might say, "I see that the credits are rolling on the screen in the main theater." If this assertion is really true—that is, if you really *can* see what is happening on the main screen by looking at the viewing room monitor—then this case provides yet another counterexample to the causal constraint.

4. Explanatory genealogy and perception

At this point, we have a puzzle. The cases I discussed in Section 3 suggest that there are counterexamples to the causal constraint, so it ought to be rejected. But the cases I discussed in Section 2 suggest that there are *also* counterexamples to the reliability constraint, so it ought to be rejected as well. Even so, we have seen that veridicality and immediacy of a given experience are not sufficient to make that experience a perception. An extra ingredient is required. What is the extra ingredient?

Let us say that a fact F is an *explanatory ancestor* of a fact G (and G is an *explanatory descendent* of F) if F is part of the causal explanation of G. And F is an *explanatory cousin* of G if there is some third fact H such that F is an explanatory descendent of H and G is an explanatory descendent of H. So there are at least two different kinds of explanatory relationships: the ancestor-descendent relationship, and the cousin-cousin relationship. I suggest that the important feature of genuine perception, in the cases that we

have considered thus far, is not *causal contact* (as the causal condition states) but rather *explanatory relatedness*. This, I believe, is the missing ingredient that we have been looking for.

The causal constraint implies that if a given seeming is an explanatory descendent of a given fact, then the seeming might qualify as a perception of the fact—but if a given seeming is an explanatory *cousin* of a given fact, then the seeming *cannot* qualify as a perception of the fact. That this constraint is too narrow is shown by the hawk case and the movie projector case. In those cases, a seeming is an explanatory cousin of a fact, yet the seeming is a perception of the fact. In light of this, I suggest that we broaden the causal constraint, and endorse the following explanatory relatedness constraint:

> (1'') For S to perceive that P, S must bear an appropriate explanatory relationship with the fact that P.

The explanatory relatedness constraint is silent about which explanatory relationships suffice to make perception possible. But presumably the appropriate relationships will include *some* forms of cousin-cousin relationships, and some forms of ancestor-descendent relationships. Appropriate relationships should include, for instance, the ancestor-descendent relationship found in the Fake Barn Country case from Section 2, as well as the cousin-cousin relationship found in the hawk case and the movie projector case from Section 3. And it seems to me that we *also* ought to consider as appropriate the cousin-cousin relationship found in the Fork Model that I described above, in Section 1. The primary difference between the Fork Model and the movie projector case is that the Fork Model involves a non-causal explanatory connection, whereas all the explanatory connections in the movie projector case are causal. But I cannot see why this should make an important difference. After all, facts that have non-causal explanations (e.g., patterns) are perceivable.[55] Given this, and given the structural similarity between the Fork Model and the cases discussed in Section 3, it seems that proponents of (1'') should grant that the Fork Model indicates a possible way to achieve non-naturalist moral perception.

[55] See T. Chappell 2008 for a clear discussion of the perceivability of patterns.

Conclusion

All of this leaves open the question of *why* we would be inclined to endorse implications of the explanatory relatedness constraint. We asked this question about the causal constraint; we should ask it about the explanatory relatedness constraint, as well. Provided that a given experience is veridical, why care whether that experience is explanatorily related to its object?

I think that the right answer to this question will refer to the distinctive role of perception as an *encounter* with the world. (This role is missing, I believe, in an account of perception as an especially *reliable* way to have veridical experiences.) To look at a green object is both to perceive it and to encounter it; I cannot encounter the green object in the same way by, say, reading about it. In a similar way, when I perceive *that* an object is green, I encounter the fact that it is green. And in such a case, the fact that the object is green is explanatorily related, in a certain very direct way, to my experience. Thus it seems that explanatory relatedness is crucially involved in the kind of encounter with the world (and the facts it contains) that characterizes perception.

But it turns out that the kind of encounter that characterizes perception can occur in virtue of a wider range of different kinds of explanatory relationships than proponents of the causal constraint expect. This, I believe, is the significance of the cases that I discussed in Section 3. Nevertheless, it does seem as if *some* kind of explanatory relation between an experience and its object is required for genuine perception; this is captured in the explanatory relatedness constraint. If I am right to think that non-naturalist moral perception is compatible with the explanatory relatedness constraint, then this would support the view that moral facts on a non-naturalist view can be encountered perceptually in much the way that facts about the color of objects can be encountered perceptually.

Of course, none of these points show that non-naturalist moral perception ever *actually occurs*. My argument here has been merely that non-naturalist moral perception is possible. In particular, my

position is that if the Fork Model were true, *then* non-naturalist moral perception would be actual. Whether we can ever verify the truth (or falsity) of the Fork Model is a different question.

Knots, Dilemmas, and Designers: A Critique of Metaethical Constructivism

Introduction

There are parallels between metaethical constructivism and the biological design hypothesis. According to the biological design hypothesis, all of life is a result of the choices or attributes of some intelligent being—a designer. Different design proponents posit different designers and different modes of design. It is possible, for instance, to construct a design hypothesis according to which everything—including all of life—is a lunatic's daydream. But most design proponents are creationists, who hold that God intentionally designed and created the world, along with all living things in it. Standard evolutionary theorists, by contrast, reject the design hypothesis, and hold that biological phenomena are a result of natural selection and other processes of biological evolution—blind processes that are independent of conscious choice or intelligent design.

Cases of apparently suboptimal biological design raise a problem for some design hypotheses, such as traditional varieties of creationism. Consider the panda's thumb—which is not really a thumb at all, but a bone spur with which pandas strip bamboo branches in preparation for eating.[56] Stripping branches with this crude implement is inefficient. The panda's thumb serves its purpose, but not very well. Evolutionary theory can handily explain this. Ex hypothesi, evolution does not reliably produce optimal designs; it reliably produces only incremental improvements to existing designs. So evolutionary theorists are not surprised that evolved organisms, such as the panda, exhibit suboptimal design. But God as traditionally conceived—omnipotent and loving—would presumably be able and willing to give his creatures a perfect, elegant, and efficient design. So traditional creationism seems unable to explain why the panda's thumb is as it is.

Metaethical constructivism is a *kind* of design hypothesis. Constructivists hold that moral facts are fixed by the normative psychology of (actual or hypothetical) intelligent beings. There is a metaethical

[56] See Sober 2008 pp. 126-128 and Gould 1980.

constructivist analog of creationism: theological voluntarism, which holds that moral facts are fixed by God's evaluative attitudes, such that an action is morally required just because God approves of that action. We can empirically test creationism by asking whether biological phenomena such as the panda's thumb are as they would be if creationism were true; likewise, we can test theological voluntarism by asking whether *moral* phenomena are as they would be if theological voluntarism were true. For instance, God as traditionally conceived holds a view of punishment that is captured by the biblical retributivist slogan "An eye for an eye." Thus theological voluntarism of the traditional sort runs into serious trouble if, e.g., a criminal justice system oriented around reparations turns out to be morally preferable to a retributivist system of punishment.

Although particular versions of the biological design hypothesis, such as traditional creationism, are empirically testable, it is unclear whether the biological design hypothesis itself is empirically testable. For it is unclear whether there is *any* conceivable set of biological observations that would be at odds with *every* version of the design hypothesis. We can imagine, for instance, a designer whose aims are strange enough that she intended for the panda's thumb to be precisely as it is.

But the fact that the biological design hypothesis seems to be empirically untestable does not show that our evidence cannot justify us in rejecting it. *Some* versions of the design hypothesis, such as traditional creationism, are empirically testable (and fare poorly). Meanwhile, versions of the design hypothesis that are fully consistent with available empirical evidence usually turn out to have various theoretical vices. For instance, the hypothesis that the world is a lunatic's daydream seems to be explanatorily impotent. No one has proposed a version of the design hypothesis that possesses a mixture of empirical *and* theoretical virtues that is comparable to evolutionary theory. So, given our current empirical evidence, we should reject the design hypothesis in favor of evolutionary theory.

I hope to show that we ought to think about metaethical constructivism in some of the same ways that we already think about the biological design hypothesis. As I have suggested, particular versions of metaethical constructivism such as theological voluntarism are testable in a certain sense: We can ask whether the moral truth is as it would be if theological voluntarism were true. By contrast, constructivism

itself, such as the design hypothesis in its most general form, does not seem to be testable in this way. No matter what the moral truth turns out to be, it might conceivably be fixed by the normative psychology of *some* real or hypothetical being. But this does not mean that our total evidence cannot possibly justify us in rejecting constructivism.

In this connection, I am particularly interested in what I call *moral knots*. The most familiar kind of a moral knot is a moral dilemma. A moral dilemma is a case in which an agent has an absolute moral obligation to do A, and has an absolute moral obligation to do B, but cannot do both A and B. For instance:

> Two-obligation case: Jen has an absolute moral obligation to give $200 to Steve, and has an absolute moral obligation to give $200 to Pete, but Jen has only $200.

But moral dilemmas are just one instance of a wider type of conceivable moral conflict. Consider this case:

> Three-obligation case: Jen has an absolute moral obligation to give $200 to Steve, an absolute moral obligation to give $200 to Pete, and an absolute moral obligation to give $200 to John, but Jen has only $400.

The common feature of these two cases is the defining feature of what I call a moral knot. A moral knot is a case in which an agent cannot satisfy all of her absolute moral obligations, and (thus) cannot avoid violating at least one of them.

There are other conceivable kinds of moral knots. Consider the "Ought implies can" principle (OIC), which states that if an agent has an absolute moral obligation to do A then the agent can do A. If this principle is false, then there is a case like the following one:

One-obligation case: Jen has an absolute moral obligation to give $200 to Steve, but Jen has no money.

In this case, Jen is unable to avoid violating her *sole* absolute moral obligation, so she is unable to avoid violating *at least one* of her absolute moral obligations. So this case would count as a moral knot. Of course, there is no guarantee that such a case is even possible. Perhaps there are *no* possible cases that would count as moral knots. Or perhaps some kinds of moral knots are possible while others are impossible. The possibility or impossibility of moral knots will depend on what the moral truth turns out to be.

My plan is as follows. In sections 1-2, I lay some groundwork by discussing the concept of an absolute moral obligation. In Section 3, I discuss an example of what I call a *situation of infinite choice*. In a situation of infinite choice, an infinite series of different actions is available to the agent. The agent can perform *any* of the actions in the series, but cannot perform *all* of them. Yet for each action in the series, the agent has a moral reason to do otherwise, and has no reason not to do otherwise. I argue that such situations are possible and that they count as moral knots. In Section 4, I consider one version of metaethical constructivism: the ideal observer theory, according to which the fact that an agent has an absolute moral obligation to do A is constituted by the fact that an ideal observer would advise the agent to do A. I argue that if the ideal observer theory is true, then situations of infinite choice do not count as moral knots. Thus, in sections 3-4, I assemble an argument against the ideal observer theory: Some situations of infinite choice count as moral knots; but if the ideal observer theory is true, then *no* situations count as moral knots; therefore, the ideal observer theory is false.

My argument in Section 4 causes problems only for the ideal observer theory, which is just one version of constructivism. There are many alternative constructivist theories. For instance, there is the ideal *designer* theory, which says that the fact that an agent has an absolute moral obligation to do A is constituted by the fact that A is required by a *moral code* (i.e. set of moral rules) that *is* or *is as if it were*

designed by an ideal designer. I argue that constructivists who are persuaded by my argument in Section 3 to reject the ideal observer theory ought to consider some version of an ideal designer theory.

My argument in sections 3-4 is, I hope, reasonably persuasive. My argument in Section 5 is more speculative. I discuss the much-discussed problem that arises from the fact that we experience different moral intuitions about certain trolley cases even though those cases seem to be morally similar to one another. I show that we can resolve this problem in an unusually appealing way if we are willing to accept dilemmatism, the view that moral dilemmas occur routinely in ordinary life. Then I address a few of the main objections to this dilemmatist solution to the trolley problem. My aim in this section is simply to show that dilemmatism *stands a reasonable chance* of being true; I do not hope to conclusively establish its truth.

In Section 6, I argue, via analogy with the biological design hypothesis, that *if* dilemmatism is true, then the ideal designer theory is probably false. By contrast, dilemmatism provides no particular reason to doubt metaethical realism, the view that the moral truth is independent of the evaluative attitudes of actual or hypothetical beings. Thus, dilemmatists ought to prefer realism over the ideal designer theory.

In Section 7, I argue (on the basis of my conclusions in sections 4 and 6) that if dilemmatism is true, then constructivists ought to choose from among various *non-ideal* constructivist theories—theories according to which the moral truth is fixed by the normative psychology of *actual* persons (or groups of people). These theories make it easy to explain the truth of dilemmatism; but they tend have a number of well-known (and quite serious) disadvantages, whereas realist theories lack those disadvantages. Thus I argue that dilemmatism would provide the basis of a case for realism. In other words, anti-realists ought to be anti-dilemmatists, and dilemmatists ought to be realists. Thus I reject a widely-held idea that moral dilemmas are somehow more amenable to anti-realism than realism.[57]

[57] See Railton 1996 pp. 160-163 for discussion of this idea, which has origins in Bernard Williams's influential discussion of the metaethical significance of dilemmas (1965).

1. The nature of absolute moral obligations

A moral knot is a certain kind of conflict between absolute moral obligations. So before I can discuss moral knots, I have to explain what I take an absolute moral obligation to be. That is what I will do in the present section.

I assume that when we engage in moral reasoning about particular situations of moral choice, we mainly aim to understand the moral dimensions of *act-types*. I think that this assumption ought to be relatively uncontroversial. Suppose, for instance, that you have just gotten your paycheck and you have to decide what percentage of it to give to charity. No one would raise an eyebrow if you announce that you are morally required to give at least 10% of your income to charity; and giving at least 10% is an act-type. Further, even if you judge that you are morally required to give *exactly* 10% of your income to charity, this would still be an act-type. There are, after all, many different ways to go about giving exactly 10%.

Of course, one cannot perform an instance of an act-type without performing an act-token. We satisfy (or violate) our moral requirements *by* performing act-tokens. Nevertheless, our moral requirements generally seem to be requirements to perform act-types. Thus, although the word "action" is ambiguous—it can refer to an act-type or an act-token—it generally seems to refer to act-types when used in contexts of moral deliberation. For instance, when we make assertions of the form "A is morally required," A usually turns out to be an act-type.

Further, I assume that a central purpose of moral reasoning about particular situations is to uncover the *moral properties* of actions (act-types)—and that we ascribe such properties when we engage in ordinary moral talk. For instance, if I say that it would be morally right to give at least 10% of your income to charity, then I ascribe a moral property—namely, moral rightness—to the act of giving at least 10% of your income to charity.

It seems that any given action might have more than one moral property. For instance, it is logically possible that a single action is both unjust and courageous, or both morally required and morally blameworthy, or both all-things-considered morally required and prima facie morally wrong, etc.

Different people will give different lists of moral properties that actions can have, and different people will hold different views about how the moral properties of actions relate to one another. Some of those relations will be more obvious than others. For instance, it is clear that the same action cannot be both courageous and cowardly, just as the same person cannot be both a bachelor and a spouse. But there are compelling arguments on both sides of the question whether the same action can be both morally required and morally blameworthy.

There is no agreed-upon list of moral properties of actions, but it ought to be widely agreed that some moral properties are *more morally important* than others. Perhaps the clearest expression of this idea is found in Ross's distinction between all-things-considered and prima facie obligations. For Ross, the fact that a given action is all-things-considered right has greater moral importance than the fact that the same action is prima facie wrong. But I suggest that even those who reject a Rossian framework will typically endorse *some* sort of a ranking of moral properties. [58]

Property monists might be seen as a counterexample to this suggestion. Property monists believe that any given action can have at most one moral property. For instance, a property monist might say that "morally required" and "morally forbidden" denote the only moral properties that actions can have, and that any given action can have at most one of these properties. But strictly speaking, property monists do endorse rankings of moral properties—although all of their rankings are single-entry. To find a *genuine* counterexample to my suggestion, we would need a view according to which (a) a given action can have more than one moral property, and (b) the various moral properties of a given single action cannot be ranked in order of moral importance. I know of no one who holds such a view.

So I assume that whatever the correct list of moral properties of a given action turns out to be, some moral properties on that list will be more important than others. This naturally raises the question of how we ought to go about determining the correct ranking of moral properties.

---

[58] Note that this suggestion does *not* presuppose consequentialism. All versions of genuine consequentialism, I believe, hold that *actions* can be ranked in order of moral status, and that these rankings depend in some way on the evaluation of actions' outcomes (where "outcome" may be broadly construed). But my assertion is only that *the moral properties of any single action* can be ranked in order of moral importance. This view is silent about how and whether we ought to rank actions in order of moral status.

To see one way of partially answering that question, let us first consider one (of many) respects in which moral reasoning differs from scientific investigation. Suppose that I want to investigate the physical properties of a particular star. There are many different interesting facts about the star that I might uncover—its mass, its age, its radius, etc. And presumably, some of these facts will be more important than others. But (apart from my own scientific goals and interests) there is no particular property of the star that I especially ought to aim to understand. Physical science leaves vast freedom to physical scientists.

By contrast, moral reasoning—especially moral reasoning about a particular situation of moral choice—seems to be much more constrained. The usual view, I believe, is that moral reasoning about a particular situation aims to uncover a single moral verdict about *what to do* in that situation. Call this the *univocal verdict model* of moral reasoning. According to the univocal verdict model, if you are engaging in moral reasoning without attempting to reach a verdict about what to do, then you are doing it wrong. Moreover, as soon as you successfully reach a verdict about what to do, then you are done with moral reasoning (at least until you encounter another situation in which moral reasoning is called for). In short, a verdict about what to do is the *ultimate aim* of moral reasoning, and everything else that occurs in the course of moral reasoning is but a means toward that end.

Some of what I say in later sections will cast some doubt on the univocal verdict model. But I grant that this model contains a significant kernel of truth. I accept that for any given action A, there is *at least one* moral property of A that moral reasoning ultimately aims to uncover. Call this the *top-ranked* moral property (or properties) of A. (It is conceivable that more than one property could be tied for first place.) Thus, I grant that aside from top-ranked moral properties, any other moral properties hold interest (for those engaged in moral reasoning) only insofar as they shed light on top-ranked moral properties. The proponent of the univocal verdict model will go further than this—she will say that once an agent has determined the top-ranked moral properties of all the actions open to her, then the agent will be able to derive a univocal verdict about what to do. This I leave as an open question for now. But I grant the importance of the distinction between moral properties that moral reasoning properly aims to uncover

(top-ranked moral properties), and moral properties that moral reasoning uncovers only in order to uncover other moral properties.

This provides a partial answer to my earlier question of how to determine the correct ranking of moral properties. It tells us that we can determine the *top-ranked* moral property (or properties) of any given action by determining which property (or properties) of that action is (are) an appropriate ultimate aim of moral reasoning.

To illustrate this view, consider a situation in which a soldier has been ordered to kill an innocent person. If the soldier disobeys, then he will be mildly reprimanded but will face no other bad consequences. In a typical Rossian view, the top-ranked moral property of killing the innocent is its *overall moral wrongness* and the top-ranked moral property of refusing to do so is its *overall moral rightness*. Each of these actions might have other moral properties (such as prima facie rightness or wrongness) but those properties are of lower moral importance, in the Rossian view. My claim here is that, regardless of whether the Rossian framework correctly identifies the moral properties of actions, the Rossian is *not* wrong in her implicit assumption that actions *have* top-ranked moral properties.

Let us say that a given moral property F of a given action A has *positive valence* if F favors doing A. Examples of moral properties that seem to have positive valence include: being *praiseworthy, admirable, just, courageous, required, prima facie right, all-things-considered right*—etc. A moral property that lacks positive valence will have either negative or neutral valence. So, for instance, in the case of the soldier, it is uncontroversial that the top-ranked moral property of killing the innocent—*whatever* that property is—has negative valence, and the top-ranked moral property of refusing to do so has positive valence.

This distinction raises a question: Are there any possible situations—actual or merely possible—in which an agent is unable to perform all available actions whose top-ranked moral properties have positive valence? This strikes me as a worthwhile question; it is the main question that I want to examine in this paper. So I will define "absolute moral obligation" in such a way as to make it easy to address that question. I define "absolute moral obligation" as follows:

> An agent has an absolute moral obligation to do A iff$_{(df)}$ the top-ranked moral property of A has positive valence.

This definition, in turn, tells us what a moral knot is. Recall that I defined a moral knot as a situation in which an agent is unable to satisfy all of her absolute moral obligations. So, given the above definition of "absolute moral obligation," a moral knot is a situation in which an agent is unable to perform all of the available actions whose top-ranked moral properties have positive valence.

It seems to me that absolute moral obligations, as I have defined them, are the main subject-matter of ordinary moral reasoning and moral disagreement. For instance, if one person says that a given wealthy individual is morally required to give at least 40% of her income to charity, and another person denies that view, then (I claim) we can gloss this as a disagreement about the content of the wealthy individual's absolute moral obligations. The idea that absolute moral obligations are under discussion in ordinary moral talk is plausible, I think, because I have given a very broad definition of "absolute moral obligation." I take this to be an advantage of my preferred definition.

2. Reasons and absolute moral obligations

Even if my definition of "absolute moral obligation" is *broad enough*, as I contend, it might still be *too* broad. In particular, perhaps we ought to include in our definition some details about how absolute moral obligations are related to *reasons*.

I have borrowed the phrase "absolute moral obligation" from Ross, but I do not accept all of his views about the nature of absolute moral obligations. Ross seems to have endorsed a definition of "absolute moral obligation" that straightforwardly entails the following view:

> Overall Reasons: An agent has an absolute moral obligation to do A iff the balance of the agent's moral reasons favors doing A.[59]

A *moral reason* for my purpose here is a particular kind of *normative* reason. A normative reason is a fact or consideration that favors a given action. A normative reason counts as *moral* if it is the sort of consideration that can be appropriately taken into consideration in moral reasoning.

Whether Overall Reasons is true or false, I think we ought to carefully avoid any definition of Overall Reasons that implies it. We want people to be able to reject Overall Reasons without contradicting themselves.

Imagine a *close case* in which an agent has very strong moral reasons to do A, and has moral reasons *almost* as strong to do B—but the agent has a strong non-moral reason to do B, and has no non-moral reason to do A. A proponent of supererogation might hold that in such a case, A is neither required nor forbidden, but supererogatory.[60] However, if we accept Overall Reasons as a definitional truth, then *by definition* the agent has an absolute moral obligation to do A in the close case. This forces the proponent of supererogation into an awkward position in which there is an absolute moral obligation to do A even though A is not morally required. I think it would be better simply to allow the logical possibility that the agent has no absolute moral obligation to do A in close cases.

So I think we should leave the truth of Overall Reasons as an open question, at least at the level of definition. Nevertheless, the *spirit* of Overall Reasons is very attractive. Overall Reasons posits a tight link between absolute moral obligations and reasons, and we want such a link to exist. When we

---

[59] Brink 1994 (pp. 215-217) apparently takes Ross to endorse a version of Overall Reasons. However, it is not clear to me whether Ross endorses Overall Reasons as a *definitional* truth, as against a *substantive* claim. An alternative reading is that for Ross, absolute moral obligations—synonymous with overall moral obligations—are just *toti-resultant* by definition. For Ross, a toti-resultant property is one that "belongs to [that action] in virtue of its whole nature and of nothing less than this" (Ross 1988 p. 28). It is *logically* possible for our obligations to be toti-resultant without being due to the balance of our moral reasons. In any case, whether Ross holds that Overall Reasons is a definitional truth or a substantive one, I accept that he endorses a version of it.

[60] This is consistent with a view of supererogation defended in Portmore 2003. Of course, it is not the *only* account of supererogation. It is conceivable, for instance, that actions are supererogatory only in certain cases in which the agent's reasons to do A are *equally strong* as the agent's reasons to do B. So, while the Portmore-style account of supererogation is among the most plausible accounts, it is certainly not the only available one.

encounter people who intend not to fulfill their absolute moral obligations, we want (and expect to be able) to offer them reasons to change their plans. Overall Reasons ensures that we will always be able to do this.

But Overall Reasons is not *required* in order to have a tight link between reasons and absolute moral obligations. Consider the following view:

> Minimal Reasons: An agent has an absolute moral obligation to do A *only if* the agent has at least one reason to do A.[61]

Minimal Reasons is considerably weaker than Overall Reasons, but it succeeds in guaranteeing that agents will always have a reason to fulfill their absolute moral obligations.

Even so, we should again want to avoid any *definition* of "absolute moral obligation" that entails Minimal Reasons. (Of course, this is compatible with the truth and the appeal of Minimal Reasons.) Imagine a voting advocate who holds that each citizen is morally required to vote in important elections. To justify her view, she points out that if the number of voters in a democracy is too low, then disaster surely follows. However, she is ready to grant that nothing of importance depends on whether any given citizen votes or fails to vote. In that case, she might believe (or be prepared to accept) that any given citizen has no reason at all to vote. She might say: "I can offer any given citizen a reason to *accept the moral rule* that requires voting. The reason to accept that rule is the fact that disaster follows if too few people show up to vote. But I cannot offer any given citizen a reason *to vote*. I allow that there is no such reason, because I allow that nothing of importance depends on whether a given citizen, taken as an individual, votes or not." There may be serious problems for this voting advocate's view, but I cannot see incoherence in it. But there *would* be incoherence in it if we accept a definition of "absolute moral obligation" that entails Minimal Reasons. So we should not accept such a definition.

[61] This is a version of moral rationalism as defended in Shafer-Landau 2003, ch. 7-8.

I suspect that there is *no* link between reasons and absolute moral obligations that we should accept as definitional. If this is right, any such link ought to be taken as a substantive claim. Consider the following view:

> Unilateral Reasons: If an agent has a very strong moral reason to do A, and has no reason (moral or non-moral) to do otherwise, then the agent has an absolute moral obligation to do A.

I believe that Unilateral Reasons is even more attractive than Minimal Reasons. Nevertheless, I think that we ought to avoid a definition of "absolute moral obligation" that entails even it.

Imagine a religious believer who holds that moral requirements are given by a set of religious rules and prohibitions. Then the believer faces a quandary if she ever has a very strong moral reason to violate one of her faith's rules and no reason to do otherwise. Of course, the believer might claim that such a case is not possible. She might say that the mere fact that an action violates one of her faith's rules provides her with a reason to do otherwise. Or she might say that all of her moral reasons are reasons to act in accordance with her faith's rules, so it is not possible for her to have a moral reason to violate any of those rules. But she might instead take a different view. She might say that she has an absolute moral obligation to follow her faith's rules, and this obligation is morally binding *despite* being completely independent of any and all of her reasons (whether moral or non-moral). So she might happily allow the possibility of a counterexample to Unilateral Reasons: a case in which she has an absolute moral obligation to adhere to her faith's rules despite having a very strong moral reason to violate those rules and no reason to adhere to them.[62]

Such a view might be absurd. I am inclined to think so. But I do not see why we ought to define "absolute moral obligation" in such a way as to make this view incoherent *as well as* absurd. So in my preferred view, Unilateral Reasons is a substantive principle.

---

[62] Kierkegaard, in his famous discussion of the biblical story of Abraham and Isaac (2006), sometimes seems to hold the view that Abraham has an absolute moral obligation to sacrifice Isaac *even though* Abraham has no reason to do so and has very strong moral reasons to do otherwise. If so, then Kierkegaard rejects Unilateral Reasons.

I have reviewed a few possible relations between reasons and absolute moral obligations, and have argued that each ought to be taken as substantive rather than definitional. I conclude that our best option is to accept the very spare definition that I offered in the previous section: an absolute moral obligation to do A *just is* the fact that the most important ("top-ranked") moral property of A has positive valence.

Nevertheless, certain relations between reasons and absolute moral obligations are very plausible. In particular, I take it that Unilateral Reasons is an *extremely* plausible principle. It seems to me that if we think that there is *any* configuration of reasons sufficient to establish the existence of an absolute moral obligation, then we ought to believe that Unilateral Reasons is true. But despite its high degree of plausibility, Unilateral Reasons has interesting implications. In the next section, I will employ Unilateral Reasons in an argument for the view that moral knots are possible.

3. Jove's knot

Consider the following scenario:

> There exists an omnipotent being named Jove, and there exists nothing else. Jove has the power to create any world he likes. In fact, he cannot avoid creating *some* world, for even if he refrains from actively creating a world, he would thereby *passively* create a world containing only Jove. Jove wants to be the most morally responsible world-creator that he can be. Unfortunately, as Jove holds before his mind the host of worlds, he sees that for each there is a better one. Although he can create any of them, he can't create the best of them—because there is no best.[63]

[63] This case, slightly modified, comes from Howard-Snyder and Howard-Snyder 1994. This case is at the heart of a debate in the philosophy of religion over the "problem of no best world," a challenge to traditional theism. See Kraay 2010 for a useful overview of the problem of no best world, which I will not discuss in this paper.

In the present section, I aim to show that Jove's case counts as a moral knot. Then, in Section 4, I will argue that this result provides the basis for an argument against one prominent version of constructivism: the ideal observer theory.

Before I begin, I ought to briefly address the issue of whether Jove's case is even metaphysically possible. Jove's case presupposes *No Best World*—the view that for any possible world there is a better one. We can easily imagine theories of value that would make No Best World false. For instance, suppose that justice is the sole value that a world can possess, and suppose that there is a perfectly just possible world. Then it follows immediately that there is a best possible world. However, it seems difficult to defend the view that justice is the sole value that a world can possess; more broadly, it is hard to imagine that there is any world that cannot be *somehow* improved. For instance, it seems possible to increase the number of happy people in any given world, no matter how many happy people it already contains—and it is hard to see how such increases would not count as improvements. So I think we can fairly assume that No Best World is true, and thus Jove's scenario is metaphysically possible, even if such a scenario is unlikely to ever actually occur.

My argument for the view that Jove's case counts as a moral knot is as follows:

(1) Unilateral Reasons: If an agent has a very strong moral reason to do A, and has no reason (moral or non-moral) to do otherwise, then the agent has an absolute moral obligation to do A.

(2) For any world W, Jove has a very strong moral reason to create a world better than W.

(3) For any world W, Jove has no reason (moral or non-moral) against creating a world better than W.

(4) Therefore, for any world W, Jove has an absolute moral obligation to create a world better than W.

(4) implies that Jove has infinitely many absolute moral obligations. He can satisfy as many of those obligations as he likes, but he cannot satisfy all of them; he is doomed to violate at least one of them. (In fact, he is doomed to violate infinitely many of them.) Thus, if (4) is true, then Jove's situation counts as a moral knot.

I have little to offer in defense of (1), Unilateral Reasons, which I regard as extremely appealing. As I suggested in the previous section, if there is *any* configuration of reasons sufficient to establish the existence of an absolute moral obligation to do A, it would presumably be the existence of a very strong moral reason to do A in conjunction with a total absence of reasons to do otherwise. I believe this observation is sufficient to establish Unilateral Reasons as plausible.

My case for (2) is a bit more complicated. To begin with, note that (2) does not follow immediately from No Best World. Suppose that the hierarchy of possible worlds is asymptotic: There is a level of goodness X such that no possible world is better than X, but there is a possible world that contains any arbitrary level of goodness less than X. Then No Best World would be true, but (2) would probably be false. To see this, consider a world W_a that is very close to X. Jove would be able to create a better world than W_a, but he would not be able to creat a *much* better world than W_a. So it would be hard to motivate the view that Jove has a *very strong* moral reason to create a world better than W_a.

But the hierarchy of possible worlds is not plausibly asymptotic. Imagine a world W_1 that contains one trillion happy people. Jove could choose to create this world, or he could choose instead to create W_{10}, which contains ten trillion happy people. It seems fair to say that Jove has a quite strong moral reason to choose W_{10} rather than W_1. (After all, nine trillion happy lives are at stake in his decision.) By the same token, Jove has a very strong moral reason to choose W_{100} rather than W_{10}, and to choose W_{1000} rather than W_{100}—etc. For any world containing arbitrarily many happy people, there is an alternative

world containing vastly more happy people; and it seems that Jove will always have a very strong moral reason to choose the latter world over the former one. If this is right, then (2) is true.[64]

Here it might be objected that (2) presupposes some form of consequentialism. Non-consequentialists will typically insist that *not all* of our moral reasons are reasons to choose a better outcome over a worse one. However, very few non-consequentialists maintain that we *never* have a reason to choose a better outcome over a worse one. And in fact, all sorts of theorists—non-consequentialists as well as consequentialists—are traditionally drawn to the view that an agent like Jove, who can choose to create any world he likes, has a very strong moral reason to choose a much better world rather than a much worse one.[65] This view lies at the basis of the problem of evil for theism—a problem whose difficulty does not depend on consequentialist assumptions.

Let us now turn to (3)—the view that for any world W, Jove has no reason (moral or non-moral) against creating a world better than W. It is important that this claim not be confused with the view that for any given pair of worlds Jove has no reason against creating the better of the pair. This latter claim is, I think, obviously wrong. For instance, we can imagine a world W_b in which there is no pain, and therefore no opportunity for courage, and a world W_w in which there is pain and ample opportunity for courage. Plausibly, W_w is considerably worse than W_b; nevertheless, there is apparently a reason against creating W_b—that reason being that W_b includes no opportunities for courage. So, if (3) carried the

[64] There will be some who say that there is no moral reason to bring a happy person into existence. Benatar (2008) defends such a view. In fact, I think that Benatar will deny No Best World: I think that Benatar's extreme anti-natalism implies that the best possible world is one in which there is no conscious life. But Benatar's view is exotic; I am not bothered by the fact that those who share his view will be unconvinced by my argument here. True, some who disagree with Benatar will still hold that there is an optimal non-zero number of people. However, it seems to me that these people still ought to accept (2). They might be persuaded to agree, for instance, that (a) Jove has a strong moral reason to improve the lives of the finite number of people who exist in a given world, and (b) it is always possible to make large improvements in people's lives, no matter how good their lives already are. This illustrates that there are many ways to avoid the conclusion that the hierarchy of worlds is asymptotic.

[65] However, it *is* significant that this line of argument does show that consequentialists *among others* ought to believe that Jove faces a moral dilemma. Ruth Marcus (1996, p. 225) speaks for many, I believe, when she suggests that consequentialism is a "plausible candidate for a moral system free of dilemma," and that "[i]it is clear that for dilemmas to arise some deontological principle is required." See also Mintoff 1997 for an argument for the view that for consequentialists, situations of infinite choice (like Jove's case) do not count as genuine dilemmas. If I am right, then both Marcus and Mintoff are wrong.

implication that there cannot be a reason against creating world W_b, then (3) would deserve to be controversial.

But (3) is a more sensible view. It says that Jove has no reason against choosing *some* world that is better than W_w. This claim would be false if (for instance) W_w were *the very best* possible world containing opportunities for courage (such that every better world lacks opportunities for courage). In that case, Jove would have a clear reason against creating a world better than W_w, since doing so would amount to creating a world without opportunities for courage. But it would be odd if there is a best possible world containing opportunities for courage. In general, for any given world that contains a given range of goods, we seem able to imagine an even better world that contains not only the original range of goods but also contains additional goods. If this is right, then (3) is highly plausible.

But there is at least one interesting objection to (3). It runs as follows. Take a given world—say, W_{1000}. According to (2), Jove has a very strong reason to create a world that is better than W_{1000}. This seems to provide Jove with a reason against creating W_{1000}. But there is nothing special about W_{1000}; we ought to say that that Jove has similar a reason against creating W_{1001}, and a reason against creating W_{1002}—etc. Thus, (i) for any given world, Jove has a reason against creating that world. It seems to follow from this that (ii) Jove has a reason against creating *any world at all*. And from (ii) it seems to follow that (iii) Jove has a reason against creating a world that is better than W_{1000}. In that case, (3) is plainly false: It is not the case that for any world W, Jove has no reason (moral or non-moral) against creating a world better than W.

However, I object to the move from (i) to (ii). Recall that Jove has to create some world—there is no way for him to avoid doing so. So (ii) implies that Jove has a reason to do something that he is unable to do. This is like the view that I can have a reason to bring about world peace or to build a triangle with four sides. I object to (ii) on the grounds that an agent cannot have a reason to do A unless the agent can do A.[66] But note that (i) does not attribute to Jove a reason to do anything that he cannot do: Jove is

[66] This assertion depends on a highly plausible "reasons imply can" principle. For a useful discussion of that view, see Streumer 2010.

perfectly capable of refraining from creating W_{1000}, and perfectly capable of refraining from creating W_{1001}, etc. So I can (and do) accept (i) while rejecting (ii).[67]

This completes my defense of (1), (2), and (3). I conclude that Jove's situation counts as a moral knot. But it is worth emphasizing that Jove's situation is just an instance of a *general type* of situation. In situations of this type, there is an infinite series of different actions are available to the agent. The agent cannot perform all of the actions in the infinite series. Yet for each action, the agent has a moral reason to do otherwise, and has no reason *not* to do otherwise. This is what I call a *situation of infinite choice*. Even if I have failed to show that Jove's situation counts as a moral knot, there might easily be other metaphysically possible situations of infinite choice that do count as moral knots.[68]

4. Moral knots and the ideal observer theory

I believe that if Jove's case counts as a moral knot, then this should worry certain constructivists. In particular, it ought to worry constructivists who accept the following view:

> Ideal observer theory: The fact that an agent has an absolute moral obligation to do A is constituted by the fact that an ideal observer would advise the agent to do A.[69]

[67] This means that I have to reject the following principle: If an agent has a reason against doing A, and has a reason against doing B, then the agent has a reason against doing *either* A or B. I believe that I can justifiably reject this principle. Suppose that if I go through Door #1, then I will be mauled by tigers, and if do anything other than go through Door #1, then I will be splashed with uncomfortably cold water. In this case, it is not at all obvious that I have a reason against *either* going through Door #1 or not going through Door #1. What precisely would that reason be? In any case, such a reason would violate the "reasons imply can" principle, since I *cannot avoid* either going through Door #1 or not going through Door #1.

[68] For instance, Slote (1989, p. 114) presents another situation of infinite choice, in which an immortal wine connoisseur has to select a finite number of bottles for himself, and must make do forever with only that many bottles. Slote considers this case to be a candidate for a certain kind of *rational* paradox; but perhaps with some modification it could be made into a candidate for a moral knot.

[69] For Firth (1952), the ideal observer is omniscient (with regard to non-moral facts), dispassionate, fully rational, and so on. But this is only the most famous view of what an ideal observer would be like. Some will argue contra Firth that an ideal observer ought to have *heightened* emotional capacities, rather than be dispassionate. Others will argue that an ideal observer should not be omniscient. For instance, a Rawlsian might hold that an ideal observer would be in the original position: ignorant of her own place in the world. So there are many different

Here I aim to show that the ideal observer theory implies that Jove does *not* face a moral knot. Thus, if Jove does face a moral knot, as I argued in the previous section, then the ideal observer theory is false.

Whatever the truth about Jove's moral situation, Jove would create a world; it is built in to the description of Jove's case that Jove has to do so. And of course this is true not only of Jove; *anyone* in Jove's situation would have to create a world. Since a person in Jove's situation would have to create a world, she would have to choose a world to create. The way in which Jove, or anyone else, would make this choice is unclear. Just as there is no best world for Jove to create, there is no best decision procedure that he can use to select a world for creation. (For instance, he can select a world at random; or he can choose a cut-off point, and set aside all the worlds below that cut-off point, and select a random world from the remaining worlds; or he can choose an *even higher* cut-off point – etc.[70]) For these reasons, there is probably no single action that Jove (or anyone) *would* do.

By the same token, there is probably nothing in particular that an observer—whether ideal or not—would *advise* Jove to do. I suppose that if any given person were asked repeated to give advice to an agent in Jove's situation (without being able to remember her advice on previous occasions), she would select different worlds for creation on each occasion.

If one accepts the ideal observer theory, then one must have something to say about cases in which there is nothing in particular that an ideal observer would recommend. For instance, suppose that I have a single piece of pie and I can give it to either Martha or Jim, and suppose that both Martha and Jim like pie equally. (For some reason I cannot split the piece between them.) To determine what an ideal observer *would* advise in this situation, we ought to imagine the nearest possible world in which an ideal

---

ways to be an ideal observer theorist. Consequently, a very wide range of constructivists count as ideal observer theorists—including some who do not explicitly identify as such. Consider, for instance, the kind of constructivism recently defended by Sharon Street (2008). Arguably, Street's constructivism amounts to an ideal observer theory according to which the truth of any given moral judgment J, relative to a given agent S, is constituted by the fact that an ideal observer would judge that J withstands scrutiny from the standpoint of all of S's other moral judgments.

[70] The fact that there is no best decision procedure available to Jove is well-known in the literature on the problem of no best world; it is pointed out already by the Howard-Snyders (1994), and is generally conceded by their respondents.

observer is asked to provide such advice. But perhaps the nearest possible world in which an ideal observer advises me to give the pie to Jim is no closer than the nearest possible world in which an ideal observer advises me to give the pie to Martha. In that case it seems that there is nothing in particular that an ideal observer would recommend. So, the ideal observer theorist ought to say that there is nothing in particular that I am absolutely morally obligated to do—either choice is permissible.

If this is what the ideal observer theorist should say about the Martha-and-Jim case, then presumably she ought to say the same thing about Jove's predicament. As I have suggested, there is probably no single action that *anyone*, even an ideal observer, would advise Jove to do. If that is so, then anyone who is fully committed to the Ideal observer theory ought to say that Jove does not face a moral knot. For the ideal observer theorist, the view that Jove faces a moral knot assigns *far too many* moral obligations to Jove—as a matter of fact, it assigns *infinitely* many obligations to Jove, whereas according to the Ideal observer theory Jove has no obligations at all. So anyone who believes that Jove faces a moral knot ought to deny the ideal observer theory—and anyone who is deeply wedded to the ideal observer theory ought to deny that Jove faces a moral knot.

This result has significance for some non-constructivists who reject the ideal observer theory. In particular, there are non-constructivists who believe that an ideal observer's advice is not *constitutive* of the moral truth—yet an ideal observer's advice provides a *perfectly reliable guide* to the moral truth. If I am right that Jove faces a moral knot, then these non-constructivists ought to change their view. But given that the ideal observer is inessential to their metaphysics, this might not be a particularly dramatic shift for them.

By contrast, constructivists who accept the ideal observer theory are more likely to regard it as a central feature of their view, since they hold that moral obligations are (presumably non-causally) *brought about* by the choices ideal agents would recommend. So the abandonment of the ideal observer theory would require them to undertake a drastic theoretical overhaul. This ought to motivate them to find fault with the premises of the argument that I offered in Section 2. For reasons that I offered above, I think that

the premises of that argument are highly plausible—but then again, ideal observer theorists are likely to regard their own view as highly plausible. So they may be in a difficult spot.

However, in the end, I suspect that ideal observer theorists ought to resolve this difficulty by rejecting the ideal observer theory and shifting to an alternative constructivist view. There are many versions of constructivism to choose from. Consider, for instance, the following view:

> Ideal designer theory: The fact that an agent has an absolute moral obligation to do A is constituted by the fact that A is required by a moral code (i.e. set moral rules) that is or is as if it were designed by an ideal designer.

The difference between the ideal observer theory and the ideal designer theory lies in the *roles* played by an ideal agent. The former theory involves an ideal agent who offers moral advice in particular situations. The latter theory involves an ideal agent who lays down a moral code that is applicable in multiple different situations.

As far as I can see, the ideal designer theory is fully compatible with the idea that Jove's situation counts as a moral knot. Imagine an ideal designer who is considering the option of including in her moral code a rule like Unilateral Reasons—a rule that requires an agent to do A in any situation in which the agent has a very strong moral reason to do A and has no reason to do otherwise. On the one hand, the ideal designer will notice that in the vast majority of real-world situations in which Unilateral Reasons has an application, its guidance is usable and (from the ideal designer's perspective) desirable. This is an advantage of Unilateral Reasons. On the other hand, the ideal designer will notice that when Unilateral Reasons is applied in highly unusual cases—namely, situations of infinite choice like Jove's case—then it produces a moral knot. This, perhaps, is a disadvantage of Unilateral Reasons. But the advantage seems to outweigh the disadvantage. So it is easy to imagine that an ideal designer could include Unilateral Reasons in her moral code *despite* the fact that Unilateral Reasons leads to moral knots in certain highly

unusual circumstances. Thus, even if the ideal observer theory runs aground on Jove's knot, the ideal *designer* theory seems able to handle Jove's knot. This is a major advantage of the latter theory over the former one, I think.

But even ideal designer theorists have some knot-related problems. Below, in Section 5, I will argue that if dilemmatism is true, then the ideal designer theory is probably false. But first I need to take a quick detour. In the next section (Section 4), I will present a *tentative* line of argument in favor of dilemmatism.

5. The possibility of moral dilemmas

Recall that a moral dilemma is a case in which an agent has an absolute moral obligation to do A, and has an absolute moral obligation to do B, but cannot do both A and B. To see why anyone might think that such cases are possible, let us start by considering an argument in favor of moral dilemmas that appeals to what I call *tough cases*. This line of argument is popular among proponents of dilemmas, but I think it is ultimately unpersuasive.

The distinguishing feature of a tough case is that we *feel torn* between more than one incompatible course of action when we consider such cases, and this feeling seems like evidence of some sort of significant moral conflict. Sophie's choice (in which a mother must select one of her two children to be murdered, or else both will be murdered) is an example of a tough case. Another is the case of Sartre's student, who has to decide whether to stay home and care for his sick mother (M) or leave home to join a resistance movement (R).

Let us take the case of Sartre's student (although any other tough case would serve equally well). Variations of the following argument[71] appear in numerous places throughout the literature on moral dilemmas:

[71] See McConnell 1996, pp. 37-39.

(5) If the student in Sartre's case does M, then he would appropriately feel guilty for failing to do R—yet if the student does R, then he would appropriately feel guilty for failing to do M.

(6) An agent can appropriately feel guilt only if the agent has violated an absolute moral obligation.

(7) Thus, the student has an absolute moral obligation to do M, and an absolute moral obligation to do R—so the student faces a moral dilemma.

Neither of the premises of this argument is obvious. Those who wish to deny (6) can argue that guilt can be appropriate in cases where something of great value is lost, even if it is lost non-culpably—or (similarly) in cases where a serious pro tanto moral rule is violated, even if it is violated permissibly. Moreover, even if (6) is granted, I am unaware of any convincing reason to believe (5). Some proponents of dilemmas have mentioned that it would seem inappropriate to *criticize* the student for feelings of guilt. But this does not show that such feelings would be appropriate. It might often be inappropriate to criticize others' inappropriate feelings. For instance, it could be inappropriate to criticize a child's irrational (hence inappropriate) fear of a harmless snake.

There is much more to say about tough cases than I can say here. But I maintain that in the end, tough cases simply do not provide clear evidence, one way or the other, as to whether moral dilemmas are possible (or actual). So I will instead consider a different line of argument—a line of argument that is at least intriguing and may also be promising, although it has almost no precedent in the literature.[72]

To start, consider a pair of trolley cases. In Switch_{20}, 20 people will be struck and killed by a runaway trolley—unless a certain agent throws a certain switch, in which case the trolley will be diverted to a side-track, where it will kill a single person. In Push_{20}, 20 people be struck and killed by a runaway

[72] But see Paske 1990 for a version of the position that I will develop here.

trolley—unless a certain agent pushes a certain football player from a bridge, in which case the football player will be struck and killed by the trolley, and the trolley will stop before it reaches the 20 people.

Here is an inconsistent tetrad:

I_1: The agent in $Switch_{20}$ faces an absolute moral obligation to throw the switch.

I_2: The agent in $Push_{20}$ faces an absolute moral obligation *not* to push the football player.

Difference Requirement: I_1 and I_2 are both true only if there is a morally relevant difference[73] between $Switch_{20}$ and $Push_{20}$.

No Difference: There is no morally relevant difference between $Switch_{20}$ and $Push_{20}$.

At least one of these propositions must be false. I am going to argue that if we reject the Difference Requirement, then we ought to believe that $Switch_{20}$ and $Push_{20}$ count as moral dilemmas.

But first I have to explain why someone might *want* to reject the Difference Requirement. Imagine yourself in the following predicament. You believe that I_1 and I_2 both possess a high degree of intuitive plausibility—and you trust your intuitions, so you are very reluctant to reject either I_1 or I_2. Nevertheless, try as you might, you cannot find a morally relevant difference between $Switch_{20}$ and $Push_{20}$. This is not due to simple thoughtlessness on your part. You have considered all of the main views regarding such a difference, and you found them all wanting. For instance, perhaps you used to be a proponent of the Doctrine of Double Effect. So you used to believe that it is morally significant that the agent in $Push_{20}$ kills the football player as a *means* to save the 20, whereas the agent in $Switch_{20}$ kills the football player as a mere side-effect of an action aimed strictly at saving the 20. However, you eventually

[73] A "morally relevant difference," for my purpose here, is the sort of difference that could explain why the absolute moral obligations of an agent in one case differ from the absolute moral obligations of an agent in another case.

came to believe that there is no principled moral distinction between killing as a means and killing as a mere side-effect. And so it went with all other potentially relevant differences between $Switch_{20}$ and $Push_{20}$. In the end, you decided that it probably only *seems as if* there is a morally relevant difference between these cases. Thus you reluctantly came to accept No Difference. Nevertheless, you continue to experience strong and unshakeable intuitions in favor of I_1 and I_2.

Obviously, not everyone is in a predicament like this. Peter Singer is not, since he rejects I_2. Judith Thomson is not,[74] since she rejects I_1. And a host of others—people such as Michael Otsuka, Frances Kamm, Philip Quinn, Ralph Wedgwood, etc.—claim to know of a morally relevant difference between cases like $Push_{20}$ and $Switch_{20}$, so they are not in this predicament either. Nevertheless, it is not remotely uncommon for people to find that I_1, I_2, and No Difference all seem true. Indeed, the fact that these three claims all have a certain degree of plausibility is undoubtedly a primary reason why pairs of cases like $Switch_{20}$ and $Push_{20}$ have managed to attract so many philosophers' attention in the first place. So it is worth asking what a person who is strongly drawn to these three claims ought to believe.

Consider what I call the *Parity Solution*. According to the Parity Solution, all of the following judgments are true:

I_1: The agent in $Switch_{20}$ faces an absolute moral obligation to throw the switch.

S_1: The agent in $Switch_{20}$ faces an absolute moral obligation *not* to throw the switch.

I_2: The agent in $Push_{20}$ faces an absolute moral obligation *not* to push the football player.

S_2: The agent in $Push_{20}$ faces an absolute moral obligation to push the football player.

[74] Here I refer to Thomson's surprising change of mind about trolley cases (2008). Thomson is now of the view that it is morally wrong to throw the switch in Switch-style cases, and thus there is no morally relevant different between these cases and Push-style cases. So her current view is the polar opposite of Peter Singer's view. She and Singer agree that the cases are morally similar, but radically disagree with regard to *how* they are similar.

If the Parity Solution is true, then I_1 and I_2 are both true. Yet, given the Parity Solution, No Difference is also true—there is no moral difference between the agent's obligations in $Switch_{20}$ and the agent's obligations in $Push_{20}$—and therefore the Difference Requirement is false. The Parity Solution thus allows us to endorse I_1 and I_2 while at the same time enabling us to call off the search for a morally relevant difference between $Switch_{20}$ and $Push_{20}$.

These considerations are the makings of an argument in favor of the possibility of moral dilemmas. Such an argument would proceed in three steps. Step 1: Take I_1, I_2, and No Difference as premises, and draw the logical conclusion that the Difference Requirement is false. Step 2: Observe that the Parity Solution handily explains *why* the Difference Requirement is false, so conclude that the Parity Solution is likely to be true. Step 3: Notice that if the Parity Solution is true, then $Switch_{20}$ and $Push_{20}$ both count as moral dilemmas—so conclude that moral dilemmas are possible.

Whether or not a fully developed version of this argument can be persuasive depends on a number of factors. Crucially, it depends on whether the proponent of the Parity Solution can offer a satisfactory theoretical explanation of *why* the Parity Solution might be true. In order to describe what such an explanation might look like, I will draw an analogy.

To begin with, consider that there might be multiple normative domains. Morality is one such domain. But there might also be *aesthetic* obligations (obligations to create beauty) and *prudential* obligations (obligations to pursue one's own self-interest). It seems that obligations from different domains can conflict with one another. Suppose that Cinderella is going to a ball and can wear either of two different pairs of shoes: the slippers (which are most beautiful) or the sneakers (which are most comfortable). She must wear exactly one of these pairs of shoes. (It is out of the question to go barefoot and these are the only shoes around.) It is in Cinderella's self-interest to be comfortable, so she has a prudential obligation to wear the sneakers; but as the slippers are most beautiful, she has an aesthetic obligation to wear the slippers. Cinderella has two obligations and she must violate at least one of them. This looks like some kind of a dilemma—but what kind? Let us say that an agent is *aesdentially* obligated to do A exactly if the agent is either aesthetically obligated to do A or prudentially obligated to do A.

Then we have a name for the sort of dilemma faced by Cinderella. Cinderella faces an *aesdential dilemma*: Every action available to Cinderella would violate at least one of her aesdential obligations.

Cross-domain dilemmas, such as aesdential dilemmas, seem not to be problematic. So perhaps the moral domain is like the aesdential domain: not really a *unified* domain at all, but instead a union of two different (sub-)domains. Imagine, for instance, that the moral domain contains two different kinds of obligations: obligations to avoid causing harm to innocent persons, and obligations to maximize the good. Call the former *NC obligations* and the latter *C obligations*. Further suppose that an obligation counts as an absolute moral obligation iff it is *either* an NC obligation or a C obligation. Then moral dilemmas are probably fairly common. They occur whenever an action that maximizes the good causes harm to innocent persons.[75]

The view I am developing here contains two moral principles:

C: An agent has an absolute moral obligation to do A if A maximizes the good.

NC: An agent has an absolute moral obligation not to do A if A causes harm to innocent persons.

Call the conjunction of these principles "C-NC." C-NC is a moral theory in its own right; its success or failure as a theory will depend on whether it favorably compares to alternative theories. Since there are many alternative theories to choose from, I cannot complete such a comparison here. But I will make a few suggestions about what such a comparison would look like.

A main feature of C-NC is that it implies the existence of what I call *shadow obligations*. A shadow obligation is an absolute moral obligation to do A that is not revealed by intuition. S_1 and S_2 posit shadow obligations in $Switch_{20}$ and $Push_{20}$. (Most of us experience no intuition that the agent in $Switch_{20}$ has an absolute moral obligation not to throw the switch—and most of us experience no intuition that the

[75] Here, of course, we need a theory of value in order to know what it is to *maximize the good*. So a full theoretical explanation for the Parity Solution would need to offer a theory of value—just as any fully developed version of consequentialism would need to do so. But here I think we can remain vague. The outcome in which twenty people die seems obviously worse than the outcome in which one person dies. Any theory of value that ratifies this axiological intuition will do for present purposes.

agent in $Push_{20}$ has an absolute moral obligation to push the football player.) So the proponent of C-NC needs to explain why these obligations fail to show up on our intuitive radar.

Perhaps the need to offer such an explanation puts C-NC at a disadvantage relative to other theories. But this disadvantage is mitigated by the fact that most of us already believe that there are many shadow obligations. It would be surprising, after all, if *all* of our absolute moral obligations were revealed to us through intuition. Here we ought to distinguish between the relatively ordinary idea that our intuitions are *reliable* (such that if we intuit that A is obligatory, then A is probably obligatory) and the more ambitious idea that our intuitions are *sensitive* (such that if A is obligatory, then we probably intuit that A is obligatory). C-NC, to its credit, seems to be fully consistent with a view of our intuitions as reliable. For instance, C-NC implies that our intuitions in favor of I_1 and I_2 are both true. C-NC, perhaps unexpectedly, possesses a high degree of *intuitivity*—that is, C-NC explains and implies a very wide range of our intuitions about particular cases. [76]

And note that C-NC is a *very simple* theory. C-NC's combination of simplicity and intuitivity is rare among moral theories; it is often assumed that we must choose either intuitive-yet-complex theories (like Frances Kamm's intuitionist non-consequentialism[77]) or simple-yet-counterintuitive theories (like act-utilitarianism). But if we accept C-NC, we do not have to choose between the twin virtues of simplicity and intuitivity. That is a major advantage of C-NC. Still, it must be admitted that C-NC is

---

[76] I have no room to substantiate this claim here—it would require me to discuss our intuitions about a wide range of different cases. But let me note that a counterexample to my claim would have to take one or the other of two forms:

> (a) An agent seems to violates an absolute moral obligation by doing A rather than B *even though* A maximizes the good and causes no harm to innocent persons.
> (b) An agent seems to have an absolute moral obligation to do A rather than B *even though* A causes harm to innocent persons without maximizing the good.

There are *some* cases that might be like this. For instance, consider a famous case due to Williams: A husband can either rescue his wife from drowning or rescue two strangers from drowning, but not both. If the husband saves the two strangers and lets his wife drown, it seems like he might violate an absolute moral obligation in doing so—even though it is at least arguable that the husband maximizes the good in doing so while merely allowing (not causing) harm to his wife. To resolve this difficulty, however, the proponent of C-NC could simply add a third principle to cover role-related moral duties. An increase in the number of principles from two to three is in the *spirit* of C-NC.

[77] Frances Kamm (2006) has made the most thorough attempt to develop a highly intuitive non-dilemmatist version of non-consequentialism. The theory she has developed has many virtues, but no one believes that simplicity is among them.

probably *not* consistent with a view of our intuitions as sensitive. This is a cost of C-NC, although I suspect it is not a large cost.

C-NC, like any view according to which moral dilemmas are possible, has to address various quasi-logical problems raised by moral dilemmas.[78] Let us consider one such problem, which comes in the form of an argument containing two premises:

> OIC: If an agent has an absolute moral obligation to do A, then the agent can do A.
>
> Agglomeration: If an agent has an absolute moral obligation to do A, and has an absolute moral obligation to do B, then the agent has an absolute moral obligation to do both A and B.

To see how OIC and Agglomeration rule out the possibility of moral dilemmas, assume a moral dilemma. An agent has an absolute moral obligation to do A and has an absolute moral obligation to do B even though the agent cannot do both A and B. Given Agglomeration, the agent has a moral obligation to do both A and B. But given OIC, the agent can do both A and B. This contradicts our assumption that A and B form a dilemma.[79]

Thus proponents of C-NC must reject either OIC or Agglomeration. I recommend that proponents of C-NC ought to reject Agglomeration. The usual problem with this move is that Agglomeration seems to be accurate as a general *rule of thumb*.[80] So if we reject Agglomeration, it is necessary to replace it with an alternative principle. For instance, we might try this alternative:

---

[78] Here I refer to problems raised by various principles that belong to so-called "deontic logic." I call these problems *quasi-logical* because I think that the principles that raise these problems are not genuinely *logical* principles at all. I take Agglomeration and OIC, for instance, to be substantive assertions.

[79] See Brink 1994 for discussion of this type of argument.

[80] Pietroski 1993 says this. The solution I am pursuing here resembles Paske's solution (1990).

> Agglomeration*: If an agent has an absolute moral obligation to do A, and has an absolute moral obligation to do B, *and the agent can do both A and B*, then the agent has an absolute moral obligation to do both A and B.

If Agglomeration* were true, it would make moral dilemmas possible while preserving the accuracy of Agglomeration as a general rule of thumb. But Agglomeration* is hard to justify, since it adds an apparently ad hoc clause.[81]

However, proponents of C-NC can replace Agglomeration with a non-ad hoc alternative. According to C-NC, the moral domain is really a union of two different sub-domains—one that is governed by C, and another that is governed by NC. Nothing stops the proponent of C-NC from accepting these two principles:

> Agglomeration-C: If an agent has a C obligation to do A, and has a C obligation to do B, then the agent has an absolute moral obligation to do both A and B.

> Agglomeration-NC: If an agent has an NC obligation to do A, and has an NC obligation to do B, then the agent has an absolute moral obligation to do both A and B.

But the proponent of C-NC ought to reject this principle:

> Agglomeration-C/NC: If an agent has a C obligation to do A, and has an NC obligation to do B, then the agent has an absolute moral obligation to do both A and B.

The rejection of Agglomeration-C/NC can be justified by the analogy between absolute moral obligations (as they are understood by C-NC) and aesdential obligations.[82] Thus the proponent of C-NC can reject

[81] Zimmerman (2007 p. 214) makes this point.

Agglomeration-C/NC on principled grounds, while endorsing Agglomeration-NC and Agglomeration-C in order to explain why Agglomeration seems to be reliable as a rule of thumb.

I have only described one quasi-logical problem for moral dilemmas. There are many others; I cannot address them all here. But I hope that what little I have said suggests that proponents of C-NC are likely to be able to provide *some* sort of an answer to such problems.

A different objection to C-NC is that it is useless for moral decision-making. Recall the univocal verdict model of moral reasoning from Section 1. We tend to think that any minimally satisfying moral theory must provide us with some guidance about *what to do*. But in cases like Switch_{20} and Push_{20}, C-NC provides no such guidance. In fact, an agent in Switch_{20} or Push_{20} who is confused about what to do will likely be even *more* confused after she consults with a proponent of C-NC.

However, there are in fact many cases in which C-NC *does* provide guidance about what to do. Imagine a case in which an agent must decide whether to do either A or B. Suppose that A has two counts against it: A does harm to an innocent person (and thus violates NC) *and* A fails to maximize the good (and thus violates C). By contrast, B has no counts against it: B maximizes the good while avoiding harm to any innocent people. Then, in *this* case at least, C-NC provides clear guidance.

And consider another type of case. Imagine that the agent must decide whether to do J, K, or L. Suppose that J violates NC, K violates C, and L violates *both* NC and C. In this case, C-NC provides to the agent no clear guidance about what to do, but it *does* provide to the agent clear guidance about what *not* to do. C-NC clearly recommends not to do L. This negative guidance is less than we might hope for, but it is better than nothing.

So it is simply not the case that C-NC cannot give guidance. Instead, there are *some* cases in which C-NC provides clear *positive* guidance; other cases in which C-NC fails to provide clear positive

---

[82] Remember Cinderella. Suppose that she has a *prudential* obligation to do A and an *aesthetic* obligation to do B. This is consistent with the possibility that Cinderella has neither a prudential obligation to do both A and B, nor an aesthetic obligation to do both A and B. Therefore, it does not follow that Cinderella has an *aesdential* obligation to do both A and B.

guidance, but still provides clear *negative* guidance; and yet other cases (like $Switch_{20}$ and $Push_{20}$) in which C-NC provides no guidance at all.

Still, I think there is an interesting objection to C-NC in this vicinity. If we accept the univocal verdict model of moral reasoning—according to which a verdict about what to do is the ultimate aim of moral reasoning—then proponents of C-NC have to say that it is just not possible to reason successfully about cases like $Switch_{20}$ and $Push_{20}$. This implication seems objectionable.

By contrast, suppose that we accept a different view of moral reasoning. Suppose we say that moral reasoning simply aims to uncover the *most important* (i.e. top-ranked) moral properties of actions. We might *hope* that by revealing such properties, we will gain some guidance about what to do; but it remains logically possible that moral reasoning about a particular case might be successfully completed even though no clear verdict about what to do is reached. If we accept *this* view, then it is possible to accept C-NC while also saying that it is possible to reason successfully about cases like $Switch_{20}$ and $Push_{20}$.

Thus the proponent of C-NC seems to have two main options. She can accept the univocal verdict model of moral reasoning and *also* accept the distasteful implication that we are incapable of reasoning successfully about cases like $Switch_{20}$ and $Push_{20}$. Or she can reject the univocal verdict model of moral reasoning and thereby gain the right to claim that we can reason successfully about cases like $Switch_{20}$ and $Push_{20}$. I think the latter option is preferable, but both options are on the table.

I think the defensibility of C-NC will ultimately come down to two factors. First, it will depend on whether we end up being pushed toward moral dilemmas by pairs of cases like $Switch_{20}$ and $Push_{20}$. If we are able to reject either I_1, I_2, or No Difference, then the Parity Solution is unattractive, and we have no need for a theory like C-NC. Second, the defensibility of C-NC will depend on whether there are fatal problems associated with it. If, for instance, it turns out that moral dilemmas are logically incoherent, or that intuition is perfectly sensitive after all, or that the moral domain is not divisible into multiple sub-domains—then C-NC becomes indefensible. So the fate of C-NC and theories like it is quite uncertain. But I hope that I have said enough here to suggest that C-NC stands a chance of being true.

6. Dilemmatism and design

C-NC implies dilemmatism. According to C-NC, a moral dilemma occurs whenever an action that maximizes the good causes harm to innocent persons, and such situations seem to arise fairly often—much more often than we would like. So if C-NC stands a chance of being true, as I have argued, then dilemmatism also stands a chance of being true. And C-NC is not the only path to dilemmatism, which is after all a weaker claim than C-NC. Thus I think that I am entitled to assert that dilemmatism *could well be* true. Given this, it is worth investigating what would follow from it if it *were* true. That is what I will do in the next two sections. Here I will argue that dilemmatism is incompatible with one version of constructivism—the ideal designer theory—but is compatible with metaethical realism. In the next section (Section 7) I will argue that metaethical constructivism in general has difficulty accepting dilemmatism.

Recall the biological design hypothesis that I mentioned in the Introduction. Although the panda's thumb looks like a case of suboptimal design, it is entirely possible to imagine a designer with obscure purposes who would want the panda's thumb to be exactly as it is. Moreover, the panda's thumb is a much-discussed example precisely because clear cases of suboptimal design are relatively scarce in nature. Most species seem in most respects to be exceedingly *well*-designed. Human beings, for instance, seem quite well-made (even if not apparently *optimally* well-made) for the purpose of bringing more human beings into existence. So there is some initial plausibility in the idea that human beings (and other creatures) are the product of a designer who is highly skilled, although perhaps also slightly odd.

However, matters change when we shift from *individual* species to the *ecosystem* comprised by all species. As far as I can see, there is no evident purpose in the ecosystem as a whole. In fact, different organisms within the ecosystem seem to operate at cross-purposes. Bears make it harder for salmon to achieve their apparent purpose; human bear-hunters make it harder for bears to achieve their apparent purpose—etc. The ecosystem looks like a clock in which each gear cannot properly turn without

preventing other gears from turning. It seems that no designer worth his salt would make such a clock. Similarly, it is hard to imagine a designer worth his salt who would create anything like the chaotic diversity of cross-purposes that we observe in nature.

This critique of the design hypothesis is importantly different from what I take to be the usual "no-designer-worth-his salt" critique, which focuses on apparent *defects* in the design of particular organisms. This alternative critique focuses on the *apparent purpose* of various organisms.[83] So this critique grants to the design proponent that we can detect a certain kind of design in nature. For instance, it grants that bears seem exceedingly well-designed to do things like hunt prey, find mates, care for offspring, etc. Likewise, human beings seem exceedingly well-designed to do things like use tools, engage in philosophical thinking, kill bears, etc. According to the present critique, the problem for the design hypothesis arises not from a lack of apparent purposes, but from a conflict between apparent purposes.

Of course, we can imagine a designer who would create a world containing organisms that function at cross-purposes. A creator who approves of disorder might do so. So might a slow-witted creator who simply failed to anticipate that creatures like bears and humans would be at odds. Or perhaps there is a collective of multiple designers, each of whom have different aims. Such versions of the design hypothesis are immune to the critique from cross-purposes. Nevertheless, the critique from cross-purposes seems potentially very damaging to *standard* versions of the design hypothesis, like creationism.

Now recall the ideal designer theory, which I introduced above in Section 4. According to the ideal designer theory, the fact that an agent has an absolute moral obligation to do A is constituted by the fact that A is required by a moral code that *is* or *is as if it were* designed by an ideal designer. I want to argue that if dilemmatism is true, then the ideal designer theory is susceptible to a critique from cross-purposes, just as standard versions of creationism are.

---

[83] See Sober 2008 pp. 126-128 for further discussion of the "no designer worth his salt" critique. I am fairly confident that the present alternative critique—the "critique from cross-purposes"—is not original to me, but I do not know whether it has been articulated anywhere in the literature.

We should expect if a given moral code is a product of design, then the rules it contains will have been designed with certain purposes in mind. And the purposes of rules typically seem obvious. If there is a rule that requires us to avoid doing harm to innocent persons, then the apparent purpose of that rule is to guide us away from actions that harm innocent persons. And if there is a rule that requires us to maximize the good, then the apparent purpose of that rule is to guide us toward actions that maximize the good. It is easy to see evidence of design in *either one* of these rules. But it is harder to see evidence of design in a moral code that contains *both* of these rules.

A list of instructions that contains both of these rules would be puzzling. If you were given such a list, you might think it was created by a prankster in order to cause confusion. Or perhaps it was created by a slow-witted person who failed to anticipate that her rules might come into conflict in certain cases. Or perhaps the rules were each contributed by different, independent individuals. Such explanations seem prima facie plausible. But it does not seem plausible to think that such a list was created by an ideal designer. We expect an *ideal* designer to anticipate the possibility of conflict between her rules and to modify or add to them in order to remove such conflict. Consequently, we expect that a moral code that is (or is as if it were) designed by an ideal designer would not contain or imply requirements that *routinely* conflict with one another. Dilemmatism and the ideal designer theory thus seem to be incompatible.

By contrast, consider how a realist ought to regard dilemmatism. For the realist, the moral truth exists independently of the psychology of all real or hypothetical beings, including the attitudes of ideal designers of moral codes. Realism is consistent with the view that our absolute moral obligations are without purpose, so it is consistent with the view that our absolute moral obligations are at apparent cross-purposes. Here it is useful to revisit the way in which evolutionary theorists regard biological phenomena.

Much of biology seems to serve us well. Our bodies, for instance, are often useful to us. But for the evolutionary theorist, the usefulness of our bodies is a fortunate side-effect of the constitution of the world in which we happen to live, not a result of anyone's intentional design. Thus evolutionary theorists are unsurprised that our bodies do not serve us *perfectly* well—as when we succumb to disease, injury, or old age.

Similarly, even for the proponent of C-NC, our absolute moral obligations often serve us well. In particular, our absolute moral obligations often provide us with usable guidance about what to do. But realists can say that the usefulness of our absolute moral obligations for decision-making purposes is just a fortunate side-effect of the way in which moral reality happens to be constituted; it is not as if our absolute moral obligations were made by someone *in order* to give us guidance. So a realist need not be surprised if our absolute moral obligations often fail to provide us with a univocal verdict about what to do, as in moral dilemmas. In fact, in a thoroughgoing realist view, what is truly surprising is that moral reality often *is* useful for our decision-making purposes—perhaps in something like the same way that our undesigned biological ecosystem is surprisingly amenable to human goals.

7. A dilemmatist case against constructivism

We can draw a three-part distinction among constructivist moral theories. First, there are *ideal constructivists*, who hold that the moral truth is fixed by psychological features of an idealized being (or beings)—versus *non-ideal constructivists*, who hold that the moral truth is fixed by psychological features of some *non*-ideal and presumably actual being (or beings). Second, there are *single-source constructivists*, who hold that the moral truth is fixed by psychological features of a *single being* (whether ideal or non-ideal)—versus *multiple-source constructivists*, who hold that the moral truth is fixed by psychological features of multiple beings. Third, there are *convergentist constructivists*, who hold that there is a single moral truth that applies to all agents—versus *divergentist constructivists*, who hold that divergent moral truths apply to different agents.

The ideal observer theory and the ideal designer theory are the main versions of single-source ideal constructivism. The former theory takes an ideal agent's *advice* to be constitutive of the moral truth; the latter theory takes an ideal agent's *moral code* to be constitutive of the moral truth. If my arguments in this paper are successful, both of these theories are incompatible with dilemmatism.

Could a *multiple-source* ideal constructivist accept dilemmatism? This is arguable. Suppose that omniscience and benevolence are the essential features of an ideal agent. Perhaps there are many different hypothetical beings that have just those qualities. These ideal beings might offer different advice or recommend different rules. So imagine, for instance, a constructivist theory according to which there is an absolute moral obligation to do A iff A is required by *at least one* hypothetical ideal designer's moral code. Such a theory would entail that moral dilemmas occur in any case in which A is required by one hypothetical ideal designer's moral code, and B is required by another hypothetical ideal designer's moral code, although the agent cannot do both A and B.

However, it seems to me unlikely that *suitably* idealized agents would disagree extensively enough to make moral dilemmas widespread. A suitably ideal designer would foresee that other ideal designers' moral codes would conflict with her own, and this seems likely to influence her in such a way as to minimize the occurrence of moral dilemmas.

So, I maintain, dilemmatism is at odds with ideal constructivism—whether single-source or multiple-source. What about *non-ideal* constructivism? First of all, note that single-source non-ideal constructivism is not plausible. A version of single-source non-ideal constructivism would hold that the moral truth is fixed by the deliverances of a single non-ideal person. For instance, there might be a version of this view according to which *my own moral views* fix the moral truth (whereas the moral views of everyone else are irrelevant to the moral truth). No one accepts a view like this.

So it seems that non-ideal constructivists ought to embrace some form of multiple-source, non-ideal constructivism. Such a position can be either convergentist or divergentist. However, convergentist versions of non-ideal multiple-source constructivism turn out to be difficult to defend. It is hard to see how a single moral truth that applies to all agents might be fixed by the attitudes of multiple non-ideal agents.

Thus, if dilemmatism is true, it looks as if constructivists ought to endorse some form of non-ideal multiple-source *divergentist* constructivism. And in fact, these kinds of theories do handle dilemmatism quite well. As an example of this kind of view, consider social constructivist theories,

according to which the moral truth relative to a given society is fixed by the moral code accepted in that society. The moral code accepted in any given society tends to be a hodge-podge of different rules and ideas cobbled together over time by different contributors, so it is unsurprising if they occasionally (or not-so-occasionally) come into conflict with one another. Given this, a social constructivist approach can easily predict and explain the frequent occurrence of moral dilemmas.[84]

But the divergentist idea that moral truth varies from culture to culture—or from agent to agent—comes with a variety of famous disadvantages. For instance, social constructivists have a hard time explaining how we can justifiably criticize the moral beliefs of different cultures, and how individuals from different cultures can have genuine moral disagreements. Realism avoids these disadvantages by positing a single moral truth that is independent of the attitudes of any given individual or culture. A dilemmatist who wants to avoid such disadvantages ought to be a realist; and a constructivist who wants to avoid such disadvantages ought to oppose dilemmatism.

Conclusion

This paper is intended as a contribution to the debate between metaethical constructivists and metaethical realists. That debate is *intramural* within cognitivist non-nihilism. Constructivists and realists are both cognitivists, since they both hold that moral judgments are beliefs (or belief-like states); and they are both non-nihilists, since they both hold that some of our moral judgments are true. Their primary disagreement concerns the *explanation* for the truth (or falsity) of any given moral judgment. Constructivists, as we have seen, believe that the truth of a moral judgment depends on the psychological features of some real or hypothetical intelligent being or beings, whereas realists deny this. A good number of philosophers accept *some* form of cognitivist non-nihilism and therefore have a stake in the realism/antirealism debate. So my arguments in this paper do have fairly broad significance. But it bears emphasizing that my arguments do not apply to those who either reject cognitivism, or accept some kind of nihilism.

[84] Dahl 1996 makes this point in a usefully clear way.

The upshot of my arguments in this paper is that the possibility and actuality of moral knots provide ammunition for realists against constructivists. I argued in sections 3-4 that even moral knots in purely hypothetical situations of infinite choice cause difficulties for *some* constructivists, namely ideal observer theorists. And constructivists ought to be *especially* wary of dilemmatism, the view that moral dilemmas occur routinely in ordinary life, since (as I argued in sections 6-7) few tenable constructivist theories are compatible with dilemmatism. Moreover, dilemmatism is not just an imaginary threat to constructivism. Dilemmatism stands a reasonable chance of being true, according to my argument in Section 5. Thus I conclude that constructivists ought to be ready to argue against dilemmatism, and realists (*qua* realists) ought to hope for a persuasive argument in favor of dilemmatism.

Bibliography

Audi, Robert. 2005. *The Good in the Right*. Princeton University Press.

---. 2010. "Moral Perception and Moral Knowledge" *Proceedings of the Aristotelian Society* 84.

Bealer, George. 1998. "Intuition and the Autonomy of Philosophy" in DePaul and Ramsey 1998.

Bonjour, Lawrence. 1980. "Externalist Theories of Empirical Knowledge." *Midwest Studies in Philosophy* 5.

Bedke, Matthew. 2009. "Intuitive Non-Naturalism Meets Cosmic Coincidence," *Pacific Philosophical Quarterly* 90.

Benatar, David. 2008. *Better Never to Have Been: The Harm of Coming Into Existence.* Oxford.

Brandt, Richard. 1954. *Hopi Ethics: A theoretical analysis*. University of Chicago Press.

Brink, David. 1994. "Moral Conflict and Its Structure." *Philosophical Review* 103.

Conee, Earl and Feldman, Richard. 1998. "The Generality Problem for Reliabilism." *Philosophical Studies* 129.

Christensen, David. 2007. "Epistemology of Disagreement: The Good News." *Philosophical Review* 119.

DePaul, Michael and Ramsey, William. 1998. *Rethinking Intuition*. Lanham: Rowman & Littlefield.

Doris, John and Plakias, Alexandra. 2008. "How to Argue about Disagreement" in Sinnott-Armstrong 2008b

Elga, Adam. 2007. "Reflection and Disagreement." *Nous* 41.

Foot, Philippa. 1967. The Problem of Abortion and the Doctrine of the Double Effect." *Oxford Review* 5.

Greene, Joshua. 2008. "The Secret Joke of Kant's Soul" in Sinnott-Armstrong [2008c].

Greene, Joshua et al. 2001. "An fMRI Investigation of Emotional Engagement in Moral Judgment." *Science* 14.

Gould, Stephen. 1980. *The Panda's Thumb*. WW Norton & Co.

Haidt, Jonathan. 2001. "The Emotional Dog and its Rational Tail." *Psychological Review* 108.

Hare, RM. 1981. *Moral Thinking*. Oxford University Press.

Harman, Gilbert. 1977. *The Nature of Morality*. Oxford.

Huemer, Michael. 2006. *Ethical Intuitionism.* Palgrave Macmillan.

Kamm, Frances. 2006. *Intricate Ethics*. Oxford.

Kelly, Thomas. 2006. "The Epistemic Significance of Disagreement." *Oxford Studies in Epistemology* 1.

Kierkegaard, Soren. 2006. *Fear and Trembling*, Evans and Walsh (eds.). Oxford.

Kraay, Klaas. 2010. "The Problem of No Best World," in Draper and Taliaferro (eds.), *A Companion to Philosophy of Religion* (2nd edition). Oxford.

Mason, HE, editor. 1996. *Moral Dilemmas and Moral Theory*. Oxford.

Marcus, Ruth. 1996. "More about Moral Dilemmas" in Mason 1996.

McBrayer, Justin. 2010. "Moral Perception and the Causal Objection," *Ratio* 23.

McGrath, Sarah. 2007. "Moral Disagreement and Moral Expertise." *Oxford Studies in Metaethics* 3.

Mintoff, Joe. 1997. "Slote on Rational Dilemmas and Rational Supererogation." *Erkenntnis* 46.

Paske, Gerald. 1990. "Genuine Moral Dilemmas and the Containment of Incoherence." Journal of Value Inquiry 24.

Pears, DF. "1976. The Causal Conditions of Perception." Synthese 33.

Pietroski, Paul. 1993. "Prima Facie Obligations, Ceteris Paribus Laws in Moral Theory." Ethics 103.

Portmore, Doug. 2003. "Position-Relative Consequentialism, Agent-Centered Options, and Supererogation." Ethics 113.

Railton, Peter. 1996. "The Diversity of Moral Dilemma" in Mason 1996.

Ross, David. 1988. The Right and the Good, Stratton-Lake (editor). Oxford.

Shafer-Landau, Russ. 2005. *Moral Realism: A Defence.* Oxford.

---. 2009. "A Defence of Categorical Reasons." Proceedings of the Aristotelian Society 109.

Singer, Peter. 2005. "Ethics and Intuitions." *The Journal of Ethics* 9.

Sinnott-Armstrong, Walter. 2002. "Moral Relativity and Intuitionism." *Philosophical Issues* 12.

---. 2008a "Framing Moral Intuitions" in Sinnott-Armstrong [2008b].

---. 2008b *Moral Psychology Volume 2*. Cambridge: MIT Press.

---. 2008c *Moral Psychology Volume 3*. Cambridge: MIT Press.

Slote, Michael. 1989. *Beyond Optimizing.* Harvard.

Sober, Elliott. 2008. *Evidence and Evolution*. Cambridge.

Sosa, Ernest. 1999. "How Must Knowledge Be Modally Related to What Is Known?" *Philosophical Topics* 26.

Stich, Steven. 2006. "Is Morality an Elegant Machine or a Kludge?" *Journal of Cognition and Culture* 6.

Street, Sharon. 2006. "A Darwinian Dilemma for Realist Theories of Value." Philosophical Studies 127.

Thomson, Judith. 1976. "Killing, Letting Die and the Trolley Problem." *Monist* 59.

Streumer , Bart. 2010. "Reasons, Impossibility, and Efficient Steps." *Philosophical Studies* 151.

Thomson, Judith. 1976. "Killing, Letting Die and the Trolley Problem." *Monist* 59.

---. 2008. "Turning the Trolley." *Philosophy and Public Affairs* 36.

White, Roger. 2005. "Epistemic Permissivism." *Philosophical Perspectives* 19.

Williams, Bernard. 1965. "Ethical Consistency." *Proceedings of the Aristotelian Society* 39.

Zimmerman, Michael. 2007. *The Concept of Moral Obligation*. Cambridge.