Of the Terrible Doubt of Appearances

An Essay in Moral Epistemology

by

David Richard Morrow

A dissertation submitted to the Graduate Faculty in Philosophy
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy,
The City University of New York
2009

This manuscript has been read and accepted for the
Graduate Faculty in Philosophy in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

11/21/2008	David Rosenthal
Date	Chair of Examining Committee
03/30/2009	Iakovos Vasiliou
Date	Executive Officer

Dissertation Committee

Catherine Wilson (advisor)

Michael E. Levin

Jonathan E. Adler

David M. Rosenthal

Jesse J. Prinz

THE CITY UNIVERSITY OF NEW YORK

Abstract

Of the Terrible Doubt of Appearances:

An Essay in Moral Epistemology

by

David R. Morrow

Advisor: Catherine Wilson

Recreational killing strikes most of us as wrong. Such "moral appearances," in which the world appears to us to be a certain way, morally speaking, play an important role in moral epistemology, usually in the guise of "moral intuitions."

Moral appearances are natural phenomena, however, and scientists are discovering the psychological mechanisms underlying them. Recent research suggests a "developmental sentimentalist" model of moral appearances, on which moral appearances arise from "moral sentiments," which develop through a process of emotional conditioning.

This naturalistic account of moral appearances allows us to explain our moral appearances without supposing that their intentional content is true. This explanatory irrelevance gives us a *prima facie* reason to discount moral appearances when deciding which moral claims to endorse. Sensibility theory and rational intuitionism attempt to validate the use of moral appearances in the face of their explanatory irrelevance. I argue that neither theory succeeds.

But it seems that moral appearances cannot be discounted altogether, for it is unclear how we could justify moral claims without them. I introduce the notion of

"practical coherence" as a basis for deciding between alternative systems of evaluative claims, including both moral and nonmoral claims. I assume that evaluative claims have, as at least one function, the prescription of actions. A system of evaluative claims is practically coherent to the extent that, given current circumstances, performing the actions prescribed by any one evaluative claim in the system increases, or at least does not reduce, the probability of being able to perform the actions prescribed by other claims in the system. Because the practical relations between different actions are determined by the world, not by what we think, practical coherence ties evaluative systems to the world. This dependence on both the values that we hold and the attitude-independent relations among various actions yields an unusual combination of limited ethical relativism and moderate moral realism. But more importantly, practical coherence leads to a multidisciplinary method of ethical inquiry that will allow us to devise more satisfying answers to the central question of ethics: How should one live?

Of the terrible doubt of appearances,
Of the uncertainty after all, that we may be deluded,
That may-be reliance and hope are but speculations after all,
That may-be identity beyond the grave is a beautiful fable only,
May-be the things I perceive, the animals, plants, men, hills, shining and flowing waters,
The skies of day and night, colors, densities, forms, may-be these are (as doubtless they are) only apparitions, and the real something has yet to be known,
(How often they dart out of themselves as if to confound me and mock me!
How often I think neither I know, nor any man knows, aught of them,)
May-be seeming to me what they are (as doubtless they indeed but seem) as from my present point of view, and might prove (as of course they would) nought of what they appear, or nought anyhow, from entirely changed points of view;
To me these and the like of these are curiously answer'd by my lovers, my dear friends,
When he whom I love travels with me or sits a long while holding me by the hand,
When the subtle air, the impalpable, the sense that words and reason hold not, surround us and pervade us,
Then I am charged with untold and untellable wisdom, I am silent, I require nothing further,
I cannot answer the question of appearances or that of identity beyond the grave,
But I walk or sit indifferent, I am satisfied,
He ahold of my hand has completely satisfied me.

— Walt Whitman, *Leaves of Grass* (1867)

Acknowledgements

Few scholarly works come into being without the support of family, friends, and colleagues. A dissertation, which marks the end of an academic's formal training, provides a particularly good opportunity to thank the numerous people who have played important roles in that training. In my own case, far too many people have helped me in far too many ways for me to thank them all. But there are some who ought not to go unmentioned.

A number of teachers have been especially important in my philosophical education: Susan Wolf, Catherine Wilson, Michael Williams, Jerry Schneewind, David Rosenthal, Michael Levin, Vaughn Huckfeldt, Daniel Callcut, and Stefan Baumrin deserve special mention in that regard.

This dissertation itself has benefited tremendously from the help of my dissertation committee. My advisor, Catherine Wilson, provided steady support, offered helpful suggestions, and asked insightful questions that pointed to new and better ways of approaching the material. Michael Levin and Jonathan Adler both read drafts of every chapter and provided detailed, thoughtful comments. In addition to serving on the final dissertation committee, David Rosenthal provided helpful feedback on an earlier version of Chapter Two. Jesse Prinz generously agreed to serve on my committee from afar, even though he was not yet on faculty at CUNY.

Chris Alen Sula also deserves special thanks, not only for reading and commenting on every chapter of this dissertation, but for being an invaluable philosophical interlocutor and friend throughout my time at The Graduate Center. I would not have learned half as much here without him.

The participants in the Cognitive Science Symposium at the CUNY Graduate Center, the participants of the First Ohio State University Ethics Conference, and two

anonymous reviewers at *Philosophical Psychology* all gave helpful feedback on various versions of what is now Chapter Two. A modified version of Chapter Two is forthcoming in *Philosophical Psychology* under the title "Moral Psychology and the 'Mencian Creature'." Thanks to Taylor Francis & Company for permission to reprint that material here.

Chapter Three benefited from the feedback I received from commentators and audience members at conferences at SUNY Albany and Ohio State in the spring of 2008. Special thanks to Hilary Kornblith and Robert Audi for their comments at those conferences.

Finally, I cannot exaggerate the debts that owe to my parents, Dan and Felicia, and to my wife, Melissa. Not only did their emotional and financial support make my education inestimably easier, but it is because of their salutary influence that I still insist, despite years of academic training, on trying to find something useful in philosophy. Like everyone that I thanked above, they have made me a better philosopher than I would otherwise be. But far more importantly, they have also made me a better person.

Table of Contents

1 Moral Appearances

When Bruce Wayne came of age, he had a problem. Having inherited a vast fortune and a multinational corporation from his late parents, this last scion of the Wayne family had to figure out how to spend his life. He had his pick of possibilities. He chose an unusual one: he became Batman.

Most of us don't have that option. But all of us face that same question: What should we do with our lives? This is not just a question about what career to pursue or whether to become a masked vigilante. It is about how we should live and about what we should value. And even once we make that choice, we face subsidiary questions about what to do in some particular set of circumstances, or how best to realize the values we do hold—either in a specific decision or in the grander scheme of life.

For most of us, answering this question involves more than mere soul-searching. Deciding how to live usually involves more than merely figuring out what one most *wants* to do, or what would bring one the most pleasure. We all recognize that some choices are better than others insofar as they satisfy our particular needs and desires; but for many of us, at least, there is also a suspicion that some choices may be better in a broader sense. We may wonder whether we have the "right" desires—whether we value the right things.

One goal of this dissertation is to figure out if that suspicion is well grounded. Another goal is to figure out whether "moral intuitions" are useful in figuring out which desires or values are the "right" ones to have. The third goal is to sketch a method for figuring out which are the "right" desires or values—a method that eschews moral intuition and yet shows how some desires or values can be "better" than others.

The purpose of pursuing these three goals is to point the way to better answers to a simple, but important, question: How should one live?

The most important interpretation of this question is the broadest one. We need to know how one should live, all things considered, not just how one should live "in the moral sense," or "from a prudential point-of-view," or from any other partial perspective. It is not enough to know that, morally speaking, Bruce Wayne ought to become Batman, but prudentially speaking, he ought to live out his life as a billionaire playboy. The question 'How should one live?' needs answering in a way that will tell Bruce Wayne whether, at the end of the day, he should don a cape and stalk the streets of Gotham.

1. A prescriptivist assumption

We might as well begin with an assumption. (There are always assumptions.)

One of the central assumptions behind this dissertation is that evaluative terminology, in at least one of its primary uses, is prescriptive. That is, to evaluate something positively is to recommend it; to evaluate something negatively is to recommend against it. More precisely, to evaluate something positively is to recommend achieving it, promoting it, acquiring it, imitating it, etc., and to evaluate something negatively is to recommend against doing any of those things. For instance, to call a book good is to recommend that a reader read it, or that a writer write a book like it (in some respect). To call a movie bad is to recommend a viewer not see it, that a screenwriter not write a movie like it, or that an actor not star in a movie like it.

This is not to say that the only thing evaluative language does is recommend things or actions. There may be other primary uses, though I don't think so, and there are certainly secondary uses. We can use evaluative language in an "inverted commas"

sense, in which we apply an evaluative term, like 'good', in the way that some other people apply it, without endorsing their recommendation of the thing we called "good." This is especially true when there are well-established and publicly known standards by which a thing is measured. For instance, a disgruntled student might talk about the "good students" who devote long hours to their schoolwork, always earn excellent grades, never cut class, and so forth. But if she disdains the system wherein they excel, this disgruntled student might not be recommending that anyone imitate the "good students" in any way.

The point of articulating this assumption is that it links questions about what is valuable to our central question of how to live. To call something or someone 'good', or to use "thicker" evaluative terms like 'just' or 'courageous', is (perhaps among other things) to endorse a particular constellation of actions related to the thing or person being endorsed. Whether calling something "good" says anything more about the thing being recommended remains to be seen.

The reasons for (and against) this prescriptive premise have been much discussed. There is a voluminous philosophical literature on the subject. The view found its most vocal advocate in R. M. Hare, though it has earlier roots in both emotivists, especially A. J. Ayer and Charles Stevenson, and American Pragmatists, especially John Dewey. Among the most important of Hare's arguments for the claim that evaluative terms are primarily prescriptive is the parable of the cannibals and the missionary.

Imagine, urges Hare, a tribe of cannibals who call men "good" based on the number of heads they collect—the more the better. Imagine, also, that a missionary comes to this tribe and deigns to enlighten them about goodness. The missionary finds the meekest men in the tribe, the ones who are kind and gentle and never kill or behead anyone. He points to them and says, "These are the good men in your tribe." He and the

cannibals disagree about what makes a man good. They disagree about what kind of man one should be.

This disagreement is only possible if evaluative terms like 'good' function differently than descriptive words. If the missionary had gathered the shortest men in the tribe and, pointing to them, said, "Here are the tallest men in your tribe," we would not construe him as disagreeing with the cannibals about height. He would be using the cannibals' word for "tall" incorrectly. The correct response is not to argue with him about who is taller, but to correct him in his use of the word. The (fictional) fact that the cannibals argue with the missionary about what makes a man "good," rather than correcting his usage, shows that the word for 'good' does not *mean* anything like "prone to beheading others." Rather, its meaning lies in prescribing or recommending various actions.[1]

We might extend Hare's thought experiment a step further, to test his conclusion that evaluative terms prescribe actions. What would the missionary think if the cannibals assented to calling the meek men "good," but continued in their homicidal ways? Suppose they continued to kill, dismember, and eat their enemies. Suppose they continued to boast of their beheadings, praise and fawn over others for their brutality, and disdain the gentle tribesmen whom they now called "good." If 'good' did not prescribe anything, then this should not perturb the missionary. But of course, the missionary would think that they were merely appeasing him—that they did not really think kindness or compassion to be good or valuable, but were merely mouthing the words he wanted to hear. Other things being equal, we do not take someone to think

[1] R. M. Hare, *The Language of Morals* (Oxford: Clarendon Press, 1952).

that something is good if he or she shows no inclination to pursue, promote, or maintain it.[2]

I do not intend to defend this prescriptivist assumption any further, but most of what follows depends on it. Readers who deny this premise will therefore disagree with much, though perhaps not all, of what I have to say. I am willing to accept that.

2. The basic problem for the epistemology of evaluation

The link between evaluative terms and prescriptions for action ties all evaluative claims to a supposedly off-hand comment in David Hume's *Treatise of Human Nature*. "I cannot forbear adding to these reasonings," wrote Hume,

> an observation, which may, perhaps, be found of some importance. In every system of morality, which I have hitherto met with…the author proceeds for some time in the ordinary way of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when of a sudden I am supriz'd to find, that instead of the usual copulations of propositions, *is*, and *is not*, I meet with no proposition that is not connected with an *ought*, or an *ought not*. [But] as this *ought*, or *ought not*, expresses some new relation or affirmation, 'tis necessary that it shou'd be observ'd and explain'd; and at the same time that a reason shou'd be given, for what seems altogether inconceivable, how this new

[2] *Cf.* C. L. Stevenson, "The Emotive Meaning of Ethical Terms," *Mind* 46, no. 181 (1937): 19.

> relation can be a deduction from others, which are entirely different from it.[3]

Hume's observation has, in fact, been found of "some importance." It is often taken to underwrite the basic epistemological problem in ethics (and value theory more generally). This problem is sometimes called the "is–ought gap." It is often summarized in the admonition, "You can't infer an 'ought' from an 'is'," which is sometimes referred to as "Hume's Law." Some philosophers have elevated the is–ought gap, which is, in Hume's eyes, an epistemological gap, into a metaphysical one—a distinction between the realm of fact and the realm of value.

Whatever the merits of that metaphysical division, Hume's Law is correct as a point of basic logic. The conclusion of a valid inference may not contain terms that do not appear in any of the premises. Thus, no inference can be cogent if 'ought' appears in its conclusion but not its premises. From premises solely about what is, we can draw no conclusion about what ought to be.

Putative counterexamples to Hume's Law rely on unstated premises. For instance, it seems reasonable to say that countries ought not to engage in nuclear war because millions of civilians are certain to die in the process. The conclusion of this inference contains an 'ought', while the only stated premise does not. Yet this is not a genuine counterexample to Hume's Law, because basic principles of logic require an unstated premise in the argument—e.g., 'Countries generally ought not to do things that are certain to kill millions of civilians', or at least 'If nuclear war will kill millions of civilians, then countries ought not to engage in it.' This has nothing in particular to do with morality or the difference between 'is' and 'ought'. As Anscombe notes, Hume's

---

[3] David Hume, *A Treatise of Human Nature* (1740), III.i.1.27.

Law applies just as much to inferences from 'is' to 'owes' or 'is' to 'needs'.[4] The rule applies just because, for any inference in which the conclusion contains a term absent from the stated premises, we must supply an additional, unstated premise that contains the missing term.

Because the unstated premises linking descriptive premises (about what "is") to prescriptive conclusions (about what "ought" to be) must contain prescriptive terms (e.g., 'ought'), they are prescriptive claims. Thus, these unstated premises, which bridge the gap between 'is' and 'ought' in ordinary moral reasoning, cannot be derived from purely descriptive premises. What makes it especially hard to justify 'ought' statements is that there is little or no consensus on these "bridge premises," as there is about premises connecting claims about what someone "needs" to other kinds of claims.

One might think that certain "thick" evaluative terms like 'justice' or 'deceit' might help. 'One ought not to act unjustly' seems uncontroversial, and so could serve as a useful bridge premise. However, the prescriptivist assumption stated above puts *all* evaluative claims on the "ought" side of the is–ought gap. Though the "surface grammar" of 'Deceit is wrong' is descriptive, Hume's Law implies that 'Deceit is wrong' cannot be derived from purely descriptive premises because 'wrong' implies a set of 'oughts'.

To simplify discussion of 'ought'-statements and statements containing evaluative language, I will stipulate a technical meaning for the word 'evaluation'. An "evaluation," as I will understand it here, is a claim in which either subject and predicate are joined by an 'ought' or 'should'; or a claim in which a subject is joined to an evaluative predicate with 'is' (or some grammatical variation on 'is'); or anything logically equivalent to such a claim. Thus, 'The United States ought to lift the embargo on Cuba', 'Fidel Castro is evil', 'All Communist dictators are evil', 'If someone is a

---

[4] G. E. M. Anscombe, "Modern Moral Philosophy," *Philosophy* 33, no. 124 (1958).

Communist dictator, then he or she is evil, 'We should go to that new Cuban restaurant', and 'You are an excellent salsa dancer' are all evaluations.

The basic problem for the epistemology of evaluations is that Hume's Law creates a special case of an ancient epistemological quandary. Any argument for an evaluation requires an evaluation as a premise. But the conclusion of any argument is justified only if the premises are justified. Thus, we can give a cogent argument for an evaluation only if we are justified in believing some further evaluation. If we try to justify our second evaluation with an argument, then we will need a third evaluation as a premise. If we try to justify that third evaluation with argument, we will need a fourth, and so on indefinitely. But none of the arguments succeed in justifying their conclusions until this regress ends, which it never does. Hence, unless we can find some way to escape this regress, we can never justify any evaluation.[5]

In epistemology generally, philosophers typically respond to this challenge in one of three ways. "Skeptics" admit that we cannot justify our beliefs—that we do not know anything. "Foundationalists" claim that some premises can be justified without argument. In non-evaluative cases, at least, many philosophers maintain that we can justify some premises through sensory experience or rational reflection without relying on argument. Many empiricists hold (or have held) that my belief that there is something blue in front of me is "self-justifying." Many rationalists hold that my belief that two times three is six is "self-evident" (once I understand the words involved). Alternatively, "coherentists" claim that in a sufficiently rich "web of beliefs," our beliefs can justify one another without any of them serving as "foundations" for the rest. Even coherentists, however, generally allow a special role for sensory experience. Although

---

[5] *Cf.* Michael Williams, *Unnatural Doubts* (Princeton: Princeton University Press, 1996), 60–68.

sensory experiences are not "self-justifying," they provide an essential link to the external world, without which we would know little or nothing.

These three traditional responses to the general epistemological regress problem suggest analogues in the epistemology of evaluations.

3. Desire and the untenability of evaluative skepticism

Skepticism is a peculiar philosophical position. Never have so many tried so hard to refute a position that is held by so few. Even philosophers who take the arguments for skepticism seriously act as if they know things. In particular, they act as if their senses give them knowledge. In practice, desires undermine evaluative skepticism much as sensory experiences undermine skepticism about the external world.

Total evaluative skepticism, according to which nothing is good or bad in any sense, is a deeply unappealing position. We cannot help but have desires, and other things being equal, having a desire generally leads us to think that something is good or that we ought to act in a particular way. Thus, desires cause us, at the very least, to *act* as if we accepted certain evaluations. Someone could insist that he is *merely acting* as if he accepted particular evaluations, but this is a kind of skepticism that is not worth arguing with. As Peirce says, "Let us not pretend to doubt in philosophy what we do not doubt in our hearts."[6]

Total evaluative skepticism may even be self-defeating. To suppress all desire requires great effort, which no one would undertake without a desire to do so.

Just as conflicting sensory experiences provoke revision of our beliefs about the external world, so conflicting desires provoke revisions of our evaluations. Someone sees

---

[6] C. S. Peirce, "Some Consequences of Four Incapacities Claimed for Man," *Journal of Speculative Philosophy* 2 (1868).

a stick "bend" as it enters the water, but feels that it is straight to the touch. It cannot be both bent and straight. One experience must be rejected. Someone wants to take a month's vacation, but also wants to keep her job. Her supervisor tells her she will be fired if she leaves for a month. One desire must be suppressed, and the prescription it entails rejected.

Thus begins the epistemology of evaluation.

But if desire derails evaluative skepticism, there is reason to wonder whether it alone can underwrite a non-skeptical evaluative epistemology. Conflicting desires may provoke us to revise our evaluations. They may stimulate us to develop a coherent set of desires. But many of us have a nagging suspicion that this is not enough—that there is still some sense in asking whether we have the "right" desires, or whether there are some things that we ought to desire independently of our current desires. In fact, many philosophers think that "ethical" thought begins only when we set aside our own desires and think about evaluations from some larger perspective.

Where does this suspicion come from?

4. Moral appearances and moral phenomenology

Some choices seem to depend merely on our desires. If I prefer ravioli to linguini, there is usually nothing beyond my preferences to guide the decision between them. But some choices present themselves as something more. As Maurice Mandelbaum reports his experience in *The Phenomenology of Moral Experience*, "in some choices we feel that one of the alternatives places a *demand* upon us, that we are obliged, or bound, to act for it. This feeling of obligation appears as being independent of preference, as many of the

alternatives within our experience do not."[7] Furthermore, these obligations do not arise just from the expectations of others. We do not feel them as obligations of convention or etiquette. They are perceived as independent not only of our own desires, but of others' desires, too.

Mandelbaum is right, I think, that many people have experiences of this kind. It is through this kind of experience that we come to suspect that there is more to evaluative epistemology than thinking hard about what we most desire and how we can best achieve those desires. Thus, any attempt to answer our central question—How should one live?—should take these experiences into account.

Mandelbaum suggests that "morality" enters our experience through these perceived external obligations. This kind of experience underwrites the distinction between evaluations grounded in our own desires and "moral evaluations."

But in accepting the importance of the experiences that Mandelbaum mentions, we must be careful to avoid begging any substantive questions about morality or moral psychology. There are two points to notice in particular. First, Mandelbaum says that we "feel" obliged or bound by certain choices. But there is significant philosophical disagreement about how we come to have these experiences. We should not make assumptions about the psychological origin of perceived obligations. Second, we should not assume that these experiences tell us anything about the world. That is, we should not assume that the appearance of an external demand is caused by some *actual* external demand. Phenomenology can mislead. It may be that there is nothing by which to guide our evaluations except our own desires and the expectations of others, and that "morality" is as much a social construct as etiquette.

---

[7] Maurice Mandelbaum, *The Phenomenology of Moral Experience* (Glencoe, IL: The Free Press, 1955), 50.

To avoid begging these questions, we need a term that allows us to remain neutral about both the psychological nature of these moral experiences and the role that these experiences should play in the epistemology of evaluation. The Stoic and Skeptical philosophers of antiquity had a term to describe experiences in this neutral way: φαινοηενα or *phainomena*, which is usually translated as 'appearance'. As Sextus Empiricus explains it, some feelings are "forced upon [us] by appearances," regardless of whether we accept them as accurate reflections of the way the world is.[8] As an example, he admits that "it appears to us that honey sweetens (we concede this inasmuch as we are sweetened in a perceptual way); but whether…it is *actually* sweet is something we [continue to] investigate."[9] Thus, to report an appearance is merely to "say what is apparent to [oneself] and report [one's] own feelings without holding opinions [about the way the external world is], affirming nothing about externals objects."[10] In reporting an appearance, then, we report an experience that is forced upon us, as opposed to one conjured up by an act of thought, but we neither accept nor reject the belief with the propositional content suggested by those experiences.

Borrowing this expression from the Hellenistic philosophers, we could use the term 'moral appearances' to denote the experiences that Mandelbaum describes. To have a moral appearance is to have an experience *as of* something having a certain moral character; someone has a moral appearance when a certain action (or type of action, or character trait, etc.) *seems* or *appears* to that person in a particular way, *morally* speaking. Having (or reporting) a moral appearance is not the same as and does not entail *judging* that something is a particular way, morally speaking. Just as it is possible

---

[8] Sextus Empiricus, *Outlines of Skepticism*, trans. Julia Annas and Jonathan Barnes (Cambridge: Cambridge University Press, 1994), I.13.
[9] Ibid., I.20, emphasis added.
[10] Ibid., I.15.

for a stick to appear bent to someone when it is half-submerged without that person judging that the stick is bent, it is possible for an action (character trait, state of affairs, etc.) to appear morally wrong (virtuous, good, etc.) to someone without that person judging that it is wrong.[11] For instance, it appears to some that the United States should not have firebombed Dresden. That is, upon hearing or reading of the firebombing, they have an experience that they might characterize as "recognizing that the firebombing was morally wrong." But even some of those people to whom the firebombing appears ghastly may withhold moral judgment about it, even if they are as powerless to dispel the moral appearance as all of us are to dispel some optical illusions.

The use of 'appears' (or 'seems') to characterize the relevant experience avoids Mandelbaum's commitment to explaining the phenomenology of morality in terms of "feeling." It is thus neutral between sentimentalist and rationalist views of moral psychology.

Talk of "appearances" also avoids the assumption that such experiences tell us anything about the world. Perhaps some of them are veridical. Perhaps the world sometimes is as it appears to be, morally speaking. But we can talk of moral appearances without taking sides on that issue.

A philosophically adequate characterization of moral appearances requires some care. The most philosophically neutral way to characterize them is by appealing to their intentional content. Intentional content is the part of a thought that can be expressed in a 'that' clause. For instance, the intentional content of the belief that there are mountains on the moon is '*that* there are mountains on the moon.' We can avoid

---

[11] Although I will usually discuss appearances of "actions" being "right" or "wrong," I do not mean to exclude appearances of other moral predicates (e.g., 'good', 'unfair', 'courageous', etc.) applying to other morally evaluable subjects (e.g., state of affairs, character traits, etc.).

assuming positions on contentious issues by using 'moral appearance' to refer to any mental state the intentional content of which, if expressed, would amount to a moral claim. This means that a mental state with the intentional content 'that there are mountains on the moon' would not qualify, but an intentional state with content like 'that parricide is wrong' or 'that a virtuous person usually repays kindness with kindness' would qualify.

Distinguishing moral from nonmoral claims, however, is a difficult task. It may be impossible to give a theoretically neutral account of what it is for a claim to be a moral claim. One test of any definition of 'moral' is that it sort claims in (at least rough) accordance with accepted usage. If I give a definition of 'moral' on which 'Rabbits like carrots' is a "moral" claim but 'Shooting rabbits for fun is wrong' is a "nonmoral" claim, then my definition must be rejected. In other words, the purpose of a definition is to systematize our intuitive categorization of claims as "moral" and "nonmoral." As with most words, our ability to apply these terms to a wide range of cases precedes our ability to define them.

It will therefore suffice, for my purposes, to point to a range of exemplars of moral claims. The boundary between those central cases and nonmoral claims is almost certainly vague. But we need not rely on borderline cases in discussing the nature and epistemological status of moral appearances, so this vagueness should not trouble us. Good examples of moral claims include: 'Torture is wrong', 'It was morally permissible for the United States to waterboard Khalid Sheik Mohammed', 'Stalin was evil', 'It is morally unacceptable that millions of children die of easily preventable diseases', 'It is unfair to force someone to atone for their ancestor's misdeeds', and 'You should not covet your neighbor's ox'. Claims like 'Murder is wrong' or 'Ingratitude is wrong' are also good exemplars of moral claims, since 'murder' and 'ingratitude' have a moral dimension

built into them. An act of killing is not murder if it is excusable, and only when gratitude is owed to someone does failure to show gratitude amount to ingratitude. Thus, there is no uncertainty as to the kind of wrongness—moral or otherwise—that such claims attribute to murder and ingratitude.

Even with this intuitive notion of 'moral' in mind, however, it will not do to define moral appearances as mental states with intentional content that, if expressed, would amount to a moral claim. Since the purpose of this term is to pick out a set of mental states that play a special epistemological role in moral discourse, the term must be restricted to (token) mental states that are not reached by conscious inference.[12] That is, it must be restricted, as a matter of stipulation, to mental states that are reached noninferentially. Thus, we need a theoretical term that refers to all *noninferentially derived mental states whose intentional content, if expressed, would amount to a moral claim*. These mental states are *moral appearances*.

Beyond the requirement that it would amount to a moral claim if expressed, there is no limitation on the intentional content of a moral appearance. Any moral claim might *appear* true or false to someone. That is, for any moral claim *p*, it might appear to someone that *p*. Furthermore, the person having the moral appearance might go on to adopt any propositional attitude, such as belief or doubt, toward the intentional content of the appearance, as long as she is *initially* in a mental state that would be characterized as "recognizing that *p*" if *p* turned out to be the case.

---

[12] Assuming that mental states are individuated in part by their content, one person might come to think that *p* without conscious inference, while another infers *p*. Only the former counts as a moral appearance. Assuming that mental state types are individuated by attitude and content, but not by whether they are reached inferentially, this entails that one token of a particular type may be a moral appearance while another token of that same type is not.

To put less fine a point on it, moral appearances are theoretical cousins of intuitive moral judgments. To make an intuitive moral judgment—that is, a noninferential token judgment about a moral claim—involves having a moral appearance, although the converse need not be true, depending on whether one takes 'judgment' to imply an affirmation, explicit or not, of a certain claim. Talk of moral appearances is virtually interchangeable with talk of intuitive moral judgments, keeping in mind the caveat that one can have a moral appearance without endorsing its intentional content.

5. The role of moral appearances in moral epistemology

Because moral appearances present certain actions (etc.) as *demanded of us*, independently of our desires, they transform the epistemology of evaluation. If we take moral appearances seriously, as most people do, then we cannot merely consult our own desires when thinking about what we should do. We must take the seemingly categorical demands of our moral appearances into account, too. In practice, moral appearances enter into evaluative epistemology in several ways.

In the simplest cases, we evaluate something noninferentially. We either desire it, despise it, or have a moral appearance about it, and we act accordingly. In the nonmoral case, we may watch a movie and find it enjoyable. We evaluate it positively. Moral examples abound, too, though we rarely think about them. A stranger unknowingly drops their wallet and begins to walk away, and it simply appears to us that we should call out to him. We hear of a senseless homicide during a robbery, and it simply appears to us that it was morally wrong. Or we read about Kurt Vonnegut's Paul Lazzaro feeding

a steak filled with metal barbs to a dog and then watching the dog die painfully, and it simply appears to us that Lazzaro is vicious.[13]

At a slight remove from such examples, we sometimes use analogies with simple cases to help us evaluate harder ones. For instance, Judith Jarvis Thomson famously imagines a case in which a woman wakes up one morning to discover that she has been surgically attached to a famous violinist without her consent. The violinist needs to share the woman's kidneys for nine months. It appears morally permissible for the woman to detach herself from the violinist, even if that means the violinist's death. It seems that she is not obligated to endure nine months of discomfort and inconvenience to save the violinist's life (though it would appear heroically kind of her to do so). This intuitively clear case, argues Thomson, is enough like the case of a pregnancy resulting from rape that consistency demands that we allow the woman to abort the pregnancy.[14] This argument rests on the moral appearance that it would be permissible for the woman to detach herself from the violinist.

Moral appearances also play a role in the justification of moral principles. Some moral appearances concern principles directly. For instance, it simply seems that recreational killing is morally wrong. We do not need to derive this from examples or by argument. More commonly, ethicists seek to justify moral principles by prompting moral appearances about several different cases and then advancing a principle that sorts those cases accordingly.

Imagine, for instance, an emergency room doctor confronted with six seriously injured patients. One requires such extensive care that, should she choose to help him, the others will die in the meantime. If she concentrates on the other five, she can save all

---

[13] Kurt Vonnegut, *Slaughterhouse-Five* (New York: Delacorte Press, 1969).
[14] Judith Jarvis Thomson, "A Defense of Abortion," *Philosophy and Public Affairs* 1, no. 1 (1971).

of them, but only at the expense of allowing the first to die. It appears morally permissible, perhaps even obligatory, for her to act in such a way that the one person dies but the other five are saved. This case parallels the famous fictional scenario in which the doctor can save five lives by killing a healthy patient and distributing his organs to five dying patients. But it appears morally wrong, in this case, to act in such a way that one person dies but five others are saved. This, along with similar pairs of cases, is often used to argue that there is a moral difference between killing someone and letting someone die.[15]

More broadly, moral appearances underwrite the moral platitudes that provide the data, such as it is, for moral philosophy. For instance, the platitude that people are not morally responsible for events beyond their control rests on a broad set of moral appearances. In practice, a great deal of moral philosophy consists of attempts to systematize moral appearances and the principles and platitudes that they support. Moral appearances exhibit a philosophically tantalizing mix of systematicity and inconsistency. Philosophers, it seems, cannot resist trying to dispel the unruly appearances to reveal a rational order beneath.

This activity is packaged in various ways. Sometimes, for instance, platitudes or principles are claimed to derive from the "meaning" of moral terms. But these meanings are themselves derived largely from moral appearances, for the moral appearances govern most people's application of moral terms. If it is part of the meaning of "morally responsible," for example, that a person is morally responsible only for events within his

---

[15] See, e.g., Philippa Foot, "Killing and Letting Die," in *Abortion: Legal and Moral Perspectives*, ed. Jay L. Garfield and Patricia Hennessey (Amherst: University of Massachusetts Press, 1984); Judith Jarvis Thomson, "Killing, Letting Die, and the Trolley Problem," *Monist: An International Quarterly of General Philosophical Inquiry* 59, no. 2 (1976); *cf.* James Rachels, "Active and Passive Euthanasia," *New England Journal of Medicine* 292 (1975). Rachels uses a pair of moral appearances to argue that there is no difference between killing someone and allowing them to die.

control, this is only because in cases in which a person cannot control some event, the inappropriateness of punishment and blame is a moral appearance.

But is the activity worth doing? Should we rely on moral appearances in deciding which evaluations to endorse? If so, under what circumstances?

There are various reasons that one might think moral appearances worth using in evaluative epistemology. First, they provide a source of evaluative premises other than one's own desires. They thereby provide a starting point on the "ought" side of the is–ought gap—especially when we want to ask about whether we *should* desire what we actually desire, in which case facts about which things we desire fall on the "is" side of the is–ought gap. Second, we might think that moral appearances connect us with "moral reality." Perhaps they provide us some insight into an otherwise inaccessible realm of facts. (This was the view of much Western philosophy during the Christian era, when moral appearances were seen as the products of *synderesis,* the *affectio iustitiae*, the dictates of conscience, the "Common Notions" of morality, the "moral sense," or some other form of divinely inscribed Law.) Third, even if we doubt the existence of "moral facts," we might insist that in virtue of having and caring about moral appearances, we are "inside the moral system," which allows no exit.[16] Being (stuck) inside the system, we must discern its rules as best we can.

But before joining in the milennia-long effort to understand and systematize our moral appearances, it might be wise to think more carefully about the nature and proper epistemic role of moral appearances.

---

[16] Ronald Dworkin, "Truth and Objectivity: You'd Better Believe It," *Philosophy and Public Affairs* 25, no. 2 (1996); *cf.* Bernard Williams, *Ethics and the Limits of Philosophy* (Cambridge, MA: Harvard University Press, 1985), Ch. 10.

6. Two questions about moral appearances

Much of this dissertation will focus on two interrelated questions: What are moral appearances, psychologically speaking? And what role should they play in the epistemology of evaluation?

Being psychological states, mental appearances are natural phenomena. As such, we can give a scientific account of them. We can develop a psychological theory explaining what they are, how they come to be, and what their effects are. We can explore the origin of particular moral appearances, such as a specific juror's moral appearance that the defendant's actions were morally reprehensible. We can also explore the general mechanisms by which the capacity for moral appearances develops in human beings.

But it is important to keep in mind what can and cannot be learned from empirical investigation. It is up to psychology to tell us about the psychological nature and origin of moral appearances. Discerning the moral and epistemological consequences of psychologists' findings requires philosophical argument. But the empirical question does bear on more philosophical questions in two ways.

First, a scientific account of moral appearances may render some moral epistemologies more or less plausible. It may do this by making general metaethical positions more or less plausible. Consider, for instance, a skeptical view on which moral claims cannot be justified because there are no moral truths for them to capture. Early emotivists, among others, held this view. If moral appearances turn out to arise from the same kinds of brain processes that generate, say, mathematical insights, this view will be less plausible than if moral appearances are the product of emotional reactions. This is not to say that the psychological facts will settle the metaethical or epistemological questions, but only that they will affect the plausibility of certain positions.

Second, and more importantly, attending to the psychological account of moral appearances forces us to avoid such vague abstractions as "moral sensibilities" or "rational insight." Faced with a detailed account of how the brain produces moral appearances, we must judge whether our philosophical notions of moral psychology map onto scientific notions, and whether the psychological processes can deliver what our philosophical theories demand of them. Refusing to fall back on appeals to abstractions, and insisting instead on scientifically assessable psychological claims, may enable us to make progress.

But rather than argue in the abstract about whether moral psychology is relevant to ethical theory, it would be more effective to do some moral psychology and then demonstrate the relevance of specific moral psychological claims. So this line of inquiry must be postponed for later.

The second major question about moral appearances—what role should they play in the epistemology of evaluation?—is harder. Because it depends heavily on the answer to the first question, there is little purpose in discussing it in the abstract here. But it is worth noting that moral appearances have traditionally occupied a central role in evaluative epistemology. Many philosophers have treated them—or in some cases, a select subset of them—as essential to discovering moral truth, which is taken to play an important role in answering the Socratic question about how to live.

In what follows, I will argue that such philosophers were largely mistaken to assign such importance to moral appearances.

2 The psychology of moral appearances

What shocked the police about Kitty Genovese's murder was not its brutality, but the dozens of neighbors who supposedly witnessed the murder and did nothing.[17] One account reports thirty eight of Genovese's "neighbors [standing] at their windows in fascination [without] so much as lifting the phone" while her assailant stabbed her to death.[18] Psychology textbooks frequently use her case to intrigue readers, and with good reason. The tale of Kitty Genovese's neighbors evokes horror and contempt in those who hear it.

What shocks the rock star Bono about the death of the world's poor is not only the conditions in which they die, but the fact that hundreds of millions of people know of their plight and do nothing.[19] More often than not, urgent letters from UNICEF or CARE, asking for small donations to save dying children, end up in the garbage. Though this horrifies some, few feel the same contempt for those who ignore CARE's plea that they do for those who ignored Kitty's cries.

Paired cases like these often carry great weight in moral epistemology. Though structurally similar, they elicit different moral appearances—that is, different noninferential moral reactions. Though ethicists have long puzzled over such cases, psychologists have begun to explore them, too. Psychologists and empirically-minded philosophers have produced a flurry of important publications on the subject in the last eight years. Unlike the ethicists' debate over the appropriateness of these reactions,

---

[17] Martin Gansberg, "37 Who Saw Murder Didn't Call the Police; Apathy at Stabbing of Queens Woman Shocks Inspector," *New York Times*, March 27, 1964.
[18] Elliot Aronson, *The Social Animal*, 5th ed. (New York: Freeman, 1988).
[19] Bono, "Foreword," in *The End of Poverty*, ed. Jeffrey D. Sachs (New York: Penguin, 2005).

however, the psychologists' controversy concerns a purely descriptive question: By what psychological processes do humans make moral appearances? Answering this question constitutes the first step into an inquiry into the nature of moral appearances, and thus a precursor to an inquiry into their proper epistemic role.

Most accounts of the psychological origin of moral appearances fall very roughly into two categories. Marc Hauser calls them "Kantian" and "Humean" models of moral judgment. Hauser has recently added a third kind of model, which he calls "Rawlsian." (These monikers are only suggestive labels. The "Humean" model, for instance, is not *Hume's* model, but only a descriptive model of the formation of moral appearances that has some special affinity with Hume's thought.) Drawing on recent work in psychology, psychopathology, and cognitive neuroscience, I argue that the correct model of moral appearances cannot fit within this tripartite division. Using Hauser's framework to organize the first part of the discussion, I argue that the Humean and the Rawlsian models capture important features of the processes by which we form moral appearances, but that none of the three models accommodates all of the available evidence. I then propose a new descriptive model of moral appearances that better accommodates the available evidence. This model shares important elements with models proposed, respectively, by Jorge Moll and colleagues, by Shaun Nichols, and by Jesse Prinz.[20] I call this model a "Mencian" model of moral psychology, since it draws important inspiration from the thought of the great Confucian philosopher Mencius. Like Hauser's models, my model is a purely descriptive one. It aims to account for the way we do form moral appearances, not for the way that we should form or appraise moral appearances.

---

[20] Jorge Moll, Ricardo de Oliveira-Souza, and Paul J. Eslinger, "Morals and the Human Brain: A Working Model," *NeuroReport* 14, no. 3 (2003); Shaun Nichols, *Sentimental Rules: On the Natural Foundations of Moral Judgment* (New York: Oxford University Press, 2004); Jesse Prinz, *The Emotional Construction of Morals* (New York: Oxford University Press, 2007).

I will postpone discussion of any philosophical implications of this research until later chapters.

2. The State of the Debate in Moral Psychology

During the 1960s and '70s, Lawrence Kohlberg's rationalist, developmental theory dominated moral psychology. According to Kohlberg, children develop through stages of moral reasoning. Kohlberg devised his theory by examining participants' conscious reasoning. At various stages, their reasoning focused on which actions are punished, on convention, on contractarian considerations, and finally, in the morally elite, on abstract moral principles. In the early 1980s, two quite different trends began to unsettle the Kohlbergian paradigm. Carol Gilligan argued that because women focus on relationships rather than rules, Kohlberg's account did not fit half the human race. Robert Zajonc 's work led to a different kind of objection. By showing that affective processes often have powerful effects on "higher" cognitive processes, Zajonc led many psychologists to develop more sentimentalist views of moral judgment.[21] This encouraged psychologists to seek less rationalist explanations of moral appearances, looking behind the veil of verbal reports and interviews on which Kohlberg based his work.

As Marc Hauser puts it, these competing approaches to moral psychology view humans as very different kinds of "creatures."[22] The rationalist paradigm treats us as creatures who reach moral judgments via explicit reasoning. According to the sentimentalist paradigm, we are creatures whose moral appearances issue from intuitive,

---

[21] Jonathan Haidt, "The New Synthesis in Moral Psychology," *Science* (2007).
[22] Marc D. Hauser, *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong* (New York: ecco, 2006).

emotional reactions to the world. Both kinds of creatures are conceivable. But which kind of creature are we?

The answer, according to Hauser, is, "Neither." We are a different kind of creature altogether, one that forms moral appearances based on innate, consciously inaccessible principles like the principles that underlie grammatical judgments.[23] Hauser takes the inspiration for this view from an analogy that John Rawls draws between moral and grammatical knowledge in the first chapter of *A Theory of Justice*, and so he christens this creature a "Rawlsian creature." By comparison, he labels the other creatures—the rationalist and sentimentalist creatures advocated by other psychologists—as "Kantian" and "Humean" creatures, respectively. (Though each "creature" model shares some particular claim with the philosopher that inspired it, the labels do not imply that features of the psychological models can be inferred from the philosophies of their respective namesakes.)

Hauser summarizes the difference between his own model and the other models in terms of the causal processes that issue in moral appearances. If we are Kantian creatures, then our perception of an action prompts conscious reasoning from explicit principles, which leads to a moral judgment. If we are Humean creatures, then our perception of an action triggers an emotional response, which directly causes a moral appearance. If we are Rawlsian creatures, however, then our perception of an action leads to an unconscious analysis of the action in terms of its causes and consequences. This analysis causes a moral appearance via the application of unconscious principles, and that prompts emotional responses and conscious reasoning.

---

[23] Ibid; Marc Hauser, Liane Young, and Fiery Cushman, "Reviving Rawls' Linguistic Analogy: Operative Principles and the Causal Structure of Moral Actions," in *Moral Psychology*, ed. Walter Sinnott-Armstrong (Cambridge, MA: MIT Press, 2008).

These three "creatures"—the Kantian, Humean, and Rawlsian creatures—provide a helpful way of organizing the current literature on the formation of moral judgments. Each model makes specific predictions about human moral psychology, and evidence from behavioral psychology, functional neuroimaging, clinical neuroscience, and psychopathology all bear on the validity of the models. By examining this evidence, we can discover what each model gets right and where each model falls short. Thus, even if we abandon this particular division of theories in the end, they provide a helpful starting point.

Ultimately, we should move beyond Hauser's tripartite division of theories of moral judgment. Though the evidence shows that we are not Kantian creatures, it lends partial support to both the Rawlsian and Humean models. That is, the Kantian creature model makes predictions that are incompatible with the evidence. There is evidence that the Rawlsian model captures some of the important features of our moral psychology, but there is also evidence against it; and much of the evidence against the Rawlsian model favors the Humean model. The Humean model also captures some, but not all, of the important features of our psychology. Thus, Hauser's division of the field, while it helps advance the debate, is not entirely adequate. Only by blurring the distinction between Humean and Rawlsian models can we develop an adequate theory of the generation of moral appearances.

3. Evidence That We Are Not "Kantian Creatures"

The "Kantian" creature, as Hauser characterizes it, forms moral judgments by consciously reasoning from explicitly held principles.[24] The only way to make sense of moral appearances, as noninferential mental states, is as nonconscious inferences from explicitly held principles. Even allowing for nonconscious inference on the Kantian model, if humans really were Kantian creatures, then people would typically be able to report the principles behind their judgments, if not the reasoning that led from their principles to the judgment or (when final judgment is withheld) to the appearance. But evidence from the last two decades or so shows that when people are asked to report the principles underlying their moral judgments, they typically fail. Hence, we are not Kantian creatures.

Humans have a very hard time recognizing the principles that determine their moral judgments. Through his studies of "moral dumbfounding," Jonathan Haidt has shown that the principles people cite are at least sometimes causally inefficacious in their moral judgments. To make matters worse for the Kantian model, Hauser has found that other principles, which people do not cite, are causally efficacious in their moral judgments. Since the Kantian creature model requires us to *consciously* deduce our moral judgments from principles that we *explicitly* endorse, it cannot explain this dissociation between the principles that people cite and the principles that actually underlie their judgments. Hence, it fails as a description of the way that people make moral judgments.

---

[24] The "Kantian creature" model emphasizes the role of reason in moral judgment, and so it is natural to name it after Kant. But it is important to remember that the "Kantian creature" model is not *Kant's* model of moral psychology. Features of the model cannot be inferred from Kant's ethics or psychology. A similar point applies to Hauser's other models, though they do resemble their respective namesakes a bit more closely.

Haidt has come to express his view on the role of moral appearances in moral cognition in terms of the "intuitive primacy principle," according to which moral judgment is primarily controlled by quick, automatic, emotion-driven moral appearances, and "[m]oral reasoning…is usually a post-hoc process in which we search for evidence to support our initial intuitive reaction."[25] Haidt has found that people generally offer quick moral verdicts but have great trouble articulating the principles behind those appearances. In cases where the experimenter demonstrates errors in the participants' post-hoc reasoning, most participants are "morally dumbfounded" rather than persuaded to change their mind. That is, they stand by their original appearance while admitting that they do not know how to justify it.[26] These findings suggest, contrary to the Kantian creature model of moral judgment, that moral appearances are not the product of explicit principles.

Haidt emphasizes that his Social Intuitionist model does not imply that reason has no role in moral thought.[27] He argues that reasoning between individuals can alter moral judgments and that some people are capable of reasoning their way beyond intuition. His model does entail, however, that for most people, most of the time, moral judgments depend on moral appearances, which are not themselves caused by reasoning from conscious principles.

---

[25] Haidt, "The New Synthesis in Moral Psychology."
[26] Jonathan Haidt, "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* 108, no. 4 (2001); Hauser, *Moral Minds*, Hauser, Young, and Cushman, "Reviving Rawls' Linguistic Analogy: Operative Principles and the Causal Structure of Moral Actions."
[27]Haidt, "The Emotional Dog and Its Rational Tail"; Jonathan Haidt, "The Emotional Dog Gets Mistaken for a Possum," *Review of General Psychology* 8 (2004); Jonathan Haidt and Fredrik Bjorklund, "Social Intuitionists Answer Six Questions About Moral Psychology," in *Moral Psychology*, ed. Walter Sinnott-Armstrong (Cambridge, MA: MIT Press, 2008).

Haidt began to develop his Social Intuitionist model after asking people about eating their pet dog. Haidt and his colleagues presented Americans and Brazilians with stories about a family who cooked and ate their pet dog after it was run over by a car; a woman who cleaned toilets with the national flag; a man who had sex with a chicken carcass before cooking and eating it; siblings who kiss "passionately;" and a son who breaks his promise to visit his mother's grave. Many participants judged most of the actions described to be wrong and said that they should be stopped or punished.[28]

But Haidt noticed an interesting pattern in their responses. Most participants would render judgments quickly, only to grope for a supporting reason. Haidt offers this example: "a subject might say, hesitantly, 'It's wrong to eat your dog, because…you might get sick'."[29] Even if the interviewer undermined that reason by reiterating part of the story—for instance, by noting that the family fully cooked the dog—participants rarely changed their judgment. Instead, they often admitted to being unable to justify their reaction, a phenomenon that Haidt dubbed "moral dumbfounding." This led Haidt to hypothesize that the reasons that participants offered were rationalizations of, rather than causes of, their moral appearances.[30] In support of this claim, Haidt has found that participants' affective reactions to the story were better predictors of their moral judgment than were their claims about harmful consequences.[31]

---

[28] Jonathan Haidt, S. Koller, and M. Dias, "Affect, Culture, and Morality, or Is It Wrong to Eat Your Dog?," *Journal of Personality and Social Psychology* 65 (1993).
[29] Haidt and Bjorklund, "Social Intuitionists Answer Six Questions About Moral Psychology."
[30] Haidt, "The Emotional Dog and Its Rational Tail"; Haidt and Bjorklund, "Social Intuitionists Answer Six Questions About Moral Psychology."
[31] Jonathan Haidt, "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* 108, no. 4 (2001). Jonathan Haidt and Matthew A. Hersh, "Sexual Morality: The Cultures and Emotions of Conservatives and Liberals," *Journal of Applied Social Psychology* 31, no. 1 (2001).

Haidt replicated his findings on moral dumbfounding in two laboratory studies.[32] The hypothesis that people typically have an automatic affective reaction to morally salient scenarios has also received direct support from a recent functional neuroimaging study.[33]

Skeptics rightly doubt that Haidt's studies alone show that the intuitive primacy principle is correct, or that the principles people cite are rarely causally efficacious in the formation of moral judgments. Haidt has intentionally focused on disgusting or unusual norm violations, such as sibling incest or eating a deceased pet. Perhaps in more central cases, such as murder or armed robbery, principles do play a bigger role. Testing for moral dumbfounding in such central cases is impossible, however. Testing for moral dumbfounding requires describing an action, soliciting a moral judgment and a justification for that judgment, showing that the justification does not apply to the case as described, and seeing whether the judgment changes. But to test whether, say, the principle that all persons have a right to life is causally efficacious in our moral judgments of killing, the experimenters would need to describe a case in which the victim has no right to life. Such a case would be so far removed from the central case of our condemnation of murder that it would not provide an empirical test of the intuitive primacy principle as it applies to murder. Similar problems arise for testing for moral dumbfounding with respect to other canonical moral judgments.

There are, however, other reasons to think that moral dumbfounding is not a fringe phenomenon, occurring only in bizarre cases like those that Haidt identifies. First, one lesson of Zajonc's "affective revolution" in the 1980s, which Haidt hails as an

---

[32] Ibid; Jonathan Haidt, Fredrik Bjorklund, and Scott Murphy, "Moral Dumbfounding: When Intuition Finds No Reason," unpublished manuscript (Charlottesville, VA: University of Virginia, 2004).
[33] Qian Luo et al., "The Neural Basis of Implicit Moral Attitude—an IAT Study Using Event-Related fMRI," *NeuroImage* 30 (2006).

impetus to the current research in moral psychology, is that more primitive and affective processes exert far more influence over "higher," more "cognitive" processes than we like to admit. Second, as I will argue below, recent neuroimaging research shows that the formation of moral judgments, including judgments about canonical violations like killing and stealing, automatically recruits affective neural circuits. Thus, we know that affect, which can powerfully influence higher-level processes, is intimately involved in core cases of moral cognition.

Haidt's research is also complemented by data from Marc Hauser's Online Moral Sense Test. Hauser's test asks participants to pass judgment on various moral dilemmas and explain their responses. By 2006, over 30,000 participants from 120 countries had pondered various moral dilemmas. For instance, some participants compared a classic moral dilemma, the "trolley problem," to a famous variation of that dilemma, the "footbridge problem." In the classic trolley dilemma, participants judged whether it would be permissible to flip a switch that would redirect a runaway trolley so that it kills one person rather than five. In the "footbridge" variation, participants judged whether it would be permissible to push a large man from a footbridge onto a railway track in order to stop a runaway trolley from killing five people further down the track. This pair of dilemmas appears frequently in psychological tests of moral judgment, and it is well established across a range of cultures that most people judge it permissible to flip the switch in the classic version but impermissible to push the fat man in the footbridge variant.[34]

---

[34] Joshua D. Greene et al., "An fMRI Investigation of Emotional Engagement in Moral Judgment," *Science* 293 (2001a); Greene et al., "The Neural Bases of Cognitive Conflict and Control in Moral Judgment," *Neuron* 44 (2004); Piercarlo Valdesolo and David DeSteno, "Manipulations of Emotional Context Shape Moral Judgment," *Psychological Science* 17, no. 6 (2006); Elisa Ciaramelli et al., "Selective Deficit in Personal Moral Judgment Following Damage to Ventromedial Prefrontal Cortex," *SCAN* 2 (2007);

Analyzing responses from about 5,000 early participants, Hauser and his collaborators found results that parallel Haidt's in important ways. Not only was there fairly strong agreement in participants' moral judgments about each case, but participants exhibited the same logically inept patterns of justification. Hauser et al. sorted participants' justifications into three major categories: "Discounted" justifications were blank or "included added assumptions." "Sufficient" justifications needed only to "correctly [identify] any factual difference between the two scenarios and [claim] the difference to be the basis of moral judgment." "Insufficient" justifications "failed to identify a factual difference between the two scenarios" or admitted to being unable to explain their judgments. After eliminating the discounted justifications from their analysis, Hauser et al. found that 70% of subjects offered insufficient explanations when comparing the classic trolley and footbridge dilemmas. Analysis of another pair of dilemmas found a similar level of insufficient justifications.[35]

Hauser's lab at Harvard continues to collect and analyze data from the Moral Sense Test, which has recently been translated into Chinese, Spanish, and Hebrew. Their research to date supports the finding that people are rarely able to articulate the principles that actually underlie their moral judgments.[36]

If we really were Kantian creatures, then we would expect people to have access to the principles underlying their moral judgments and appearances. Haidt's and Hauser's research shows that this is generally false. Moral appearances play a fairly

---

Michael Koenigs et al., "Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgments," *Nature* 446 (2007); Hauser, Young, and Cushman, "Reviving Rawls' Linguistic Analogy: Operative Principles and the Causal Structure of Moral Actions."

[35] Marc Hauser et al., "A Dissociation between Moral Judgments and Justifications," *Mind & Language* 22, no. 1 (2007).

[36] Fiery Cushman, Liane Young, and Marc Hauser, "The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm," *Psychological Science* 17, no. 12 (2006); Hauser, *Moral Minds*; Hauser, Young, and Cushman, "Reviving Rawls' Linguistic Analogy."

decisive role in determining moral judgment. Furthermore, people cite principles that are causally inefficacious in the formation of their moral appearances, but do not cite principles that are causally efficacious. The best explanation of this double dissociation is that people do not have conscious access to the principles guiding their moral appearances.

These findings do not by themselves establish the relative roles of emotion and reason in moral judgment. The Kantian creature that is ruled out by this research is only one possible rationalist model of moral judgment. Nor does such research directly undermine the claim that we can or ought to reach moral judgments by rational deliberation or insight into rational principles. What it does establish is that for most people, most of the time, moral appearances are caused by something other than rational principles of which they have, or can easily achieve, conscious awareness, and that these appearances tend to dominate moral judgment.

4. Evidence That We Are, At Least Partly, Rawlsian Creatures

Hauser has proposed an alternative to the Kantian creature, which he calls a "Rawlsian creature," and for which he has produced some compelling evidence. This model cannot account for all of the existing data, but it does capture an important aspect of our moral thought by providing a role for principles in the formation of moral appearances.

Hauser takes his inspiration for the model, along with its name, from an analogy that John Rawls draws between moral theory and linguistics. Citing Chomsky's work in linguistics, Rawls suggests that the task of moral theory is like the task of "describing the sense of grammaticalness that we have for the sentences of our native language.... This is a difficult undertaking which...is known to require theoretical constructions that far

outrun the ad hoc precepts of our explicit grammatical knowledge."[37] Thus, just as linguists search speakers' judgments of grammaticality for underlying patterns, Hauser wants to probe people's moral judgments to discover underlying principles.

Hauser extends Rawls' analogy between moral and grammatical judgments even further along Chomskian lines.[38] Chomsky argues that all humans have an innate, universal grammar, which provides a set of principles underlying the language of each human culture; Hauser claims that all humans have an innate "moral grammar," which provides a set of principles underlying the moral code of each culture. Chomsky explains variation between languages in terms of variation along well-defined "parameters," such as whether adjectives precede or follow the nouns they modify; Hauser explains variation between moral codes in terms of variation in parameters, as well, such as whether adultery warrants an exception to the universal (though rarely or never exceptionless) prohibition on killing. The result, for both theorists, is a set of tendencies shared by all psychologically normal humans, which varies systematically between cultures.

With this analogy in mind, there are two features of Hauser's model that are essential to determining whether, or to what extent, humans are Rawlsian creatures. First, Hauser believes that there are principles, shared by all humans, that shape the appearances that people have about moral dilemmas, but that these principles are rarely accessible to consciousness. This differentiates the Rawlsian creature from the Kantian creature because the principles driving the Rawlsian creature's appearances are unconscious. Second, Hauser believes that the human mind applies these principles, and thus forms a moral appearance, prior to generating the emotional reactions or post-

---

[37] John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971).
[38] Hauser, *Moral Minds*; Hauser, Young, and Cushman, "Reviving Rawls' Linguistic Analogy."

hoc chains of reasoning observed by Haidt. This differentiates the Rawlsian creature from the Humean creature, whose emotions are causes, not effects, of its appearances.

The data from Hauser's online moral sense test suggest that people's responses to moral dilemmas do follow identifiable principles, even though people are generally inept at articulating those principles. To "follow" a principle, in the relevant sense, is to respond in a way that is consistent with or would be predicted by the application of that principle. For instance, responses to trolley dilemmas are governed by the "doctrine of double effect," according to which it is morally worse to harm someone as a means to an end than as a foreseen but unintended consequence of pursuing that end. One difference between pushing someone off the footbridge (in the "footbridge" dilemma) and throwing the switch (in the "classic" trolley dilemma) is that the man pushed from the bridge is being used as a means of stopping the trolley, whereas the hapless hiker on the sidetrack dies merely as a side effect of diverting the trolley away from the five people further down the track.[39] The doctrine of double effect entails that pushing someone from the footbridge is impermissible, but throwing the switch is permissible. Most people's judgments are consistent with this principle, even though the majority of people cannot articulate this principle when asked to justify their judgments. This counts against the Kantian model, but in favor of the Rawlsian model, for it shows that although there are principles driving people's moral judgments, at least some of those principles are typically consciously inaccessible.

This argument for the existence of consciously inaccessible principles mirrors the argument for the existence of consciously inaccessible principles of grammar.

---

[39] See §3 and Hauser, *Moral Minds*; Hauser et al., "A Dissociation between Moral Judgments and Justifications."

These findings do not, however, establish the second central feature of the Rawlsian creature, for it does not rule out the possibility that people have the same appearances because they share the same emotional reactions to the situations. That is, it does not rule out the possibility that the processes that lead people to follow certain principles are affective processes. Thus, it does not establish that moral appearances are not effects of emotional reactions. This constitutes the biggest hurdle for Hauser's model, for there is compelling evidence that emotions often play a causal role in the formation of moral appearances—evidence, that is, that humans are, at least partly, Humean creatures.

5. Evidence That We Are, At Least Partly, Humean Creatures

While the Kantian and Rawlsian creatures both rely on principles, conscious or unconscious, to form their moral appearances, the "Humean" creature's moral intuitions are determined by its emotional reactions to situation at hand. Like the Rawlsian creature model, this model cannot account for all of the data, but it does capture something central to human moral psychology.

Speaking presciently for those who would advocate this theory, Hume wrote that the "hypothesis which we embrace is plain. It maintains, that morality is determined by sentiment."[40] The distinctive feature of the Humean creature is that emotion plays a causal role in the formation of its moral judgments, a role it does not play for either the Kantian or Rawlsian creatures. On this model, emotions are causes, not effects, of moral appearances.

Beginning with Haidt's discovery that emotional reactions were good predictors of moral appearances in his studies of moral dumbfounding, a number of researchers,

---

[40] David Hume, *An Enquiry Concerning the Principles of Morals* (1751; repr., Indianapolis, IN: Hackett, 1983).

using a number of different techniques, have provided reasons to believe that moral appearances are sometimes caused by emotions. Several functional neuroimaging studies have captured the brain in the midst of forming moral appearances and found associations between emotional reactions and moral thought. Clinical and behavioral studies show that this association is causal. Finally, research on psychopaths shows significant abnormalities in the content of their moral judgments, not just their moral motivation, and links those abnormalities to abnormalities in their emotional lives. These considerations, from a number of psychological subdisciplines, provide a powerful argument that we are, at least partly, Humean creatures.

5.1. Evidence from Functional Neuroimaging

Preliminary functional neuroimaging studies found a strong connection between emotional reactions and moral judgments. Working with an array of different colleagues in Brazil and the United States, Jorge Moll and Ricardo de Oliveira-Souza have conducted a series of studies exploring the neural correlates of moral judgment.[41] In these studies, participants were presented with sentences or pictures with and without moral content. Moll and de Oliveira-Souza sometimes asked their Brazilian participants

[41] Jorge Moll, Paul J. Eslinger, and Ricardo de Oliveira-Souza, "Frontopolar and Anterior Temporal Cortex Activation in a Moral Judgment Task," *Arq. Neuro-Psiquiatr* 59, no. 3B (2001); Jorge Moll et al., "The Neural Correlates of Moral Sensitivity: A Functional Magentic Resonance Imaging Investigation of Basic and Moral Emotions," *The Journal of Neuroscience* 22, no. 7 (2002); Moll et al., "Functional Networks in Emotional Moral and Nonmoral Social Judgments," *NeuroImage* 16 (2002); Moll et al., "The Moral Affiliations of Disgust: A Functional MRI Study," *Cognitive and Behavioral Neurology* 18, no. 1 (2005).

to assess sentences as “certo” (“right”) or “errado” (“wrong”).[42] Sometimes they only asked participants to view the stimuli attentively.

For example, in the study related in (2001), ten normal adult participants heard sentences like ‘The boy stole his mother’s savings’ and ‘They hung an innocent,’ as well as nonmoral sentences like ‘Walking is good for your health’ and ‘Stones are made of water’. Subjects assessed each sentence as “certo” or “errado.” Afterwards, participants were asked to rate each sentence on its moral content, judgment difficulty, and emotional valence (positive or negative), verifying that they recognized the difference between the moral and nonmoral sentences.

Later studies compensated for potentially confounding effects of the emotional impact of sentences, largely confirming the results of the initial study.

In reviewing their and others’ early studies, Moll, de Oliveira-Souza, and Eslinger summarize the brain areas recruited during moral judgment tasks. They explicitly note that many of the areas that were especially active during moral judgments are implicated in emotional processing. These areas include the medial frontal gyrus, the medial orbitofrontal cortex, the amygdala, and the insula.[43] This fits well with the conclusions of a similar review by Joshua Greene and Jonathan Haidt.[44]

A related study by Luo et al. also found recruitment of the some of these same brain regions in moral judgment tasks.[45] Heekeren et al. report similar findings during

---

[42] The authors note that the Portuguese words ‘certo’ and ‘errado’ “allow moral, factual, and structural connotations,” just as ‘right’ and ‘wrong’ do in English. See Moll, Eslinger, and de Oliveira-Souza, "Frontopolar and Anterior Temporal Cortex Activation in a Moral Judgment Task."

[43] Moll, de Oliveira-Souza, and Eslinger, "Morals and the Human Brain: A Working Model."

[44] Joshua D. Greene and Jonathan Haidt, "How (and Where) Does Moral Judgment Work?" *Trends in Cognitive Sciences* 6, no. 12 (2002).

[45] Luo et al., "The Neural Basis of Implicit Moral Attitude.”

experiments with German-speaking subjects.[46] These three studies focused on the kind of quick, automatic responses that would be involved in the formation of moral appearances.

By 2001, however, Joshua Greene and his colleagues were already beginning to complicate this picture. While Greene's functional neuroimaging work supports the basic Humean tenet that emotions play a causal role in moral judgment, it also reveals that this model only applies to certain kinds of moral judgments. Thus, we are partly Humean creatures, but not entirely.

Greene and his colleagues confronted nine fMRI-ensconced subjects with a battery of moral dilemmas, inspired by trolley dilemmas, along with a number of nonmoral dilemmas.[47] Greene had two independent evaluators divide these dilemmas into "personal" and "impersonal" dilemmas, where a "personal" dilemma involved an action that "(a) could reasonably be expected to lead to serious bodily harm (b) to a particular person or…group of people (c) where this harm is not the result of deflecting an existing threat onto a different party." On these criteria, the classic trolley dilemma, where the protagonist need only throw a switch that redirects the trolley, counts as an impersonal dilemma, whereas the "footbridge" variation, in which the protagonist must push a man to his death, counts as a personal dilemma.

---

[46] Hauke R. Heekeren et al., "An fMRI Study of Simple Ethical Decision-Making," *NeuroReport* 14, no. 9 (2003); Heekeren et al., "Influence of Bodily Harm on Neural Correlates of Semantic and Moral Decision-Making," *NeuroImage* 24 (2005).

[47] Greene et al., "An fMRI Investigation of Emotional Engagement in Moral Judgment." Functional magnetic resonance imaging (fMRI) machines provide a dynamic, three-dimensional image of changes in blood flow within the brain, which correlates with changes in neuronal activity. Thus, fMRIs provide a dynamic picture of which parts of the brain are active during different tasks. (By contrast, normal MRI machines provide a static, three-dimensional image of soft tissue, including the brain. They cannot be used to track neuronal activity.)

Greene et al. asked their participants whether it would be appropriate to act in a certain way if they found themselves facing a particular dilemma. For instance, they showed some of their participants this story:

> Your plane has crashed in the Himalayas. The only survivors are yourself, another man, and a young boy. The three of you travel for days, battling extreme cold and wind. Your only chance at survival is to find your way to a small village on the other side of the mountain, several days away.
>
> The boy has a broken leg and cannot move very quickly. His chances of surviving the journey are essentially zero. Without food, you and the other man will probably die as well. The other man suggests that you sacrifice the boy and eat his remains over the next few days.
>
> Is it appropriate to kill this boy so that you and the other man may survive your journey to safety?[48]

Greene et al. found a significant difference in the patterns of brain activity incited by the moral-personal, moral-impersonal, and nonmoral dilemmas, which were later replicated in a similar experiment with a larger sample.[49] The moral-personal dilemmas provoked strong activation of areas associated with emotional processing, while moral-impersonal dilemmas provoked weaker activation of those areas. Nonmoral dilemmas produced less activation in areas associated with emotion, and actually suppressed activity in some of those areas. The moral-impersonal and nonmoral dilemmas, on the other hand,

---

[48] A list of the dilemmas is available online in Joshua D. Greene et al., "Supplementary Material," *Science*, http://www.sciencemag.org/cgi/content/full/293/5537/2105/DC1.
[49] Greene et al., "The Neural Bases of Cognitive Conflict and Control in Moral Judgment."

generated activity in areas associated with working memory and higher cognition. Thus, emotions do play a causal role in some moral judgments, but not all.

A more recent study reveals more subtle interactions between the content of a moral dilemma and the neurocognitive processes that it activates.[50] Borg and her colleagues presented 24 participants with a range of moral dilemmas involving killing. Some of these dilemmas focused on consequentialist concerns: Different choices would lead to different numbers of lives lost. Others focused on specific deontological questions: Is there a moral difference between killing and letting die? Is there a moral difference between killing someone as a means to an end and killing them merely as a consequence of pursuing some other goal? Some dilemmas combined consequentialist and deontological features: Is it better to let many die than to kill one?

Borg et al. report that although the dilemmas they tested showed a clear difference in the areas activated by the moral dilemmas, as compared to the nonmoral dilemmas, the details varied significantly between the different kinds of moral dilemmas. The main effect of moral versus nonmoral dilemmas in their study confirmed some, though not all, of the central neuroanatomical findings of earlier studies. When considering different kinds of moral dilemmas, however, distinct patterns emerge. Dilemmas involving intentional harm to others provoked strong emotional reactions, as did dilemmas that pit subjects' general preference for inaction over action—that is, for "letting die" over "killing"—against their preference for saving more lives over saving fewer. On the other hand, emotional processing was actively suppressed when subjects considered dilemmas that focused primarily on "killing" versus "letting die." Along with dilemmas focusing primarily on the number of lives saved or lost, action-versus-inaction

---

[50] Jana Schaich Borg et al., "Consequences, Action, and Intention as Factors in Moral Judgments: An fMRI Investigation," *Journal of Cognitive Neuroscience* 18, no. 5 (2006).

moral dilemmas recruited areas associated with higher cognition rather than emotion. Thus, certain kinds of moral dilemmas engage emotional processing, while others do not. We are Humean creatures in some situations, but not in all.

5.2. Evidence from Clinical and Behavioral Studies

Subsequent clinical studies have supported the hypothesis that emotions play a causal role in some, but not all, kinds of moral judgments. These studies show that damage to emotional centers of the brain leads to abnormal responses to these moral dilemmas. Non-brain-damaged people typically judge it permissible to kill one in order to save five in Greene's moral–impersonal dilemmas, like the classic trolley dilemma, but impermissible in the moral–personal dilemmas, such as the footbridge dilemma. However, patients with damage to their orbital and ventromedial frontal cortices, areas associated with emotional processing, are less likely to distinguish between these two kinds of cases. These patients continue to judge the classic trolley dilemma as healthy participants do, but they are much more likely to approve of heaving the fat man over the railing.[51] These brain areas are precisely the areas identified by Greene as crucial to moral judgment. Most authors hypothesize that by eliminating the aversive reaction that healthy participants have to pushing the man, damage to these brain regions condemns (or frees) them to rely exclusively on utilitarian reasoning. Though there is some disagreement about exactly how to interpret these findings,[52] these clinical studies

---

[51] Ciaramelli et al., "Selective Deficit in Personal Moral Judgment Following Damage to Ventromedial Prefrontal Cortex"; Koenigs et al., "Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgments."

[52] J. D. Greene, "Why Are VMPFC Patients More Utilitarian? A Dual-Process Theory of Moral Judgment Explains," *Trends in Neurosciences* 11, no. 8 (2007); Guy Kahane and Nicholas Shackel, "Do Abnormal Responses Show Utilitarian Bias?," *Nature* 452, no. 185 (2008); M. Koenigs, "Koenigs Et Al. Reply," *Nature* 452, no. 7185 (2008); Jorge

provide strong support for the claim that emotion plays a causal role in some, but not all, moral judgments.

Others have produced behavioral evidence that emotions play a causal role in non-brain-damaged humans, as well. This behavioral evidence comes from studies involving the manipulation of emotion. By inducing positive or negative emotions in subjects prior to asking them to make moral judgments, psychologists have succeeded in modulating those judgments. Given the clinical evidence that emotional processing is causally implicated in normal moral judgment, it is reasonable to interpret this as manipulating the input into moral cognition, rather than as some kind of priming effect.

Thalia Wheatley and Jonathan Haidt hypnotized subjects to feel a pang of disgust at one of two words—'take' or 'often'. They had their subjects read moral and nonmoral vignettes, some of which contained the word to which subjects had been hypnotically sensitized. Wheatley and Haidt found that the mere presence of the "disgust word" in the story affected the severity of subjects' moral judgments. Thus, by injecting irrelevant emotion into a scenario, Wheatley and Haidt managed to color subjects' moral evaluation of an action.[53]

Piercarlo Valdesolo and David DeSteno accomplished something similar with a clip from *Saturday Night Live* (*SNL*). After having 79 participants watch a clip from either *SNL* or a documentary, Valdesolo and DeSteno presented them with a series of moral and nonmoral dilemmas, including both the classic trolley dilemma and its footbridge variation. As expected, participants who had watched *SNL* reported more positive affect than those who had watched the documentary; and as expected, given

---

Moll, "Moral Judgments, Emotions and the Utilitarian Brain," *Trends in Neurosciences* 11, no. 8 (2007).

[53] Thalia Wheatley and Jonathan Haidt, "Hypnotic Disgust Makes Moral Judgments More Severe," *Psychological Science* 16, no. 10 (2005).

Greene's findings about the relative involvement of emotion in these two dilemmas, this affected participants' response to the footbridge dilemma but not their response to the classic trolley dilemma. After watching *SNL*, participants were more willing to push the fat man off the bridge than were their documentary-watching counterparts. This is as Valdesolo and DeSteno had predicted. Theorizing that affective states serve as "multiply determined" "momentary informational signals," they predicted that the emotional afterglow of the *SNL* sketch would counteract the negative affect induced by the footbridge dilemma. As in Wheatley and Haidt's study, affective influences that were completely unrelated to the moral features of the dilemma affected participants' moral judgments, suggesting that the formation of particular moral judgment does involve emotional processing.[54]

Two other studies have also found that manipulating affect alters moral judgment.[55]

5.3. Evidence from Psychopathy

Another line of reasoning supporting the Humean contention that emotions play a causal role in moral judgment comes from the study of psychopaths. Psychopaths are not the amoralists they are sometimes made out to be; they are not people who make normal moral judgments but do not care about them. Psychopaths' moral judgments are significantly cognitively different than those of normal humans. The leading theories of psychopathy blame the condition on dysfunction in neural systems that include the

---

[54] Valdesolo and DeSteno, "Manipulations of Emotional Context Shape Moral Judgment."

[55] S. H. Schnall, Jonathan Haidt, and G. L. Clore, "Irrelevant Disgust Makes Moral Judgment More Severe, for Those Who Listen to Their Bodies," (Charlottesville, VA: University of Virginia, 2005); Fredrik Bjorklund and Jonathan Haidt, "Vivid Violations Are Morally Worse" (Lund, Sweden: University of Lund, in preparation).

emotional centers of the brain. On at least one leading theory, the abnormalities in psychopaths' moral judgments are caused by the abnormalities in their emotional lives, suggesting that emotions play a causal role in the development of the very capacity for moral judgment.

Although most psychopaths appear to be able to tell the difference between right and wrong, many researchers believe that when it comes to moral matters, psychopaths "know the words, but not the music."[56] Psychopaths can give normal answers to questions about morality, but moral statements do not mean the same thing in their mouths that they do in others'. The words may roll glibly off their tongues, but they ring hollow in their heads.

The primary abnormality in psychopaths' moral judgments is that they fail to distinguish what psychologists call "moral" rules from what they call "conventional" rules. By the age of three, normal humans judge some rules to be more serious than others and to be less contingent on the rules, conventions, or customs of an authority figure, institution, or culture. They tend to justify rules of the first kind—the serious, non-contingent rules—in terms of the effects of the prohibited action on others, whereas they justify the second in terms of the maintenance of social order, the threat of punishment, or the dictates of authorities. For instance, hitting someone is wrong because it harms them, and it would be wrong even if no authority figure prohibited it; but talking during class is wrong because the teacher needs to maintain social order, and talking would be permissible if the teacher allowed it. Psychologists call rules of the first

---

[56] J. H. Johns and H. C. Quay, "The Effect of Social Reward on Verbal Condition in Psychopathic and Neurotic Military Officers," *Journal of Consulting & Clinical Psychology* 26 (1962); Kent A. Kiehl, "Without Morals: The Cognitive Neuroscience of Criminal Psychopaths," in *Moral Psychology*, ed. Walter Sinnott-Armstrong (Cambridge, MA: MIT Press, 2008).

kind "moral" rules, and those of the second "conventional" rules.[57] As a group, psychopaths distinguish less rigorously between conventional and moral rules, and individual psychopaths are more likely than non-psychopaths to make no moral–conventional distinction at all.[58] They are more likely to see moral rules as depending on an authority who can punish infractions. They are also more likely to justify all rules in the ways that normal adults justify conventional rules, such as by appealing to the threat of punishment. This constitutes a significant cognitive difference in psychopaths' moral judgments, not just a difference in their moral motivation. It is not that they recognize the distinctiveness of moral rules but care nothing for them. They do not really recognize moral rules, *qua* moral rules, at all.

Most theories of psychopathy attribute the condition to dysfunction in all or part of the paralimbic system, a group of neurologically related brain structures that includes most important centers of affective processing.[59] Functional neuroimaging studies implicate many of these structures in moral judgment, and an analysis of early fMRI

---

[57] This is not intended as a philosophical characterization of the difference between moral and nonmoral rules. It is only meant to identify two classes of rules with respect to which normal humans respond differently. This distinction is typically associated with the work of Elliott Turiel. For reviews of the relevant literature, see J. Smetana, "Understanding of Social Rules," in *The Development of Social Cognition: The Child as Psychologist*, ed. M Bennett (New York: Guilford Press, 1993); and M. Tisak, "Domains of Social Reasoning and Beyond," in *Annals of Child Development*, ed. R. Vasta (London: Jessica Kingsley, 1995).

[58] R. J. R. Blair, "A Cognitive Developmental Approach to Morality: Investigating the Psychopath," *Cognition* 57 (1995); R. J. R. Blair et al., "Is the Psychopath 'Morally Insane'?" *Personality and Individual Differences* 19 (1995); Blair et al., "Emotion Attributions in the Psychopath," *Personality and Individual Differences* 19, no. 431-437 (1995); R. J. R. Blair, "Moral Reasoning in the Child with Psychopathic Tendencies," *Personality and Individual Differences* 22 (1997); R. J. R. Blair, J. Monson, and N. Frederickson, "Moral Reasoning and Conduct Problems in Children with Emotional and Behavioral Difficulties," *Personality and Individual Differences* 31 (2001).

[59] Kiehl, "Without Morals: The Cognitive Neuroscience of Criminal Psychopaths."

studies found increased connectivity within much of the paralimbic system during moral judgment tasks.[60]

Blair's Integrated Emotion Systems theory of psychopathy builds on this evidence to provide a neurocognitive explanation for psychopaths' failure to develop normal moral thought.[61] Expanding on his earlier Violence Inhibition Mechanism model,[62] Blair argues that psychopathic tendencies emerge because psychopaths are unable to learn normal negative associations with socially destructive behaviors. As a result, they never develop the same aversion to transgressions of moral norms that normal humans do.

In short, Blair hypothesizes that since normal people experience others' fear and distress as negative stimuli, the fear and distress caused by socially destructive actions condition normal children to react negatively to such behaviors.[63] This eventually results in a moral–conventional distinction, as children learn to avoid behaviors that are "punished" by the victim's (or observer's) distress regardless of context or the presence of explicit rules, but learn to avoid others only because they are punished by authorities. The former type of behaviors comes to be treated as moral transgressions—wrong independently of authority and because of their effects on others. The latter comes to be treated as conventional transgressions—wrong only because authorities prohibit and punish them.

---

[60] Moll, de Oliveira-Souza, and Eslinger, "Morals and the Human Brain: A Working Model."
[61] James Blair et al., "Neuro-Cognitive Systems Involved in Morality," *Philosophical Explorations* 9, no. 1 (2006); R. J. R. Blair, "The Emergence of Psychopathy: Implications for the Neuropsychological Approach to Developmental Disorders," *Cognition* 10 (2006).
[62] Blair, "A Cognitive Developmental Approach to Morality: Investigating the Psychopath."
[63] For a more detailed discussion of the role of conditioning in the development of moral dispositions, see p. 58ff below.

Deaf to the fear and distress of others, psychopaths never learn to make these distinctions. Others' distress is not an aversive stimulus to them, and so they cannot learn to react negatively to actions that cause distress. There is, therefore, no difference to psychopaths between moral and conventional rules, for their only motivation to obey either kind of rule is to avoid punishment.

This model provides good theoretical reason to believe that emotional reactions play a causal role in the development of moral thought, shaping both which actions become morally forbidden and how moral trangressions are treated.

5.4. Summary of the Evidence That We Are Partly Humean Creatures

Three lines of reasoning converge on the conclusion that we are partly Humean creatures. Functional neuroimaging studies show that certain kinds of moral judgments and moral appearances are regularly accompanied by emotional processing. Clinical and behavioral studies support the hypothesis that, when such emotional processing occurs, it plays a causal role in the formation of moral judgment. Leading theories of psychopathy suggest that psychopaths' abnormal moral judgments are caused by abnormal emotional processing, even at the developmental stage. Thus, not only has emotional processing been shown to correlate with some kinds of moral judgments, but eliminating or manipulating emotional processing alters those moral judgments, and serious alteration of emotional processing can profoundly alter a person's cognitive approach to moral judgments. This constitutes a solid argument for the claim that emotions play a causal role in some kinds of moral judgments and appearances.

6. Crossing the Humean Creature with the Rawlsian Creature

That lingering qualification—emotion plays a causal role in *some kinds* of moral judgments—is important. For although the evidence presented above suggests that we are partly Humean creatures, it also reveals that our moral psychology is not exclusively Humean. In fact, some of it suggests a Rawlsian side to our moral psychology.

Several of the functional neuroimaging, clinical, and behavioral studies described above suggest that although certain kinds of moral judgments are partly the products of emotions, others are not. Furthermore, these studies suggest that kinds of moral judgments are distinguished according to their content—that is, according to the principles in play in the situation at hand. For instance, Borg et al. found that when the only principle in play in a moral dilemma is some form of the principle "It is better to (passively) let someone die than to (actively) kill someone," emotions are not engaged; whereas when the principle of double effect is in play, emotions are engaged. If the brain distinguishes moral dilemmas according to these principles before generating emotional reactions, then representations or processes that cause us to follow those principles must be temporally and causally prior to the activation of emotional processes. This is precisely what the Rawlsian model predicts, and so with respect to this subset of moral judgments, we do seem to be Rawlsian creatures.

The Humean and Rawlsian models both capture important elements of our moral psychology, but neither can accommodate all of the evidence. We may need a model of moral judgment that combines the suggestion that our moral judgments follow consciously inaccessible principles with the idea that emotions play an essential causal role in the formation of moral judgments.

7. The Mencian Creature: A New Model of Moral Judgment

I propose a new model of moral judgment—one that recognizes both the causal importance of emotions and the existence of unconscious mental processes that mediate the effects of emotion on moral appearances and moral judgment. Along with intellectual debts to Hauser, the model owes much to theories proposed by Jorge Moll and by Shaun Nichols (Moll, de Oliveira-Souza et al. 2003; Nichols 2004). It shares important elements with Jesse Prinz's "constructive sentimentalist" account of morality.[64] It also draws on one of the central ideas in the thought of the great Chinese philosopher Mencius about the connection between the emotions and moral development. Hence, it might be called a "developmental sentimentalist" theory, or more colorfully, a "Mencian creature" model of moral judgment.[65]

Mencius argued that all people are born with various emotional tendencies that, if properly nurtured, blossom into dispositions to feel specific emotions in response to morally salient situations. The development and exercise of these emotional

---

[64] Prinz, *The Emotional Construction of Morals*.

[65] Despite similarities between Confucian and Aristotelian ethics, Mencius is a better namesake for this model than Aristotle for two reasons. First, Mencius emphasizes the natural development of emotional reactions, whereas Aristotle emphasizes the deliberate training of emotional responses and the development of *phronesis*. Second, to label the theory "Aristotelian" risks having it tied to virtues, which have become a controversial topic in moral psychology. See J. Annas, "Comments on John Doris's Lack of Character," *Philosophy and Phenomenological Research* 71, no. 3 (2005); N. Arpaly, "Comments on Lack of Character by John Doris," *Philosophy and Phenomenological Research* 71, no. 3 (2005); J. M. Doris, *Lack of Character: Personality and Moral Behavior* (Cambridge: Cambridge University Press, 2002); Doris, "Precis of Lack of Character," *Philosophy and Phenomenological Research* 71, no. 3 (2005); Gilbert Harman, "The Nonexistence of Character Traits," *Proceedings of the Aristotelian Society* 100, no. 2 (2000); R. C. Solomon, "What's Character Got to Do with It?," *Philosophy and Phenomenological Research* 71, no. 3 (2005).

dispositions—sentiments, in Prinz's terminology—constitutes proper moral development.[66]

The Mencian creature is characterized by three features:

(1) When the Mencian creature feels a "moral emotion," it tends to form a moral appearance; and *ceteris paribus*, one who has a "moral sentiment," which is a disposition to feel a range of "moral emotions" in relevant circumstances, has a dispositional moral belief.
(2) The Mencian creature develops moral sentiments through a process of conditioning in which others' emotions play a prominent role as unconditioned stimuli.
(3) The cognitive process(es) by which moral sentiments are developed and activated in the Mencian creature receive mental representations of act-tokens as input and then categorize those tokens in accordance with consciously inaccessible rules.

7.1. Moral emotions, moral sentiments, and moral beliefs

Substantial empirical research shows a close connection between emotions and moral beliefs. Behavioral and neuroimaging studies show that certain emotions occur along with moral judgments. Studies that manipulate emotion show that moral judgments can be caused by inducing emotions. Thus, it seems that having one of some set of emotions causes people to form moral appearances.

---

[66] D. C. Lau, "Introduction," in *Mencius* (London: Penguin Books, 1970); Mencius, *Mencius*, trans. D. C. Lau (London: Penguin Books, 1970).

Several further caveats are in order, however. First, not all emotions are "moral emotions," for not all emotions are involved in mental states that we would pretheoretically call moral judgments or moral beliefs. Moll et al. identify the major moral emotions as guilt, shame, embarrassment, pride, indignation/anger, contempt/disgust, pity/compassion, awe/elevation, and gratitude.[67] Alternatively, Prinz lists indignation, "righteous anger," "moral disgust," contempt, guilt, shame, admiration, gratitude, gratification, and dignity, which he groups broadly into other- and self-regarding emotions of approbation and disapprobation.[68] These emotions, or emotions from a similar list, are the emotions that prompt moral judgments. Second, feeling a moral emotion is not always sufficient for the formation of a moral judgment or occurrent moral belief. For instance, J. J. C. Smart is reputed to have admitted that certain acts prescribed by utilitarianism appeared wrong to him, though he consistently rejected those appearances in favor of the utilitarian judgment. Third, although triggering a moral emotion typically prompts a moral appearance, there may be other ways to generate such an appearance. There are certainly other ways, such as explicit reasoning, to cause an occurrent moral belief. This entails that moral judgments, or occurrent moral beliefs, are not constituted by moral emotions.

Occurrent beliefs, however, are less interesting than dispositional beliefs, which are dispositions to form occurrent beliefs. If feeling a moral emotion is typically sufficient for forming an occurrent moral belief, then having a disposition to feel a moral emotion is typically sufficient for having a dispositional moral belief. To have such a disposition (or, more accurately, a set of related dispositions) is to have what Prinz calls a "moral sentiment." A moral sentiment is a physiologically instantiated disposition to

---

[67] Jorge Moll et al., "The Cognitive Neuroscience of Moral Emotions," in *Moral Psychology*, ed. Walter Sinnott-Armstrong (Cambridge, MA: MIT Press, 2008).
[68] Prinz, *The Emotional Construction of Morals*.

feel a range of moral emotions.[69] For instance, someone with a strong dispositional belief that eating meat is morally wrong may be disposed to feel anger or disgust when she sees someone eating meat, contempt for those who approve of eating meat, and guilt about her secret love of bacon. All of these related dispositions constitute a different aspect of the belief that eating meat is wrong, and thus work together to constitute a single moral sentiment.

J. J. C. Smart again provides a counterexample to the strong statement that having a moral sentiment is entirely sufficient for having a dispositional moral belief. Smart may have a moral sentiment such that he feels, say, "righteous anger" at the thought of someone torturing a terrorist's child in order to extract vital information from the terrorist. But if he believed that utilitarianism required the action, he would consistently reject those appearances. He therefore cannot be said to have a dispositional belief that one should not torture the child in such situations, despite his moral sentiment. But cases like this are probably fairly rare.

Whether having a moral sentiment is necessary for having a dispositional moral belief is a harder question, about which I am agnostic. On the one hand, it seems that someone might be able to conclude, through conscious reasoning, that an action is morally wrong without developing a moral sentiment toward it. There are also cases of brain-damaged patients who appear to retain moral beliefs while lacking moral sentiments,[70] though the interpretation of these cases is disputed.[71] Also, if someone

---

[69] Ibid.
[70] A. R. Damasio, "Individuals with Sociopathic Behavior Caused by Frontal Damage Fail to Respond Autonomically to Social Stimuli," *Behavioural Brain Research* 41, no. 2 (1990); A. Roskies, "Are Ethical Judgments Intrinsically Motivational? Lessons From 'Acquired Sociopathy'," *Philosophical Psychology* 16, no. 1 (2003).
[71] Jeanette Kennett and Cordelia Fine, "Internalism and the Evidence from Psychopaths and 'Acquired Sociopaths'," in *Moral Psychology*, ed. Walter Sinnott-Armstrong (Cambridge, MA: MIT Press, 2008).

insisted that, say, donating money to poverty relief is morally obligatory, but never donated money (even when she could afford to do so) and never felt *any* guilt about it, we might question whether she believes donating to be obligatory.[72] Perhaps a more refined version of the theory will provide an answer to this difficult question.

7.2. The development of moral sentiments

Blair's study of psychopaths suggests a possible mechanism by which humans develop moral sentiments. Blair explains psychopaths' abnormal moral judgments as the product of their abnormal emotional development. Because they feel no distress when witnessing others' negative emotions, psychopaths fail to develop the normal negative associations with actions that distress others.

If Blair's theory is correct, it suggests that normal humans develop their moral tendencies through conditioning processes that pair certain kinds of actions with negative emotions. There are two types of conditioning: classical conditioning and operant conditioning. The Mencian creature develops moral sentiments through classical conditioning. Despite its strong association with and frequent use by classical behaviorists to induce arbitrary and unnatural behaviors, conditioning is a natural, adaptive learning mechanism. Thus, to say that moral development occurs by conditioning is not to suggest that there is anything arbitrary or unnatural about morality.

Classical conditioning, or Pavlovian conditioning, involves learning a response to a previously neutral stimulus through a temporal association between the neutral stimulus and a non-neutral stimulus (i.e., a stimulus that already elicits a particular

---

[72] Prinz, *The Emotional Construction of Morals*.

response). The non-neutral stimulus is called the "unconditioned stimulus" (US), and the response it elicits is called the "unconditioned response" (UR). The neutral stimulus is called the "conditioned stimulus" (CS), since it elicits a response (the "conditioned response" or "CR") only on the condition that it has been repeatedly associated with the US. Classical conditioning occurs when an organism repeatedly perceives the CS in relatively close spatio-temporal association with the US. Sometimes the organism then develops a CR that is similar to the UR. For instance, Pavlov's dogs began to salivate upon hearing a bell because they repeatedly heard the bell shortly before receiving food, to which they had an UR of salivating. In other cases, the organism develops a CR that is related to, but different than, the UR. For example, rats learn to freeze in fear upon being placed a container in which they have repeatedly received electric shock. They do not respond to the shocks by freezing; but because they have learned that the container signals an imminent shock, they react to it fearfully.

Though it will take further empirical study to determine exactly how this conditioning process works, some speculative examples may help elucidate the hypothesis. Suppose a young girl witnesses a bully taking her brother's cookie, causing her brother to cry. Her brother's crying may be an aversive (unconditioned) stimulus to her, eliciting an (unconditioned) negative emotional response. By hypothesis, she does not yet have an aversion to taking other people's cookies; but the act of stealing cookies has just occurred in close temporal proximity to a stimulus to which she does have an aversion—namely, her brother's crying. If such incidents are repeated, she will acquire, through conditioning, an aversion to taking people's cookies. Similarly, if a boy repeatedly observes his parents reacting with disgust to the sight of interracial couples (which we may suppose, at least for the sake of argument, is a neutral stimulus), he may develop a conditioned response of disgust to such couples.

Note that the mechanisms at work here are neither mysterious nor peculiar to morality. They involve natural tendencies to empathize with others, to find certain emotions aversive, and to develop aversions to a previously neutral stimulus when it is routinely associated with an aversive stimulus. As Mencius emphasizes, as long as these processes are given the opportunity to develop, they will yield moral sentiments.

7.3. Moral sentiments and the unconscious categorization of actions

Hauser's most significant contribution to moral psychology is the idea that moral judgment follows consciously inaccessible principles. He treats this as an objection to sentimentalist accounts of morality, but by prying more deeply into the physiological nature of moral sentiments, sentimentalist accounts can accommodate Hauser's insight. Specifically, sentimentalists need to consider the "grounds" of moral sentiments. That is, they need to consider the physiological instantiation of dispositions to feel moral emotions.

Whatever else moral sentiments require, they must involve at least a neural connection between a neural representation of a type of action and emotional circuits of the brain that allows activation of the representation of the action to excite the relevant emotional circuits. In order to develop this connection, however, the brain must recognize multiple act tokens, or individual acts, as belonging to the same type or category. Recent research into the function of the human frontal lobe, and particularly into certain areas involved in moral processing, suggest that this may be accomplished through stored neural representations of various action types.[73] If there are (semi-

[73] Jordan Grafman, "The Structured Event Complex and the Human Prefrontal Cortex," in *Principles of Frontal Lobe Function*, ed. Donald T Stuss and Robert T Knight (New York: Oxford University Press, 2000); Jacqueline N. Wood and Jordan Grafman,

)permanent neural representations of specific action types, then moral sentiments may involve a connection between a neural representation of an act *type* and various emotional circuits.

The interesting question, then, is how a perception or thought of an act token gets mapped onto the representation of one act type rather than another. Every act token instantiates indefinitely many types. Hacking into a company's computer system to skim fractions of a cent from each transaction, for example, instantiates several different act types: accumulating wealth, stealing, doing what some characters in a movie tried to do, etc. The brain cannot possibly represent all of the types that this act token instantiates, both because of processing constraints and because the act will undoubtedly instantiate types about which the person is ignorant. So there is presumably some set of heuristics that the brain uses to map act tokens onto act types such that each act token is represented as belonging to some of the types it instantiates, but not represented as belonging to others. Physiologically speaking, perceiving an act token activates neural representations of *some* types that the token instantiates, but *not* representations of *all* types that it instantiates.

This is important because when the conditioning process forges an association between a certain action type and a certain emotion, the association is, at the neurocognitive level, between a *particular representation* (or set of representations) and the relevant emotion. Thus, activating *those representations* triggers the relevant emotion, while representations of similar actions might not. For instance, if failing to help a drowning child activates one set of neural representations but failing to help distant needy strangers activates a different set, then the former representations may

---

"Human Prefrontal Cortex: Processing and Representational Perspectives," *Nature Reviews Neuroscience* 4 (2003).

become associated with a negative emotion while the latter are not. Hence, activating the former representations triggers negative emotions and moral condemnation, whereas triggering the latter representations does not. A negative association with some of the neural states used to represent FAILING TO HELP A DROWNING CHILD may not extend to the state that represents FAILING TO HELP DISTANT NEEDY STRANGERS, even though we can describe both actions so that they fall into the same type.

Hauser's insight about consciously inaccessible principles suggests that the heuristics that categorize actions operate outside of consciousness, both in that we are not aware of them and that we have limited ability to manipulate them through top-down processing. But if we are Mencian creatures, Hauser errs in attributing the strangely principled distinctions of our moral judgments to specifically moral rules. The distinctions arise form an interplay of heuristics for categorizing actions and a conditioning process that links certain types of actions with certain emotional circuits. The "moral grammar" toward which Hauser is working is more like an "action grammar," and the moral significance that attaches to certain kinds of actions comes from emotional conditioning. People appear to follow the doctrine of double effect only because the heuristics for categorizing action distinguish between actions that intentionally harm others and actions that harm others as a foreseen but unintended consequence, and thus we tend to develop different moral sentiments regarding actions of these two types. These heuristics may not rely on that principle *per se*, just as the heuristics for catching a fly ball do not rely on the principles of calculus that could be used to describe their outputs, but the output of the heuristics is at least roughly extensionally equivalent to the output of a process that uses the doctrine of double

defect.[74] How much these heuristics can be controlled by conscious reasoning or altered during development is an important and interesting empirical question.

7.4. Summary of the "Mencian" developmental sentimentalist model

This "Mencian" or developmental sentimentalist model of moral psychology emphasizes two stages in the formation of moral appearances, both of which are controlled by consciously inaccessible heuristics for categorizing actions. During the first, developmental stage, the model holds that people categorize actions using unconscious processes. Some of these types become associated with moral emotions through a conditioning process in which certain actions are paired with unconditioned emotional reinforcers. For instance, actions categorized as HITTING FORCEFULLY may become associated with indignation or anger because of the indignation and anger that they evoke in others. Eventually, this causes the representation of an act of forceful hitting to trigger indignation. In Prinz's terminology, this conditioning process leads to the development of a "moral sentiment"—a disposition, grounded in neural structure, to experience certain moral emotions upon perceiving certain types of actions.[75]

In the second stage of the process—the formation of a particular moral appearance—the representation of an act triggers an associated moral emotion, which in turn triggers a moral appearance. Other processes may intervene to cause a moral appearance in the absence of an appropriate cause, as when the manipulation of emotion infuses a situation with emotion that it would otherwise lack. Since action representations are activated according to unconscious principles that map the

---

[74] Gerd Gigerenzer, "Moral Intuition = Fast and Frugal Heuristics?" in *Moral Psychology*, ed. Walter Sinnott-Armstrong (Cambridge, MA: MIT Press, 2007).
[75] Prinz, *The Emotional Construction of Morals*.

perception of a particular action onto specific representations, whether an action token triggers a certain moral appearance will depend on how those unconscious processes sort it. This gives the effect of reliance on unconscious principles.

This sketch of the Mencian creature leaves many important questions unanswered, and much empirical work needs to be done to support the model. It is also incomplete, as it stands, in that it does not give a complete picture of what it is to have a moral belief. But by incorporating Hauser's insight about unconscious principles into a sentimentalist framework, it may represent a step in the right direction and a promising avenue for further research.

8. Conclusion

Hauser divides descriptive theories of human moral psychology into three kinds, which he represents as three different kinds of "creatures." The Kantian creature reaches moral judgments by reasoning explicitly from consciously held principles. The Rawlsian creature reaches moral appearances and moral judgments by unconsciously applying unconscious principles. The moral appearances of a Humean creature are caused by its emotional reactions. Significant psychological evidence shows that humans generally are not Kantian creatures, but that we have important Rawlsian and Humean elements in us. We do seem to apply unconscious principles to our moral appearances, and our emotions do seem to play a causal role in our moral appearances and judgments.

The "Mencian creature" model, a developmental sentimentalist model that abandons Hauser's division between Rawlsian and Humean creatures, may be a better account of human moral psychology than previous models. Like Hume, the Mencian creature model places great weight on the causal role of emotional reactions in the

formation of moral appearances. By concentrating on the neural representation and categorization of actions, however, it can explain the unconscious principles that appear to drive our moral appearances. This is surely not a final model of moral appearances, but it may be an improvement over its predecessors.

Two caveats bears repeating. These are purely descriptive models. They aim only to capture human moral psychology as it is, not as it should be. Further philosophical argument is needed to derive any normative claims from these models.

Second, the details of this model will almost certainly need revision in light of future empirical work. However, there are three conclusions that should be fairly clear from the evidence presented above. First, for most people, most of the time, moral appearances play a significant role in the formation of moral judgments. Second, emotion plays a powerful causal role in the development and formation of certain kinds of moral appearances. Third, moral appearances follow principles that are typically beyond the conscious access of the people that use them, and these principles are largely determined by heuristics for categorizing actions. It follows from these three claims that if our psychology differed in terms of how it categorizes actions or how the categorization of action is connected to emotional circuits, our moral appearances (and thus moral judgments) would differ, as well. This has important implications for the role of moral appearances in the epistemology of evaluation.

3 Doubting Moral Appearances

A cartoon man glances uneasily over his shoulder on the cover of the *New York Times Magazine* from January 13, 2008. He is watching a mother heft her baby, stroller and all, up the stairs of a New York subway station. A few deft lines in the drawing expose the question on the man's mind: Should he help the woman with her baby? Or should he hop on the train that is about to leave the station? Across the bottom of the drawing is a headline: "What Makes Us Want To Be Good?"

In the accompanying article, Steven Pinker explores the burgeoning field of empirical moral psychology, including many of the studies discussed in the previous chapter. Articles like this have proliferated in the last few years, reflecting a deep journalistic interest in new empirical work on moral psychology. But such articles can also raise a certain amount of skepticism. As Pinker notes, "The attempt to dissect our moral intuitions can look like an attempt to debunk them."[1]

Pinker is not the first to observe that psychological explanations of moral thought might shake our confidence in morality. Decades before the latest wave of empirical research into moral psychology, Gilbert Harman argued that psychological explanations of moral judgments allow us to explain the appearance of morality without supposing that there are any "moral facts." Harman's "argument from explanatory irrelevance," as I call it, is usually read as an argument about the metaphysics of morals. But it can also be construed as an argument about moral epistemology. Read this way, the argument from explanatory irrelevance shows that moral appearances do not provide independent evidence for moral claims, unless those moral claims are seen as nothing more than

[1] Steven Pinker, "The Moral Instinct," *The New York Times Magazine*, January 13, 2008.

expressions or reports of our own attitudes. This creates a serious problem for the epistemology of evaluation.

1. The argument from explanatory irrelevance

The argument from explanatory irrelevance begins from a difference between moral and nonmoral "observations." An "observation," in this context, is just "an opinion that is a direct [i.e., noninferential] result of perception."[2] In our terminology, an observation is an "appearance" in the ancient Skeptical sense of that word. Thus, an observation may or may not be accurate.

To illustrate the difference between moral and nonmoral observations, Harman compares a gang of cat-torching hooligans to a team of particle-spotting physicists. Imagine rounding a corner to see a group of teenagers dousing a cat with gasoline and setting it ablaze. This, presumably, will provoke a negative moral appearance. Such an act would immediately strike most of us as wrong. But what do we actually witness? We see the teenagers; we see the cat; we see the gasoline, its container, the match, and the flame. We see the cat running or leaping or flailing; we hear it howling; we see and hear the teenagers jeering and laughing with excitement. We may feel within us a swelling sense of anger at them and pity for the cat. But it seems that we do not see anything beyond these nonmoral facts. If we "see" the moral wrongness of the teenagers' action, it is not in the same way that we see the flames and hear the cats' cries.

By contrast, imagine physicists watching for particles passing through a cloud chamber. Supercooled vapor in the chamber condenses along the path of speeding subatomic particles, forming visible trails behind them. Based on these trails, well-trained physicists can noninferentially report the passage of particles. (Recall that even

[2] Gilbert Harman, *The Nature of Morality* (New York: Oxford University Press, 1977), 5.

though the physicists need a nonconscious inference to make the report, this does not make the report inferential, since imposing that condition would collapse the difference between inferential and noninferential judgments.) But what do the physicists actually witness? They see a mist-filled black box, in which spidery lines appear abruptly and then vanish. If the physicists "see" the subatomic particles that make the lines, it is not in the same way that they see the vapor and the trails.

These two cases share some important features. They are alike in that both cases require us to "observe" something that is, in some sense, unobservable. They are also alike in that we need to attribute some specific additional mental states to the observers in order to explain the apparent observations. For instance, in order to explain how it came to pass that we judged the hooligans' action to be wrong, we must assume that we had certain perceptions—e.g., of the cat igniting—and certain beliefs. Some of the beliefs we must posit are nonevaluative, such as the belief that burning is painful, that the teenagers knew what would happen when they put match to cat, and that the cat is not a robot designed to mimic feline expressions of pain. But we must also assume that we have some more complicated mental states, including moral dispositions, such as the disposition to judge that it is wrong to inflict pain on domesticated animals just for fun or to form negative moral appearances when we witness someone inflicting pain on domesticated animals just for fun. Psychologists like Haidt, Greene, Moll, and Hauser are beginning to fill in the details of what such a disposition consists in—how the human brain goes from recognizing certain kinds of events in the world to forming a moral appearance.[3] The types of mental states we must attribute to the physicists are much the same. We need to posit particular perceptions on the physicists' part—e.g., of the vapor trails—and certain beliefs—e.g., that the box before them is a well-built cloud chamber,

[3] See Chapter 2.

in which vapor trails indicate the passage of subatomic particles—in order to explain the physicists' observations.

There is also a crucial difference between the two cases. As Harman explains it, the difference is that we "need to make assumptions about certain physical facts to explain the occurrence of the observations [made by the physicists], but [we] do not seem to need to make assumptions about any moral facts to explain the occurrence of the…moral observations."[4] The best explanation for the physicists' observations is that subatomic particles *are* passing through the cloud chamber, leaving vapor trails. If there had been no particles, then the physicists would not have reported seeing them. In the moral case, however, we need not posit anything outside our own minds except for the nonmoral facts of the situation: A group of teenagers has set a cat on fire for fun; we saw this happen; we think that actions like that are wrong or are disposed to have negative moral appearances upon witnessing actions like that; and as a result of these events and mental states, it appears to us that the teenagers' behavior was wrong. In contrast to the physicists' case, in which their observation that a particle passed through the cloud chamber is best explained by supposing that a particle passed through the cloud chamber, we need not suppose that the teenagers' action was wrong in order to explain our observation that is was wrong. The claim that recreational cat-burning is wrong is superfluous or irrelevant to the best explanation of our moral appearances.

Philosophers disagree about what conclusions to draw from this difference between moral and nonmoral observations. Some draw strong conclusions about moral nihilism or the non-existence of "moral facts." But such conclusions may be too strong, and it is unclear what a "moral fact" is anyway. Fortunately, we can settle for a weaker, epistemological conclusion.

---

[4] Harman, *The Nature of Morality*, 6.

The fact that moral claims are superfluous in explaining moral appearances entails that, unless specific conditions are met, moral appearances do not count as evidence for moral claims.[5] This is a specific application of a more general principle: If we do not need to suppose that *p* in order to explain observation *o*, then *o* does not count as evidence for *p*. This is so because if we can explain *o* without supposing that *p*, then we can identify some set of conditions, *c*, such that *c* is independent of *p* and, *ceteris paribus*, when *c* obtains and there is an observer present, *o* occurs. (To say that *c* is "independent of *p*" is to say that *p* is irrelevant to whether *c* obtains.) More succinctly, other things being equal, when *c* obtains, *o* occurs, even if not-*p*. But if an observer would have observed *o* even if not-*p*, then *o* does not count as evidence that *p*. Therefore, if we can explain an observation without supposing that *p*, then that observation does not count as evidence for *p*. And since we can explain the moral appearance that *m*, for any moral claim *m*, without supposing that *m*, moral appearances do not count as evidence for any moral claim.

Thus, even without any confusing metaphysical talk of "moral facts," we can see that the argument from explanatory irrelevance, if sound, has important implications for ethical theory. There are, however, several ways to resist the argument from explanatory irrelevance.

2. "Cornell realism" and the explanatory relevance of moral appearances

Among the best-known responses to Harman's argument is Nicholas Sturgeon's. Sturgeon, one of the so-called "Cornell realists," insists that, *pace* Harman, our best explanations of moral appearances *do* require us to endorse "moral facts." As he freely

[5] I will address these "specific conditions" in §5. Until then, I will drop the qualification 'unless specific conditions are met' for ease of exposition.

admits, however, his response works only if we begin by endorsing some moral appearances; hence, it will not suffice to explain why we should take moral appearances as evidence for moral claims in the first place.

Sturgeon's strategy depends crucially on the Cornell realists' metaphysics of morals, according to which moral facts supervene on nonmoral facts. This means (roughly) that if an action has moral property ϕ, then it necessarily has some nonmoral property, ψ, and it could not lack ϕ without lacking ψ. The notion of supervenience is sometimes explained by saying that when one property supervenes on another, there can be no change in the former property without a change in the latter.

This metaphysical point threatens Harman's argument because it undermines the claim that, in any particular case, we might have the moral appearances that we do even if the situation were morally different. Consider Sturgeon's example: Hitler did what he did because he was morally depraved; if he had not been morally depraved, he would not have done what he did, and so would not appear depraved to us. Proponents of Harman's argument would substitute 'had character traits that we judge to be morally depraved' for 'morally depraved', claiming that he would still appear morally depraved, so long as he had those traits, even if there were nothing morally wrong with them. Given Sturgeon's claim about supervenience, however, having the character traits upon which moral depravity supervenes *just is* being morally depraved. That is, it is metaphysically impossible for Hitler to be as he was and yet not be morally depraved. Because moral properties supervene on nonmoral properties, it is impossible for the moral properties to change (e.g., for Hitler to change from depraved to not depraved) without a change in underlying, nonmoral properties (e.g., Hitler's character traits). Thus, if Hitler were *not* depraved, he would not have the traits for which we judge him to be depraved, and so we would not have the moral appearance that Hitler is depraved.

Pace Harman, Hitler's depravity *is* an important part of the explanation of our moral appearances.[6]

There is much of value in Sturgeon's argument, and there is, of course, much to say about it. The overarching problem with the argument, however, is that it operates at the wrong scale. The question is not whether a psychological account of moral appearances undermines any *particular* appearance, but whether it undermines them *as a whole*. The most interesting question in moral psychology is not why we have this or that appearance, but why we have *any* moral appearances at all. In this sense, moral appearances do differ significantly from nonmoral perceptions. The best explanation for our having *any* perceptions of the external world at all is that the external world exists. (There are alternative, skeptical explanations, of course, but they do not fare as well on some important criteria, especially pragmatic criteria, as the realist explanations.) By contrast, we can explain the fact that we have any moral appearances at all without endorsing any of those appearances.

Thus, if the question is why moral appearances *ever* count as evidence for moral claims, rather than whether *particular* moral appearances count as evidence for *specific* moral claims, the Cornell realist response will not suffice. If, *pace* Sturgeon, we do not assume that moral facts exist, moral appearances give us no independent reason to believe in them.

3. Rationalism about ethics as a response to Harman

Some philosophers, while happy to grant that moral appearances are not like observations of the physical world, deny that this has any bearing on moral epistemology.

[6] Nicholas Sturgeon, "Moral Explanations," in *Morality, Reason, and Truth: New Essays on the Foundations of Ethics*, ed. David Copp and David Zimmerman (Totowa, NJ: Rowman & Allanheld, 1985).

Perhaps the most prominent version of this response is "rationalism," understood broadly as the view that moral claims are justified through the use of reason. According to rationalists, the explanatory irrelevance of moral claims is beside the point, for (at least some) moral claims are justified in the same way that mathematical or logical claims are justified—which is to say that they are not justified by observation. Harman's argument is therefore irrelevant to moral epistemology.

There are two ways to rebut this rationalist response. The first, which Harman uses, is to note that mathematical and logical truths do bear on observations in a way that "moral facts" never do, and in this respect moral claims cannot be happily assimilated to mathematical or logical claims. The second is that moral truths cannot plausibly be justified in the way that mathematical and logical truths are justified. Thus, even if moral claims were like math and logic with respect to observation, there is a defense of those fields that is unavailable for morality.

Mathematical and logical claims bear on observations because many observations are best explained by invoking math or logic. Scientists routinely use math and logic to explain why they observe the physical phenomena that they do. Just as claims about "dark matter" or "dark energy" are justified by appealing to their role in explaining observations about things other than dark matter or dark energy, claims about math or logic can be justified by appealing to their role in explaining observations about things other than math or logic.[7] Thus, math and logic are not wholly cut off from observation, even though we can explain our judgments about abstract math and logic without supposing that any of our mathematical or logical beliefs are correct. So, Harman's argument implies different conclusions about math and logic than it does about morality.

---

[7] Harman, *The Nature of Morality*, 9.

Still, rationalists will insist that we have means other than observation for justifying mathematical and logical claims, and that these methods will work for moral truths, too. After all, we can justify the claim that four million divided by two is two million through rational reflection. We do not need to observe a set of exactly four million things being divided in half. A long tradition in ethics holds that we can also establish moral truths using rational reflection. The claim that we can justify moral claims through rational reflection is a more formidable response to Harman's argument. Because it requires a much more detailed rebuttal, I will postpone discussion of it until the next chapter.

4. Sensibility theory as a response to Harman's argument

The rationalists' general strategy is to find "companions in guilt" with which to protect morality from Harman's argument. "Mathematical facts," it was suggested, were as irrelevant as "moral facts" in explaining our observations, and yet we have alternative means of justifying mathematical claims. "Sensibility theorists," such as John McDowell and David Wiggins, suggest a different kind of property as a "companion in guilt"—namely, response-dependent properties like *being red* or *being funny*.

Response-dependent properties are properties that consist in (and thus depend for their existence on) a disposition to elicit certain responses in some (group of) perceiver(s). Consider, for instance, what Crispin Wright calls the "widely accepted and familiar basic equation for redness, a canonical response-dependent property:

> (RED) *x* is red ↔ for any S: if S were perceptually normal and were to encounter *x* in perceptually normal conditions, S would experience *x* as red.[8]

The idea behind (RED) is that what it means for something to be red is just that it elicits experiences of red in us. We can only determine which wavelengths of electromagnetic radiation count as "red" light—and, thus, which microphysical surface structures count as "red"—by determining which wavelengths of light elicit experiences of red in us. Wright includes qualifications about "perceptually normal" observers and "perceptually normal conditions" because an object is red even if it does not elicit experiences of redness in colorblind observers or under colored lights.

Explaining moral properties by direct analogy to colors produces the following schema:

> (MORAL) *x* has moral property M ↔ for any S: if S is normal and S were to encounter *x* in normal conditions, S would have a moral appearance with the intentional content '*x* is M'.

This schema, however, creates a dilemma for sensibility theorists. The schema's insistence on "normal" perceivers and "normal conditions" could be interpreted in a normative or a statistical sense. The difference between a normative and statistical interpretation of 'normal' comes out, for instance, in imagining the last sane man in a world gone mad. If he were to call himself "the last normal man on Earth," he would be wrong, in a statistical sense. If everyone else were insane, he would be wildly *abnormal*

[8] Crispin Wright, "Moral Values, Projection and Secondary Qualities," *Proceedings of the Aristotelian Society* Supplemental Vol. LXII (1988): 14.

in a statistical sense. But his claim is best understood in a normative sense: He is as everyone else *should* be.

If 'normal' is interpreted along statistical lines in (MORAL), then morality becomes a purely descriptive, democratic matter: Whatever most people judge to be right is right—and right *because* most people judge it to be so. This reading of 'normal' therefore seems to abandon the normativity of morality. Furthermore, it leaves no grounds for criticizing the majority opinion, since it is logically impossible, on this reading, for the majority opinion to be incorrect.

However, interpreting 'normal' in a normative way invites a theoretical problem of its own. On a normative interpretation of 'normal', a "normal" perceiver will be one who responds to morally salient stimuli *correctly* in (statistically) normal circumstances. (Alternatively, the normativity could be embedded in "normal circumstances": "normal circumstances" could be defined as those in which a statistically normal perceiver responds correctly.) Identifying normal perceivers (or circumstances), on this definition, requires a prior standard of correctness. In other words, we would need to know which moral appearances are correct in order to determine which perceivers are normal, which is required in order to know which moral appearances are correct. This is hopelessly circular.[9]

Most sensibility theorists prefer the second horn of the dilemma, since the first drains morality of its normativity. They typically try to avoid the need to identify "normal" perceivers and circumstances by expressing the response-dependence schema for moral properties as follows:

---

[9] Simon Blackburn, *Ruling Passions* (New York: Oxford University Press, 1998), 104–11.

> (MERIT) *x* has moral property M ↔ *x* merits, or makes appropriate, a moral appearance with the intentional content '*x* is M'.

However, as more than a few critics have noted, it is far from clear how to cash out the notion of "meriting" or "making appropriate" a moral appearance without presupposing some prior standard of morality. (MERIT) seems just as circular as (MORAL).

Sensibility theorists think they have a defense against the accusation that (MERIT) is circular. This defense relies on the response-dependence of humor, rather than the response-dependence of color. Consider this joke:

> (MUFFINS) Two muffins are baking in an oven. The first looks over to the second and says, "It's really hot in here." The second answers, "Whoa, a talking muffin!"

This joke is not funny. But told in the right context, even this joke can make people laugh.[10] That is, the joke does not merit amusement, but it can sometimes amuse people nonetheless. Furthermore, we can discriminate between jokes that merit amusement and those that do not in a way that at least seems independent of their *actual* tendency to elicit amusement. For instance, even if a joke makes someone laugh when he is in a particularly jovial mood, he might insist that it is "not really funny."

Sensibility theorists typically understand (MERIT), as stated above, as akin to claims like "Something is funny if it merits amusement" or "Something is frightening if it

---

[10] Tyler F. Stillman, Roy F. Baumeister, and C. Nathan DeWall, "What's So Funny About Not Having Money? The Effects of Power on Laughter," *Personality and Social Psychology Bulletin* 33, no. 11 (2007).

merits fear."[11] They insist that something is frightening only if it *merits* fear—only if it makes that fear *appropriate*. A cowardly person is one who feels fear inappropriately. Likewise, a morally corrupt person is one whose moral responses do not "fit" the actions or objects that elicit them—one whose moral appearances are not merited by the actions or objects they are about. By making room for such a person, sensibility theorists seem to introduce a possibility for serious criticism of moral appearances. The question, however, is whether we can extend the analogies with amusement and fear in order to make sense of the idea that some things "merit" certain moral appearances while others do not—that some peoples moral appearances do not "fit" their objects.

In the case of amusement and fear, we can make some sense of these claims. Fear is easier to explain. Some things, like hungry tigers, merit fear because they are dangerous, and we can explain danger in terms of probable harm. Other things, like stuffed tigers, do not merit fear because they are not dangerous at all. There are interesting intermediate cases, too, such as tarantulas. In general, tarantulas are no more dangerous to humans than bees are, but they do seem, in some sense, to be more genuinely frightening than bees.

Amusement is somewhat more difficult. On the one hand, we can distinguish between meriting amusement and causing amusement. Jokes like (MUFFINS) do not merit amusement, even if they sometimes amuse. Likewise, when we hear jokes that we have heard many times before, we may feel no amusement, even though we think that they merit amusement. However, amusement differs from fear in that it does not seem to be a response to some independently specifiable feature of the world, as fear is to danger. Amusement is a response to humor; but humor is simply the property of

---

[11] John McDowell, "Values and Secondary Qualities," in *Mind, Value, and Reality* (Cambridge: Harvard University Press, 1985/1998); David Wiggins, *Needs, Values, Truth: Essays in the Philosophy of Value* (Oxford: Clarendon Press, 1998).

eliciting amusement.[12] Thus, although we seem to be able to make sense of the distinction between causing and meriting amusement, it is not at all clear what the distinction amounts to.

Furthermore, we seem not to gain very much by explaining why something is funny by saying that it merits amusement, although the explanation is not *entirely* trivial. Such an explanation does rule out certain alternatives, even if no one finds those alternatives particularly appealing in the case of humor. For instance, it rules out the explanation that being funny consists in participating in the Form of the Funny or conforming to a set of Comic Rules inscribed in the nature of the universe. (If these "explanations" sound ridiculous, compare them to those sometimes given to explain moral properties.) And if it is correct, then, as David Wiggins points out, no account of the comic is complete without it.[13] But it is epistemologically useless in identifying funny things, for we cannot know what merits amusement independently of knowing what is funny.

Critics of sensibility theory have long complained that the notion of "meriting" a particular moral response, or making a particular moral response "appropriate," is even more mysterious than the notion of meriting amusement. It is especially unclear how we could know that something merits a moral appearance with the intentional content '*x* is M' independently of knowing that *x* is M. This is not just because the two are linked by a biconditional. A triangle is equilateral if and only if it is equiangular, but we can learn that a triangle is equilateral without knowing that it is equiangular, if we are unaware that all equilateral triangles are equiangular. The problem is that, while we can offer separate criteria for being equilateral and being equiangular, there do not seem to be any separate

---

[12] Wiggins, *Needs, Values, Truth: Essays in the Philosophy of Value*; D'Arms and Jacobson, "Sensibility Theory and Projectivism," in *Oxford Handbook of Ethical Theory*, ed. David Copp (Oxford: Oxford University Press, 2006).
[13] Wiggins, *Needs, Values, Truth: Essays in the Philosophy of Value.*

criteria for saying that *x* merits the appearance that *x* is M as opposed to saying that *x* is M. Thus, (MERIT) is epistemologically useless. There are no conditions in which we can learn that *x* is M by first learning that *x* merits a certain moral response, for any circumstance in which we know that *x* merits a certain response will be one in which we *already* know that *x* is M.

This has two important consequences. First, it means that (MERIT) is epistemically useless. By the time we are in a position to apply (MERIT), we already know which things are morally good, morally wrong, etc. But more importantly, the circularity of (MERIT) implies that sensibility theory cannot give us an *independent* reason to accept any moral claims. In order to know that anything merits a particular moral appearance, we would already need to know that some moral appearances are correct. But in order to know that, we would need to know some moral claims. Thus, sensibility theory, as it is typically conceived, cannot save moral appearances from Harman's argument.

5. How moral appearances could be evidence for moral claims

The Cornell realists', the rationalists', and the sensibility theorists' responses to Harman's argument constitute the most important responses in the philosophical literature. If they fail to undermine the conclusion that moral appearances do not provide evidence for moral claims, we have good reason to accept that conclusion.

However, the original statement of the claim that moral appearances do not provide evidence for moral claims came with a caveat. There are several different conditions under which we could count moral appearances as evidence for moral claims.

First, there is an exception to the general principle that if we can explain an observation *o* without supposing that *p*, then *o* is not evidence for *p*. If *p* can be inferred or constructed out of statements for which *o* is evidence, then *o* can provide evidence for

p. Harman uses statements about the "average American" to illustrate this point. We never need to appeal to the qualities of the "average American" to explain any observation we make. (We may need to appeal to people's *beliefs* about the average American, as when we explain someone's opinion of America in terms of their beliefs about the average American, but this is different than appealing to statistical facts about the average American.) Yet, our evidence for claims about the average American rests on observations of particular Americans. Each of those observations, therefore, provides some tiny amount of evidence for claims about the average American.[14]

If, therefore, we could construct moral claims out of claims supported by moral appearances, then moral appearances could provide evidence for moral claims. However, it seems that the only claims supported by moral appearances are claims about our moral sensibilities. Thus, moral appearances would count as evidence for moral claims only if moral claims could be inferred from claims about our moral sensibilities. As we saw in discussing sensibility theory, however, that project faces serious challenges.

One way to meet those challenges, however, and thus one case in which moral appearances would be evidence for moral claims, is to construe moral properties as relational properties. To say that some property *F* is a relational property is, in this context, to say that nothing is ever *F simpliciter*, but is always *F-in-relation-to-x*, for some *x*. In this case, we would say that an action, character trait, etc. has moral property *M-in-relation-to-moral-sensibility-S*, which is just to say that someone with moral sensibility *S* will experience the action (etc.) as being *M*. Since moral claims would always be relativized to a particular moral sensibility, moral appearances would provide evidence for moral claims.

---

[14] Harman, *The Nature of Morality*, 13.

This way of getting moral appearances to count as evidence for moral claims would involve a serious change in the way most of us think about morality, however. It would make morality much more like etiquette than we normally think it is. Just as nothing is rude *simpliciter*, but only rude-in-such-and-such-a-society or rude-in-such-and-such-a-context, so nothing would be morally wrong *simpliciter*, but only wrong-according-to-such-and-such-a-moral-sensibility. This strategy therefore amounts to embracing the first horn of the sensibility theorists' dilemma, robbing morality of its normative significance.

If, therefore, we count moral appearances as evidence for moral claims because we construe moral properties as purely relational properties, morality ceases to be normative. Morality will not impose external demands on us, as so many people seem to think that it does. The epistemology of evaluation would then be quite different than moral epistemology, for moral claims would cease to be evaluative claims and evaluative claims could not, in general, be justified by reference to moral claims. The only case in which they could be justified with moral claims is when the emotional pull of a moral appearance is counted as a reason for or against some evaluation; but in this case, morality plays no more significant a role in evaluative thought than do any of our other emotional dispositions. Just as we often tell ourselves to set aside our irrational fears in order to achieve some desirable goal, we would often tell ourselves to set aside our nagging moral appearances in order to achieve some good but "immoral" end.

There may be nothing inconsistent about this view, but few will adopt it.

There is a different kind of condition under which moral appearances could provide evidence for moral claims. If we could justify moral claims without relying on moral appearances, and then independently verify that someone's moral appearances reliably tracked justified moral claims, then we could count that persons' moral

appearances as evidence for moral claims. By analogy, we can determine what wavelengths of light a surface reflects using a spectrometer. This provides a reason, independent of color perception, to think that a given surface reflects specific wavelengths of light. If we know that a particular person reliably sees blue when exposed light of that wavelength, we could use that person's color sensations as evidence that a given surface does or does not reflect those wavelengths of light.

The challenge, of course, is finding a way to justify moral claims that is independent of moral appearances. If we want to construe moral claims as evaluative claims, then we will need a way to justify evaluations that is independent of moral appearances. I will turn to this project in Chapter Five, after fulfilling my earlier promise to consider whether moral claims might be justify through rational reflection.

4 Moral Appearances as Rational Intuitions

Samuel Clarke had a name for people who doubt moral appearances: "the most abandoned of all mankind." Clarke believed that we apprehend basic moral claims through the use of reason, just as we apprehend basic truths of arithmetic and geometry. Many contemporary philosophers share a similarly rationalist view of moral appearances, though most disavow Clarke's extreme version.

Rationalist views of moral appearances provide one of the most important responses to Harman's "argument from explanatory irrelevance." Harman's challenge for moral epistemology begins from a difference between moral and nonmoral "observations." If we observe protons or parrots, for instance, we can best explain our "proton appearances" or "parrot appearances" by supposing that there are protons or parrots. We can explain our moral appearances, however, without endorsing their intentional content. We need not think that torture is wrong to explain how we come to have the moral appearance that torture is wrong. Because moral claims are irrelevant in explaining moral appearances, moral appearances do not count as evidence for moral claims.

Rationalism offers one of the most important responses to Harman's challenge. Rationalists object that moral appearances should be compared to noninferential judgments about mathematical or logical claims, not to observations of the external world. We can justify claims like 'two plus three equals five' and the law of non-contradiction without appealing to observations. Furthermore, mature, educated humans seem to be able to grasp these "truths of reason" noninferentially. Thus, if some moral appearances were like our noninferential apprehension of mathematical truths,

those appearances could serve as premises by which to justify other moral evaluations, just as mathematical or logical axioms justify more complicated theorems.

The analogy between moral appearances and mathematical or logical intuitions has a long, turbulent history. It forms the basis of rational intuitionism, the best developed rationalist approach to moral epistemology. Distrust of the analogy between moral appearances and mathematical or logical axioms has generated criticism of rational intuitionism since the eighteenth century; but recent intuitionists, such as Robert Audi, Philip Stratton-Lake, and others, have argued forcefully that this criticism rests on a faulty understanding of rational intuitionism. Even with their new, sophisticated forms of rational intuitionism, however, the analogy between morality and mathematics breaks down.

The central tenet of rational intuitionism—that we can know certain moral principles because they are self-evident—breaks down along with the analogy between morality and mathematics. Contemporary intuitionists offer subtle accounts of self-evidence; but how it is that we could recognize any moral claims as being self-evident, in their sense of the term, is psychologically mysterious. Furthermore, their notion of "soft" self-evidence commits rational intuitionists to what might be called the "Thermometer View" of moral appearances—though this sophisticated view should not be confused with the simpler epistemologies of earlier intuitionists. Recent epistemological arguments about "peer disagreement" show that, because of this Thermometer View, rational intuitionists' justification for thinking that they have identified a self-evident moral principle is constantly defeated by peer disagreement, and that their intuitionist justification for endorsing any given moral principle is defeated with it. Furthermore, peer disagreements flow naturally from the central thesis of rational intuitionism rather

than from parts of the theory that could easily be abandoned. Thus, rational intuitionism cannot justify the use of moral appearances in moral epistemology.

1. The revival of rational intuitionism and the softening of self-evidence

After two centuries as one of the leading theories in moral epistemology, rational intuitionism suffered a sharp decline in the middle of the twentieth century. Shortly after W. D. Ross articulated his canonical version of the theory in 1930, the theory plunged into disrepute. There were several reasons for this: Critics conflated the more plausible versions of the theory, like Ross's, with Moorean non-naturalism, which seems to require an occult faculty of moral intuition. Quine's attack on the analytic–synthetic distinction cast *a priori* knowledge into disrepute. The rise of emotivism in the 1930s pushed cognitivist theories out of favor by offering an alternative explanation for the data about moral judgment and discourse—one that many philosophers found superior to the rationalist account. But emotivism's ability to explain the nuances of moral thought and talk is thought to be unraveling, cognitivism is again on the rise, Quine's influence has diminished, and defenders of intuitionism have carefully distinguished the traditional version of the theory from Moore's.

This change of philosophical climate has reawakened intuitionism in general, and a number of philosophers have picked up where Ross left off to champion a new, more sophisticated form of rational intuitionism. Despite numerous additions and emendations intended to deal with common objections, the core of the new intuitionists' theories remains the same: There are self-evident moral principles, and we can know them through rational reflection.[1]

[1] There are other rationalist varieties of intuitionism, of course, such as Jonathan Dancy's particularism, as well as non-rationalist varieties of intuitionism. For the purposes of

Of the various philosophers advocating a return to Rossian intuitionism, the most prominent are Robert Audi and Philip Stratton-Lake. Audi offers the most sophisticated version of the theory to date, and much of the work on rational intuitionism as a moral epistemology draws on his theory of self-evidence.[2] Stratton-Lake, in addition to editing a new collection of essays on the topic, has lucidly and concisely presented the case for renewed interest in the theory.[3]

As Stratton-Lake argues, the twentieth century opposition to rational intuitionism focused largely on its claim that we can know self-evident moral principles, but much of it depended on an old understanding of intuition that fell from favor with intuitionists by the late nineteenth century, if not earlier. This outmoded understanding, which Roger Crisp calls the "Hotline View,"[4] is clearly articulated in Samuel Clarke's 1705 Boyle Lectures:

> These [moral principles] are so notoriously plain and self-evident, that nothing but the extremest stupidity of mind, corruption of manners, or perverseness of spirit, can possibly make any man entertain the least doubt concerning them. For a man endued with reason, to deny the truth of these things; is the very same thing, as if a man that has the use of his sight, should at the same time that he beholds the sun, deny that there is any such thing as light in the world; or as if a man that understands geometry or arithmetic, should deny the most obvious and known

---

this discussion, 'rational intuitionism' refers specifically to the epistemological view that there are self-evident moral principles. Other kinds of intuitionism are beyond the scope of this discussion.

[2] Robert Audi, *The Good in the Right: A Theory of Intuition and Intrinsic Value* (Princeton, NJ: Princeton University Press, 2004); Audi, "Self-Evidence," *Philosophical Perspectives* 13 (1999).

[3] Philip Stratton-Lake, ed. *Ethical Intuitionism: Re-Evaluations* (Oxford: Oxford University Press, 2002). See especially Stratton-Lake's introduction in that volume.

[4] Roger Crisp, "Sidgwick and the Boundaries of Intuitionism," in *Intuitionism: Re-Evaluations*, ed. Philip Stratton-Lake (Oxford: Oxford University Press, 2002), 60.

proportions of lines or numbers.[5]

Four features of this view cause trouble for intuitionism. First, it treats self-evident moral principles as so obvious that any "man endued with reason" will accept them as soon as he understands them. Second, it dismisses the possibility of disagreement over self-evident moral principles, except among those suffering from "the extremest stupidity of mind" or moral corruption. Third, it suggests that anyone who recognizes the truth of a self-evident proposition will also recognize that it is self-evident. Finally, the view is deeply dogmatic. These four features give rise to two of the oldest objections to rational intuitionism—namely, that disagreement about moral principles shows that they are not self-evident and that intuitionism is nothing but a front for dogmatic subjectivism.[6]

Stratton-Lake charges critics of rational intuitionism with attributing all of these views to contemporary intuitionists, though intuitionists no longer hold any of them. Contemporary intuitionists claim only what Audi calls "soft self-evidence" for moral principles, meaning, among other things, that the principles need not be obvious to be self-evident. (Contrary to popular opinion, W. D. Ross's view is arguably closer to the

---

[5] Samuel Clarke, *A Discourse Concerning the Being and Attributes of God, the Obligations of Natural Religion, and the Truth and Certainty of the Christian Revelation* (Glasgow: Richard Griffin, 1823), 159; or see: L. A. Selby-Bigge, *British Moralists*, 2 vols. (Whitefish, MT: Kessinger, 2007), Vol. 2, §484; or D. D. Raphael, *British Moralists: 1650-1800*, 2nd Revised ed., 2 vols. (Indianapolis, IN: Hackett, 1990), Vol. 1, §227.

[6] William Paley, *The Principles of Morality and Political Philosophy* (n.p.: P. Byrne, L. White, and W. M'Kenzie, 1788), I.v; J. S. Mill, *Autobiography* (New York: Penguin, 1990), 148; John Dewey and James H. Tufts, *Ethics* (New York: Henry Holt, 1908), 321; Michael R. DePaul, "Intuitions in Moral Inquiry," in *The Oxford Handbook of Ethical Theory*, ed. David Copp (Oxford: Oxford University Press, 2006), 595; Alisdair MacIntyre, *A Short History of Ethics*, 2nd ed. (Notre Dame, IN: Notre Dame University Press, 1996), 254; R. M. Hare, *Essays in Ethical Theory* (Oxford: Oxford University Press, 1989), 147; P. H. Nowell-Smith, *Ethics* (Harmondsworth: Penguin, 1954), 26.

contemporary view than the Hotline View.[7]) As Stratton-Lake and Audi both show, this subtle theory of self-evidence undermines the simple objections from disagreement and dogmatism so often raised against intuitionism.[8] Along with a number of other improvements to the theory, this advance has again made rational intuitionism the object of serious study in moral philosophy.

This softening of self-evidence is a key element in the revival of rational intuitionism because it allows intuitionists to maintain that some moral principles are self-evident in the face of disagreement and uncertainty about those principles. Audi defines soft self-evidence negatively, as the opposite of "hard" self-evidence, which is marked by four features. A proposition is self-evident in the hard sense if:

(1) it is "strongly axiomatic," meaning that "there is nothing epistemically prior to [it]";

(2) it is "immediate," meaning that it is possible for a normally intelligent adult to have an adequate understanding of it immediately, without prolonged reflection on its meaning;

(3) it is "indefeasibly justified";

(4) it is "cognitively compelling" or "luminous," meaning that anyone who understands the proposition is irresistibly moved to accept it.

A self-evident proposition that meets none of those criteria is self-evident only in the soft sense. Propositions that meet some, but not all, of these criteria will be neither "perfectly hard" nor "perfectly soft."

Softly self-evident propositions will not leap from the page, blazing with the light of Reason, to ensconce themselves in the reader's mind. Grasping the truth of such a

---

[7] Philip Stratton-Lake, "Pleasure and Reflection in Ross's Intuitionism," in *Ethical Intuitionism: Re-Evaluations*, ed. Philip Stratton-Lake (Oxford: Oxford University Press, 2002), 114–19.
[8] Stratton-Lake, "Introduction," 18–23; Audi, *The Good in the Right*.

proposition may require sustained, careful reflection, the purpose of which is to develop an adequate understanding of the proposition. Even upon acquiring such an understanding, however, some may not believe the proposition. After all, as Stratton-Lake notes, calling the proposition self-evident is only to say that it is "know*able* on the basis of an understanding of it," not that anyone who understands it will believe it.[9] Furthermore, since recognizing the truth of a self-evident proposition and recognizing that it is self-evident are two different things, it may be possible to grasp the proposition's truth without knowing that the proposition is self-evident.

Audi and Stratton-Lake offer some self-evident propositions that exhibit at least some features of soft self-evidence. Audi suggests 'If there had never been any siblings, then there would not have been any first cousins'. Stratton-Lake mentions 'Circles are figures bounded by a line that is equidistant from its center' and 'If all As are Bs, and no Cs are Bs, then no Cs are As'. These may not be obvious at first sight, but they do seem to be self-evident upon careful reflection.

By attributing only this softer kind of self-evidence to moral principles, rational intuitionists can make some sense of disagreement and uncertainty about those principles. They can also make room for additional arguments to support those principles, sidestepping the objection that intuitionism is nothing more than a dogmatic form of subjectivism.

2. The "Thermometer View" of Moral Appearances

Despite the advantages of this softened notion of self-evidence over the Hotline View, contemporary intuitionists are still committed to a general view of intuition that I call the "Thermometer View." Even if this is an adequate account of some forms of

[9] Stratton-Lake, "Introduction," 20–21.

noninferential knowledge, however, it is psychologically and epistemologically problematic as an account of intuitive *moral* knowledge.

According to the Thermometer View, at least some humans possess a reliable cognitive capacity to detect the truth of softly self-evident propositions after they have reflected adequately on them. Just as a thermometer gives an accurate reading of an object's temperature only after given sufficient time to interact with it, free of disturbing influences, this cognitive capacity gives an accurate, noninferential "reading" of some propositions' truth only after given sufficient time to reflect upon them, free of disturbing influences. (For propositions that are neither self-evident nor self-contradictory, however, the "reading" should be only that it cannot determine the truth of the proposition.)

A thermometer is an apt analogue for the faculty of moral intuition for several reasons. Like moral intuition, a thermometer does not deliver instant results; the thermometer must be allowed time to interact with what it is measuring, just as according to the intuitionist, the ordinary person must be allowed time to reflect on a moral principle. The thermometer must be functioning properly in order to give a reliable result; outside influences must be eliminated. Likewise, our power of intuition must be functioning properly to be reliable; outside influences must be eliminated. Our justification for accepting a thermometer's reading may be defeated, even when the thermometer is correct, just as Audi says that our justification for accepting a moral intuition can be defeated, even when the intuition is correct.

The Thermometer View is essentially just a psychological restatement of the central rational intuitionist claim that humans can recognize the truth of softly self-evident principles after adequate reflection, although they sometimes fail to do so. To deny any part of the view is to deny some part of that central claim. It is possible to deny

the Thermometer View and yet maintain that there are self-evident propositions, for one is a psychological claim and the other a logical one; but self-evident claims that are inaccessible to us (except perhaps inferentially) are of no use to rational intuitionism as a moral epistemology. Thus, any rational intuitionist who holds that rational intuition actually yields moral knowledge is committed to the Thermometer View.

Unfortunately for the rational intuitionist, the Thermometer View cannot underwrite the thesis that we have noninferential knowledge of moral principles.

The view suffers from two major problems. First, the view is psychologically mysterious. It is unclear what psychological mechanisms we might have that could do what we are supposed to be able to do on the Thermometer View. Contrary to some intuitionists' claims, grasping moral principles does seem to require more than just an ability to "understand and think."[10] This is because moral principles are importantly different from the other kinds of propositions that intuitionists suggest as candidates for soft self-evidence, as I will argue below. Second, the view is epistemically inadequate. If moral appearances arise from the kind of faculty suggested by the Thermometer View, then they do not justify moral principles. This is because disagreement about moral principles constantly undermines one's justification for accepting the results of one's "moral thermometer." Thus, contemporary versions of rational intuitionism are no better off than widely rejected intuitionist epistemologies, such as Moore's, that posit occult faculties of moral intuition and denounce moral dissenters as morally corrupt or immature.

---

[10] Ibid., 22.

2.1. The psychologically mysterious "mental thermometer"

Some critics allege that rational intuitionism requires an occult faculty of mind, but contemporary intuitionists insist otherwise.[11] Indeed, one advantage of rational intuitionism is that it relies only on the notion of self-evidence, which is less mysterious than the occult faculties of intuition required by theories like Moore's.[12] Since rational intuitionism seems to require nothing more than our comparatively ordinary ability to grasp basic truths, its defenders believe it to be psychologically unproblematic. When we pry more deeply into the cognitive mechanisms that would be required for specifically *moral* intuition, however, the psychology becomes mysterious once more.

To establish that we do have the ability to grasp softly self-evident moral principles, rational intuitionists typically point out other propositions that are softly self-evident, claiming that the process by which we intuit those is reliable and that we seem—introspectively, perhaps—to use the same process to intuit moral principles. The argument, in short, is that the Thermometer View works for some self-evident propositions, and so it should work for moral principles. Our "mental thermometers" are reliable when it comes to some softly self-evident propositions, so it is reliable with respect to softly self-evident moral principles. However, there are important differences between moral principles and the other propositions to which intuitionists compare them, and these differences make the mechanisms of moral intuition mysterious. Consider the following list of allegedly self-evident but unobvious propositions from Audi and Stratton-Lake:

---

[11] See, e.g., William Frankena, *Ethics* (Englewood Cliffs, NJ: Prentice-Hall, 1973); cf. Audi, *The Good in the Right;* Stratton-Lake, "Introduction."

[12] Some philosophers find self-evidence mysterious, as well. Rather than try to settle that question, however, I will grant, for the sake of argument, that some propositions are self-evident. If there are no such propositions, so much the worse for rational intuitionism.

(a) If there have never been any siblings, there have never been any first cousins.

(b) The existence of great-grandchildren is impossible apart from that of four generations of people.

(c) Circles are figures bounded by a line that is equidistant from its center.

(d) All female mammals suckle their young.

(e) If all As are Bs, and no Cs are Bs, then no Cs are As.

All of these can be fairly easily derived from basic logical principles, quasi-logical principles (e.g., about relations about family members), or definitions (e.g., of circles).[13] They are all, in a sense, theorems of an easily formalized system. (a) follows from the definition of siblings as 'children of the same parents' and a definition of first cousins as 'the respective children of two or more children of the same parents'. (b) and (c) follow from the definition of 'great-grandchildren' and 'circle', respectively. (e) follows from basic principles of set theory and non-contradiction. In other words, it is possible to check each of these propositions by transforming specifiable and uncontroversial premises in specifiable and uncontroversial ways. Perhaps more importantly, the premises from which these inferences begin are either definitions, basic logical truths, or geometric axioms, and the necessary transformations are essentially formal transformations. Furthermore, the axioms and definitions from which they are derived are taken to be either conventional or self-evident in the hard sense.

Psychologists have not yet studied the cognitive mechanisms by which we grasp such propositions. Introspectively, however, we seem to have the capacity to apprehend them noninferentially—that is, the ability to recognize that they are true without

[13] If (d) seems an exception to this, that is because (d) is false—at least as stated. The truth that it approximates is that the females of all mammal species have mammary glands, which are typically used to suckle young; this is part definition and part empirical fact.

consciously inferring them from more basic premises, provided that we understand the words in them. There are, however, some plausible hypotheses that explain this capacity in unmysterious ways.[14] For instance, it may be that once we call specific definitions or axioms to mind (such as the definitions of 'sibling' and 'first cousin'), our brains nonconsciously complete the formal transformations needed to derive (a) from those definitions (or, at least, a reliable heuristic proxy for those transformations). On this hypothesis, the steps by which we reach these allegedly self-evident propositions are steps that we might follow if we were consciously deducing the propositions from definitions or axioms, but we are not aware of making those inferences. If the steps never enter our consciousness, our grasp of (a) will appear noninferential to us; (a) will seem self-evident. Those with an inadequate understanding of the proposition or insufficient logical acumen may fail to grasp the proposition without an explicit account of the inference, but to those who are prepared, the propositions are knowable noninferentially.

This provides one reasonably definite and plausible psychological account of how the Thermometer View works with respect to these kinds of propositions. This explanation of these propositions' self-evidence yields testable hypotheses. It predicts, first, that a person's likelihood of noninferentially grasping the truth of a proposition varies directly with the person's logical acumen and with the complexity of the proof of the proposition. (A sufficiently intelligent, well-trained being might be able to intuit the truth of the proposition that the measure of each interior angle of a pentagon is 108°, though it seems unlikely that any human can.) It also predicts that the content of a

[14] Although I offer one hypothesis here, there are surely others. Furthermore, we may have various mechanisms for grasping different kinds of propositions (e.g., arithmetic, geometric, social, etc.), and different people may use different mechanisms for grasping the same proposition. These are empirical questions, and settling them would require psychological research.

proposition will affect people's ability to grasp it noninferentially, even keeping the logical form of the proposition constant. This is because humans are naturally better at certain kinds of reasoning (e.g., social reasoning) than others (e.g., abstract reasoning or reasoning about probability).

But most importantly, this account of the cognitive processes underlying soft self-evidence entails that a different cognitive process is at work when reflecting on propositions that are not formally derivable from more widely accepted basic premises. For instance, compare these propositions with some candidate moral principles:[15]

(f) We (morally) should keep our promises.

(g) We (morally) should make amends for our wrong-doing.

None of these can be derived in the ways that (a)–(e) can. There are no uncontroversial definitions or basic logical truths from which these can be derived by uncontroversial steps. If there are unconscious inferences at work behind an intuition about one of these propositions, those inferences are not formal; they depend on contentious unstated premises. The process probably depends on the cognitive relations between concepts—understood as mental representations rather than abstract entities—involved in the thought being entertained, and on collateral beliefs, which may or may not be well justified. At best, then, this process could show that one's concepts and beliefs are related in a suitably coherent way—but this is not a reasonable guarantee of truth. Whatever cognitive capacities make the Thermometer View a plausible account of our intuitive knowledge of propositions like (a)–(e) seem inadequate for noninferentially evaluating propositions like (f) and (g).

---

[15] These are drawn from Audi's list of self-evident first-order moral principles in Robert Audi, "Prospects for a Value-Based Intuitionism," in *Ethical Intuitionism: Re-Evaluations*, ed. Philip Stratton-lake (2002). See pp. 187–195.

If this hypothesis about differing cognitive processes is correct, then it seems that our reliability in performing formal inferences nonconsciously does not show that we can reliably perform informal inferences nonconsciously. The fact that our noninferential assent to an essentially formal proposition is a reliable indicator of truth does not mean that our noninferential assent to an informal proposition is a reliable indicator of truth.

This does not, strictly speaking, show that we do not have the capacity to detect the truth of informal propositions, such as moral principles, noninferentially, but it puts a significant burden on the intuitionists. They must explain and argue for a cognitive capacity that can detect the truth of informal propositions without the tacit cognitive use of contentious premises. How such a capacity could work—and in particular, how it could be shown to detect truth rather than just certain kinds of relationships among one's own concepts—is mysterious.

To make matters worse, rational intuitionists' explanation of moral intuition must fare better than the alternative hypothesis that noninferential assent to moral principles is largely the result of acculturation and biology. That is, their explanation must be stronger than the simpler explanation that because we have come to believe certain things so strongly and understand certain concepts in specific ways, certain propositions simply appear to be true; but that had we been raised differently enough to have different concepts or treat different beliefs as fundamental, those same propositions may seem contingent or even false. Given how strongly one's social groups tend to influence one's concepts and one's moral judgments, and given the powerful sentimentalist elements in leading accounts of moral psychology, rational intuitionists have a steep disadvantage to overcome in offering that explanation.

2.2. How peer disagreement undermines "moral intuitions"

Suppose, for the sake of argument, that humans do have a reliable, if mysterious, capacity of rational intuition. Can it do the epistemological work that rational intuitionists ask of it? It seems not, for whatever justification one derives from the capacity for a particular belief would be constantly defeated by peer disagreement.[16] This is because the Thermometer View, as the name suggests, introduces a surprising element of epistemic externalism into the notion of self-evidence, which causes serious problems in cases of disagreement.

Epistemic externalism is the thesis that something other than the content of one's mental states partly determines whether one is justified in any given belief. For instance, whether one is justified in believing that a room is quiet depends, in part, on whether one's auditory system is working properly—and not just on whether one *believes* that one's auditory system is working properly. Whether one is justified in believing, on the basis of a thermometer reading, that the temperature in a room is 72° Fahrenheit depends, in part, on whether the thermometer is working. On the Thermometer View of intuition, whether one is justified in believing a self-evident proposition depends on whether one's capacity for rational intuition is functioning properly, has been exposed to enough reflection, and has not been unduly influenced by outside forces. Someone who intuits a moral principle is, in this respect, like someone who reads the ambient

---

[16] Epistemologists have been discussing "peer disagreement" vigorously over the last few years, developing several major views on it. See, e.g., Adam Elga, "Reflection and Disagreement," *Nous* 41, no. 3 (2007); Richard Feldman, *Epistemology* (Upper Saddle River, NJ: Prentice Hall, 2003); Feldman, "Epistemological Puzzles About Disagreement," in *Epistemology Futures*, ed. Stephen Hetherington (Oxford: Oxford University Press, 2006); Thomas Kelly, "The Epistemic Significance of Disagreement," in *Oxford Studies in Epistemology*, ed. Tamar Szabo Gendler and John Hawthorne (Oxford: Oxford University Press, 2005); Kelly, "Peer Disagreement and Higher Order Evidence," in *Disagreement*, ed. Richard Feldman and Ted Warfield (New York: Oxford University Press, forthcoming).

temperature from a thermometer. Having no (conscious) access to the processes leading to the conclusion, her belief in the principle or about the temperature is justified only if her capacity for intuition, or the thermometer, functioned properly.

Rational intuitionists might complain that this view of intuitionism relies on a "strongly axiomatic" view of intuition—the mistaken view that reasons cannot be given for self-evident propositions—but they would be wrong. Reasons can be given to support a self-evident proposition, and doing so could provide a strong basis for believing that our intuitive capacity is reliable with respect to moral principles, replacing the argument, criticized above, that it is reliable with respect to moral principles because it is reliable with respect to unobvious formal propositions. However, there are two points to bear in mind. First, these reasons do not count as reasons to think that the proposition is self-evident. Proving a proposition cannot show that it needs no proof. Second, even if we know that our capacity for intuition is reliable, any particular intuition that has not been supported by reasons is still like a reading taken from a thermometer. Its justification still depends on the assumption that the capacity is functioning properly *on this occasion*. Thus, if this externalism causes epistemological problems for the Thermometer View, then whenever such problems arise, intuitions can only be justified by giving reasons for them, rendering intuition epistemically useless in those cases.

In fact, this externalism does cause problems for the Thermometer View because of what epistemologists call "peer disagreement." Peer disagreement occurs when two "epistemic peers"—that is, two people who are equally epistemically virtuous with respect to the subject matter at hand—disagree about something, even after being exposed to all of the same evidence and reasoning. For instance, suppose that two equally thoughtful, intelligent, and knowledgeable political pundits reviewed the same arguments and evidence about which presidential candidate will receive the Democratic

nomination in a given election. After considering this evidence, one predicted that one candidate would win the nomination and the other predicted that another candidate would win. The epistemological question in such cases is how, if at all, the two epistemic peers ought to revise their respective beliefs upon learning that they disagree.

Epistemologists disagree about peer disagreement. Some hold what is called the "No Independent Weight View" (NIW) of disagreement, according to which at least one of the pundits is rationally entitled to maintain exactly the same confidence in her original belief. This is so even when she cannot find any flaw in the other's reasoning or premises. Others hold the "Equal Weight View" (EW), on which both pundits must "split the difference" between their views. This might amount to suspending belief altogether; it might amount to believing that one of those two candidates will win; or if belief is not considered an "all-or-nothing" matter, then each commentator might believe, to some degree, that the first candidate will win and, with equal credence, that the other will win. Other positions in the debate include Thomas Kelly's "Total Evidence View" (TE), which Kelly locates between EW and NIW, and Jonathan Adler's view that disagreement should not undermine belief, though it should sometimes diminish one's confidence in that belief.[17]

Because they are committed to the Thermometer View of intuition, rational intuitionists are committed to the Total Evidence View of peer disagreement, at least with respect to moral appearances.[18] TE occupies an intermediate position between EW and NIW. On NIW, only one's original evidence counts. On EW, one's original evidence is swamped by the fact of disagreement. By contrast, TE assigns weight to both one's original evidence and to the fact of disagreement. Specifically, according to TE, one ought to temper one's response to the original evidence only to the extent that one has

---

[17] Jonathan Adler, personal communication.
[18] Kelly, "Peer Disagreement and Higher Order Evidence."

good evidence that one's dissenting peer performed at least as well, epistemically, in evaluating the evidence. Suppose, for instance, that a mathematician discovers what he thinks is a sound proof for a famously unproven conjecture. How he should respond upon learning that a respected colleague thinks that the proof is unsound depends on what he knows about his colleague's state at the time of her assessment. According to TE, if she has not slept in three days, then he may discount her disagreement because he knows that this makes it likely that she has evaluated the current evidence inappropriately; if, however, his only reason to think that she has failed to appreciate his proof is because she disagrees with him, then he must split the difference, as per EW.[19]

Because of its externalist consequences, the Thermometer View of moral intuition commits rational intuitionists to TE with respect to moral appearances. Suppose that a cook measures the temperature of a soup, getting a reading of 125° F. In normal circumstances, we would probably grant that the thermometer's reading is sufficient evidence to justify his belief that the soup is 125°. Imagine, however, that a second cook measures the same soup at the same time with another thermometer, which we have every reason to believe is just as reliable as the first one. The second cook gets a reading of 130°. Neither cook would be rationally justified in maintaining, as NIW recommends, that his thermometer was correct. For each now knows that there are two thermometers and that at least one of the thermometers is incorrect, but neither knows *which* thermometer that is. Thus, even if the first cook's thermometer is working properly, and thus he would have been justified in believing it in the absence of the second cook, the reading of the second cook's thermometer defeats that justification. If the cooks have "higher order evidence" about the functioning of the two thermometers, such as evidence that one was touching the bottom of the pot, then they can use this

[19] Ibid.

evidence to decide which thermometer to trust and which to ignore. But without such evidence, as TE suggests, the rational thing to do would be to follow EW by suspending belief, adopting a disjunctive belief, or giving equal credence to each thermometer's reading.

It is essential to notice the externalist element of this example. The *only* reason either cook has to think that the soup is at any particular temperature is that the thermometers say so. The fact that the second thermometer belongs to another cook is irrelevant. If the first cook had measured the soup himself using the two thermometers, getting different readings from each, he would be in the same epistemic quandary. The cook cannot therefore fall back on his "original evidence," as proponents of NIW sometimes recommend. To do so would be to ignore important new evidence. Nor would it be appropriate for the first cook to continue to believe, though with reduced confidence, that the soup is 125°. He ought not to believe that the soup is 125°. Nor should he believe that the soup is 130°.

The Thermometer View is epistemically pernicious for the rational intuitionist because it puts dissenting ethicists in the same position as the cooks with divergent thermometer readings. When an ethicist "intuits" a moral principle, the only reason he can give for it is that it noninferentially appears correct to him after prolonged, serious reflection. If another ethicist, after equally serious reflection and with an equally strong understanding of the relevant concepts, "intuits" a different principle, or does not come to accept that the principle in question on the basis of reflection, both ethicists seem now to be in the same position as the cook or cooks who discover that two different thermometers have given different readings.

To save their theory, intuitionists must show either that ethicists with conflicting intuitions are not in the same epistemic position as the cooks or that there is some way

for them to escape that epistemic predicament. Like the cooks, intuitionists are unable to fall back on their reasons for their intuition, since they have none. Thus, their best hope of showing that their position is not as epistemically precarious as the cooks' is to argue that 'Stick with your own intuition' is a justifiable heuristic in the ethical case. That heuristic might be pragmatically justifiable, since reconsidering one's views take time and effort, but it cannot be epistemically virtuous. (A heuristic is "epistemically virtuous" if a majority of the beliefs acquired by using it are true.) For 'Stick with your own intuition' to be an epistemically virtuous heuristic in the face of peer disagreement, it would need to be the case that each peer's intuitions were correct more often than the other's. This is impossible. If Tweedle Dee's intuitions are right in more than half of the cases in which he disagrees with Tweedle Dum, then Tweedle Dum's intuitions must be right in less than half of those cases. Thus, for one of the two epistemic peers, 'Stick with your own intuition' is not epistemically virtuous. Furthermore, if Tweedle Dee knew that his intuitions were right (significantly) more often than Tweedle Dum's, then the pair would not be epistemic peers, and they would never find themselves in this situation.

If we grant, then, that the intuitionists are at least initially the same situation as the cooks, their only hope for saving their intuitions is to find a way out of their deadlock. The most direct way to do this would be to show that the others' intuition is incorrect. The problem for the rational intuitionist is that providing such an argument renders the intuition useless. As Audi and Stratton-Lake rightly stress, it might be possible to give reasons for a self-evident proposition. But in any case in which one provides reasons for an intuited principle, one ceases to rely on intuition, and so if one must do this in all cases, then intuition contributes nothing to the justification of one's principles. Thus, if rational intuitionists wish to rely on intuitions in cases of disagreement, then they must

rely on indirect evidence that they have intuited correctly, and those who disagree have erred.

It is not clear what sort of evidence is available, however. Temporary conditions, such as sleep deprivation or intoxication, would be reason to trust someone's capacity for rational intuition, but these do not provide plausible explanations of persistent disagreement. One promising approach is to show that many others also intuit the proposition. After all, to claim that one has intuited a proposition correctly entails that the proposition is self-evident; and it is unclear what counts as evidence that a proposition is self-evident except widespread acceptance of it by those who understand it clearly. However, supporting an intuited principle in this way faces two obstacles: First, intuitionists will need to show that many or most of those who understand their principles most thoroughly—in this case, that presumably means moral philosophers—do actually agree about them. Second, even if there is widespread agreement on a principle, the hypothesis that the principle is self-evident must stand up to the competing hypothesis that cultural or biological influences cause widespread agreement, implying that if we were raised or bred differently, we would think differently.

If these obstacles cannot be overcome, then intuitionists' justification for believing intuited moral principles will be constantly defeated by peer disagreement. To make matters worse for the intuitionists, if one's justification for believing a proposition is defeated, then so is one's justification for believing that the proposition is self-evident. Suppose it were possible to be justified in believing that some proposition is self-evident but not justified in believing the proposition itself. The contradiction here is not hard to find. The self-evidence of a proposition, in the relevant sense of "self-evidence," deductively entails that proposition. Assuming that justification is closed under deductive entailment, at least in this context, this means that someone who justifiably

believes that *p* is self-evident would, *contra* the initial supposition, also be justified in believing that *p*. Thus, peer disagreement not only leaves the intuitionists without any justified moral beliefs, but it strips them of the right to claim that any moral principles can be justified by appealing to self-evident moral intuitions.

One interesting feature of this objection is that it does not deny that there are self-evident moral principles. It only denies that we can be justified in believing we have identified one whenever there is peer disagreement about it. This is an advantage when dealing with current versions of intuitionism because, by making it hard to prove that there are self-evident principles, intuitionists have also made it hard to prove that there are no self-evident principles. If this objection succeeds, then intuitionism fails as a moral epistemology even if intuitionists are correct in their central claim that there are self-evident moral principles.

Another important feature of this objection is that it does not turn on a specific view about peer disagreement in general. It might be that in some kinds of cases, peer disagreement does not constitute a reason to withdraw one's belief. Because of the externalist consequences of the Thermometer View, however, rational intuitionists are not entitled to take that position with respect to moral appearances. Since rational intuitionists are committed to the Thermometer View, they are committed to the view that peer disagreement undermines the power of moral appearances to justify moral claims.

3. Why disagreement persists about intuitionists' moral principles

An objection from disagreement, of course, carries little weight if there is no disagreement. As W. D. Hudson notes, the disagreements that make moral judgments

appear to differ so drastically between cultures is sometimes, if not often, superficial.[20] However, the rational intuitionists cannot hide behind general agreement in basic moral values, for the question is whether there is disagreement about the specific principles that they claim to be self-evident. There is such disagreement about intuitionists' principles, even among thoughtful people, and this disagreement arises out of features of those principles that rational intuitionists cannot easily abandon. Specifically, the disagreement results from the postulation of *pro tanto* duties. Since, however, rational intuitionists face strong theoretical pressure to posit such duties, they cannot easily escape such disagreement.

As Klemens Kappel argues, the reason that principles of *pro tanto* duty engender disagreement is that they imply that a given feature of an act is *always* a right- or wrong-making feature of the act, even if it is sometimes overridden by a more important right- or wrong-making feature in particular circumstances.[21] The most plausible examples of self-evidently *pro tanto* duties are the duty not to harm others and the (imperfect) duty to benefit others, but some particularists, such as Jonathan Dancy, dispute even these examples. According to Dancy, there are no *pro tanto* duties in the relevant sense because, for any given feature of an act, that feature sometimes makes the act better, sometimes worse, and is sometimes neutral with respect to the moral value of the act.[22]

But even those who believe that harming others always makes an act worse will likely dispute one or more of the intuitionists' other, more specific principles. Are lying and promise-breaking *always* wrong-making features of an act—even in cases where everyone benefits? Does showing gratitude *always* make an act morally better—even if

---

[20] William Donald Hudson, *Ethical Intuitionism* (London: Macmillan, 1967), 52–53.
[21] Klemens Kappel, "Challenges to Audi's *Ethical Intuitionism*," *Ethical Theory and Moral Practice* 5, no. 4 (2002).
[22] Jonathan Dancy, *Ethics without Principles* (Oxford: Oxford University Press, 2004).

one's benefactor is a cruel dictator who dispenses benefits only to curry enough favor to suppress dissent?

Alternatively, some philosophers might agree that lying is generally wrong, but insist that this is so because lies generally harm others, not because they are lies *per se*. This is another way of disagreeing that lying is a wrong-making feature of an act.[23]

Because of these controversies, rational intuitionists will face disagreement as long as they posit *pro tanto* duties, but the need to provide useful, plausibly self-evident principles commits them to *pro tanto* duties. Thus, they cannot escape disagreement. They are committed to *pro tanto* duties for two reasons. First, few, if any, principles of exceptionless duty are plausibly self-evident, since such principles are either implausible or too complicated to be self-evident. Second, unless *pro tanto* duties are construed in such a way that being of the relevant act type is *always* a right- or wrong-making feature of an act, the principles that prescribe them are virtually useless.

Almost all straightforward principles of exceptionless duty are implausible. Common principles like 'One ought not to kill' or 'One ought not to lie' are riddled with exceptions. Most people admit that killing in self-defense is morally acceptable; some claim that executions, killing in war, or killing to protect one's family's honor are morally acceptable. Even more basic principles like 'One ought not to harm others' have exceptions, as when inflicting some small harm on one person is necessary to avoid great harm to others.

Most people admit that these principles have exceptions, of course, but the problem for the intuitionist is that there is no way to build in the myriad exceptions without sacrificing the self-evidence of the principle. Sidgwick, for instance, provides a painfully detailed attempt to tease out the exceptions, as he saw them, to commonsense

---

[23] Kappel, "Challenge's to Audi's *Ethical Intuitionism*," 398–99.

moral rules; but the resulting principles, bristling with complications and qualifications, are not plausibly self-evident.[24]

While intuitionists could extricate themselves from this particular difficulty by allowing that self-evident principles only express general tendencies of act-types, this would render their principles useless. Suppose that all we knew about the moral qualities of lying is that lying *generally* makes an act wrong. This provides no real guidance in any particular case, because it gives us no way to decide whether *this* case is one in which lying carries moral weight; and in order to develop more detailed theories about when lying is morally significant, we would need to be able to judge individual cases without appealing to general principles, thus obviating the need for the principles altogether. So, even if 'Lying is generally wrong' were self-evident, it would be useless as a guide to forming concrete moral judgments.

Thus, rational intuitionists do face serious inescapable disagreement about principles they hold to be self-evident. And because that disagreement defeats their justification for believing in the truth and self-evidence of their allegedly intuited moral principles, their theory cannot adequately justify the use of moral appearances in moral philosophy.

4. Conclusion

The idea that moral claims are truths of Reason is a very old one. It reaches its clearest form in rational intuitionism, of which the most sophisticated version is Robert Audi's. The theory is important because it promises one way to evade Harman's argument from explanatory irrelevance. Moral appearances may not provide "data" for moral claims in

[24] Henry Sidgwick, *The Methods of Ethics*, 7th ed. (Indianapolis: Hackett, 1981), 217–361.

the same way that empirical observations provided data for claims about the physical world, but if they are deliverances of Reason, they may be justified nonetheless.

Even in its latest incarnation, however, rational intuitionism fails as a theory of moral epistemology. It fails on two counts. First, the cognitive mechanisms that it requires are psychologically mysterious. It is implausible to suppose that the cognitive mechanisms underlying our noninferential assent to claims like 'All first cousins share at least one grandparent' also underlie our moral appearances. But it is mysterious what other cognitive mechanisms might provide us with rational insight into moral claims. Unless rational intuitionists can show that such a mechanism exists, we are not warranted in thinking that moral appearances are the product of rational processes that reliably produce true beliefs.

Second, rational intuitionism is committed to the "Thermometer View" of moral intuition. This view introduces a surprising externalist element into the theory. As a result, even self-evident moral appearances will succumb to "peer disagreement." And such disagreement is bound to persist because of disagreements over *pro tanto* duties, which cannot be expunged from intuitionist systems of ethics. Thus, even if there were some psychological mechanism for recognizing self-evident moral claims, we would still not be justified in accepting the claims on the basis of that mechanism.

The failure of rational intuitionism is a significant loss to those who wish to preserve the importance of moral appearances in moral epistemology. If moral appearances do not provide evidence for moral claims, and they do not result from the rational apprehension of moral truths, then their traditionally privileged role in the practice of moral philosophy looks untenable, as I argued in the Chapter Three. If the question of how one should live amounts to anything more than the question of what one most wants, ethicists will need some other way of figuring out what more is needed.

5 Practical Coherence

There are two ways to interpret Ivan Karamazov's thought that, without an afterlife, "everything would be permitted."[1] It might mean that no moral claim is justifiable—that moral nihilism is true. But it might mean that every moral claim is justifiable—that for any action, there is some justifiable moral claim entailing that the action is permissible. This pure relativism would be as disastrous for moral epistemology as nihilism, for it would reduce the justification of moral claims to an arbitrary choice between infinitely many equally valid moral claims.

Some may fear that discounting the use of moral appearances in moral epistemology, as advocated in the preceding chapters, leaves us in just this kind of relativistic crisis. Ethicists often use moral appearances to limit the range of admissible moral claims. Theories are rejected because they entail the permissibility of an act that plainly appears to be wrong. Moral claims are endorsed and their negations denied because they plainly appear to be correct. If we cannot use moral appearances to justify moral claims, then it might seem that we have no means of criticizing any moral claims—no means of showing that any moral claim is any better than any other.

We might think of this as the "Karamazov Problem": unless we are willing to endorse complete relativism, we must find ways to show that one moral claim should be preferred over another. We must find ways to show that not everything is permitted.

Without moral appearances, however, there is no way to show that individual moral claims are preferable to others in all circumstances. We have no way to test individual moral claims outside of a system of moral claims, just as we have no way to

[1] Fyodor Dostoyevsky, *The Brothers Karamazov*, tr. Andrew R. MacAndrew (New York: Random House, Bantam Dell, 1970), 88.

test individual claims about the physical world outside of a system of beliefs about the world. Moral systems confront the Karamazov Problem as "corporate bodies." And so the real challenge is to show that some moral system—that is, some set of moral claims—should be preferred over some others.

Ethicists use a number of methods for criticizing moral systems without comparing them to some Platonic moral reality. These methods include exposing non-evaluative error, highlighting inconsistencies between moral claims, and showing that some moral systems cohere better with our nonmoral evaluations. The result is a coherentist view of moral epistemology. This coherentist theory can be improved, however, by introducing the notion of "practical coherence," which provides a way of anchoring systems of moral and nonmoral evaluations to the world. By conceiving of ethical inquiry as the search for practical coherence, we can develop a means of doing ethics that solves the Karamazov Problem without appealing to moral appearances.

1. Some grounds for criticizing moral claims

Philosophers have proposed a number of grounds, independent of moral appearances, for criticizing or rejecting moral systems. These grounds provide a partial solution to the Karamazov Problem: they show that not all moral systems are equally good.

1.1. Nonmoral error

Even the staunchest anti-realists about morality maintain that we can sometimes criticize particular moral claims by showing that they rest on false nonmoral premises. The epistemic basis for this criticism is not that such claims are known to be false, but

rather than they receive no support from the moral and nonmoral belief system to which they belong.

Consider, for instance, the plot twists of Alfred Hitchcock's *Rear Window*. Like most people, Jimmy Stewart's character, L. B. Jeffries, takes a dim view of killing one's spouse. When Jeffries thinks that his neighbor, Lars Thorwald, has killed his wife, dismembered her corpse, and shipped it away, he thinks that Thorwald has done something morally wrong. When a friend convinces Jeffries that Thorwald's wife has not been killed, Jeffries revises his moral assessment of his neighbor. He concludes that his negative moral claim about Thorwald was incorrect.

Formally, we can think of cases like this as instances of the following pattern of reasoning:

(1) If nonmoral condition *C* is met, then *M* (where '*M*' is moral claim).

(2) *C* is met.

∴ (3) *M*.

In cases of nonmoral error, (2) is false. Thus, *ceteris paribus*, we err in concluding that *M*. When we learn that (2) is false, we withdraw our conclusion.

Of course, the falsity of (2) does not entail the falsity of (3). Thorwald might have killed someone other than his wife, and therefore might have done something morally wrong anyway. What underwrites the rejection of (3), then, is not that we have reason to disbelieve, but only that the web of beliefs to which it belongs provides no support for it. *Ceteris paribus*, a system without such "free-floating" claims is better than a system with such claims.

The limitation of this method for criticizing moral claims is that it can only criticize particular applications of general moral claims. In the schema above, both (1) and (3) are moral claims. Only (3), however, can be revised on the basis of nonmoral error. If someone believed that a person is good if they kill and scalp strangers for fun, we may not be able to dissuade them using this method. In particular, we could not point to an instance of recreational scalping, hope that it elicits a moral appearance, and then use that moral appearance to justify a counterexample to his evaluation of scalping. Such moral appearances have no justificatory power.

Thus, criticizing moral systems on the basis of nonmoral error is useful, but only in limited contexts.

1.2. Inconsistency

Inconsistency within a moral system is as much a reason to revise our moral claims as inconsistency in non-evaluative beliefs is a reason to revise those beliefs. Since moral claims are prescriptive, we only frustrate our own goals by evaluating the same thing in conflicting ways.

Exposing inconsistencies in moral systems is among the most common form of moral argument. One way to do this is to show that an object that would be evaluated positively would also, perhaps under a different description, be evaluated negatively. Another way to do this is to show that an object that would be evaluated in one way is an instance of a type which we are generally disposed to evaluate differently. In either case, we must change some moral claim if we are to escape inconsistency.

Some standard arguments for or against the permissibility of abortion illustrate the first method of pitting moral claims against one another. One standard argument against abortion is to point out that the procedure, which might seem acceptable when

described as allowing a woman to avoid an unwanted child, seems impermissible when described as murdering an innocent and defenseless child. A standard response is to redescribe the action again as preventing a fetus from developing into a child. Both of these arguments involve describing the same action in different ways to warrant a new moral claim that is incompatible with the previous one. The goal is to prompt an interlocutor to reject one of the two conflicting moral claims on the grounds that endorsing both of them leads to contradiction.

The second method of pitting moral claims against one another is slightly trickier. Consider Peter Singer's famous argument for poverty relief. Singer notes that most of think it would be monstrous to allow a child to drown in a shallow pond in order to avoid ruining our clothes. This, he suggests, is because we accept that we ought to prevent very bad things from happening if we can do so without sacrificing anything of significant moral value. He then points out that by failing to donate fairly small sums to poverty relief, we are allowing innocent people to die from easily preventable causes. This is an instance of the same action type as allowing the child to drown. Thus, to maintain consistency among our moral claims, we must say that donating to poverty relief is as stringent a duty as rescuing the drowning child. The difficulty in this argument, and in others like it, is in determining whether the two actions (states of affairs, etc.) fall under the same relevant type. If they do not, then there is no inconsistency.

Unlike the previous method, this method does give us grounds for criticizing some general moral claims (e.g., 'Generally, *x* are *M*'), rather than particular moral claims (e.g., 'This is *M*'). If we can show that a general moral claim is inconsistent with another moral claim, we have grounds for rejecting one of them. If, for instance, someone holds that all *x* are *M* and all *y* are not-*M*, where some *x* are *y*, we can show that

he is committed to saying, for some cases, that something is both *M* and not-*M*, thereby forcing him to revise or reject one of his general moral claims.

This method has its limits, as well. Like any method that demands only consistency, the method says nothing about which of two inconsistent moral claims to reject. Furthermore, it provides no grounds for criticizing any internally consistent moral system. This poses a potentially serious problem. When two moral claims conflict, we can resolve the inconsistency by rejecting either one. If endorsing cruelty conflicts with condemning suffering, endorsing suffering resolves the inconsistency as well as condemning cruelty does. What, critics may ask, could we say to the person who develops a consistent but seemingly perverse moral system? Great evils, after all, are often cloaked in elaborate moral justifications. As long as a given system of moral claims is consistent, this method of moral criticism is powerless. What could we say to Milton's Satan when he resolves, "Evil, be thou my good?"[2]

1.3. Nonmoral values

Appealing to nonmoral error and inconsistency will not show Milton's Satan the error of his ways. We can, however, can justifiably appeal to nonmoral values—what Prinz provocatively calls "extramoral" values—in order to evaluate one moral appearance or moral system as better than another.

The objection from coherent evil—the objection that Satan personifies—mirrors a common objection to coherence theories of nonmoral knowledge. It seems possible to develop a coherent system of beliefs (whether moral or nonmoral) that seems clearly false. Conspiracy theorists weave elaborate webs to protect their beliefs. Serial killers

[2] Blackburn frames the problem in this way in Blackburn, *Ruling Passions*. See: John Milton, *Paradise Lost* (1667), Book IV, line 110.

might do the same to justify their moral views. If we could criticize moral systems solely on the basis of internal inconsistency or dependence on nonmoral errors, then it seems that we would have no basis for claiming that our moral system is any better than one that we think is clearly mistaken.

There are two ways to interpret this objection. On the one hand, the point might be that our current moral system is no better than a coherent but "evil" system—"evil," at least, relative to the values we now profess. It is, in a sense, mere chance that we have opted for our current system. We could just as well adopt the evil system. On the other hand, the point might be that there could be someone who endorses a coherent, evil moral system, and we would have no rational means of criticizing such a person's moral claims. If the objection is interpreted in the former way, then it is easily refutable. If it is interpreted in the second way, it is probably not refutable, but once it is properly understood, it loses much of its bite.

The first interpretation of the objection from coherent evil fails because our current moral system coheres much better with our nonmoral values than a coherent but evil alternative would. Consider a brief list of things that nearly everyone values in a nonmoral sense (though some may also value these things in a moral sense): safety and security, having enough to eat, having meaningful relationships with others, having shelter and material comforts, having meaningful and enjoyable activities to engage in, and so forth. We also value the well-being of those to whom we feel attachment. Many of us directly value the social stability and well-being of our communities. As Hobbes observed long ago, our moral values—including especially norms about violence, honesty, fidelity, and justice—are crucial to satisfying those nonmoral values. A moral system that endorsed wanton cruelty, deceit, betrayal, and gross injustice could not sustain the nonmoral goods that we prize. This is a clear sense in which our moral system is better

than its inverse. Our moral system supports the nonmoral values that we hold; its inverse undermines them. We could not, therefore, just as easily endorse a coherent but evil system, for doing so would amount to sacrificing many things that we hold dear in a nonmoral sense.[3]

This intimate relationship between prohibitions against harm, deceit, etc. and the things that we consider to be nonmoral goods underwrites the philosophical conceit of a social contract. When we imagine the development of a social contract, we are imagining people with fixed nonmoral values, but no moral values, who are trying to decide which moral system to adopt. As contractualist arguments from Hobbes to Rawls show, this decision is not arbitrary, for choosing different moral values will profoundly affect the prospects for realizing the parties' nonmoral values. The parties to the contract have good reason to choose values like the ones that we now endorse. Our current system clearly promotes our actual nonmoral values more effectively than its inverse would.

A more challenging form of this objection involves more realistic alternatives to our current moral system—systems that differ in dramatic ways but are not the complete opposite of our own. We need not look to Milton to find characters who have coherent moral systems that we find unacceptable. Consider, for instance, the morality systems of the Mafia or of pirate bands. The prescriptions of Mafioso morality differ significantly from those of mainstream Western society. Violence and coercion are much more widely permitted than in mainstream morality, though there are still rules about when such practices are permissible. A society controlled, in whole or in part, by the Mafia is like a society with a corrupt and oppressive government, and the implications for our

[3] Jesse Prinz and David Wong also make this point. See Prinz, *The Emotional Construction of Morals*, Ch. 10, D. B. Wong, *Natural Moralities* (2006).

nonmoral values are much the same in both cases.[4] In the short term, Mafioso morality leads to increased insecurity and suffering. In the long term, it has the same constricting effect on economic development that a corrupt state does, reducing the well-being of future generations—possibly including the well-being of the Mafia's own members, relative to what they would enjoy had they been engaged in legitimate business. If we care about these things, then we do have a reason to choose mainstream morality over Mafioso morality.

If the objection from coherent evil, in its first interpretation, were correct, then the decision between our current (mainstream) moral system and its inverse or competitors like Mafioso morality would be arbitrary. Any moral system would be as good as any other. As the preceding arguments show, however, these choices are not arbitrary. Which moral system we choose significantly affects the likelihood of realizing our nonmoral values. Thus, this interpretation of the objection from coherent evil fails. Coherent but evil moral systems are not as good as the moral system we have now, for they would interfere with the realization of our nonmoral values.[5]

The second interpretation of the objection from coherent evil, however, is much harder to refute. We can imagine someone who does not share any of our nonmoral values—someone who finds constant personal insecurity thrilling, hunger insignificant, other people irksome, material comforts a bore, hardships an exciting challenge, and others' suffering a fascinating spectacle. For such a person, Hobbes' version of the social contract would seem perverse, for Hobbes' insistence on peace and justice would seem to

---

[4] Stergios Skaperdas, "The Political Economy of Organized Crime: Providing Protection When the State Does Not," in *Conflict and Governance*, ed. Amihai Glazer and Kai Andreas Konrad (Springer, 2003).

[5] There is, however, an important variant of the objection from coherent evil with which I have not yet dealt. It might seem that opportunistic moral systems, on which moral rules can be broken when it is profitable to do so, would realize nonmoral values even better than non-opportunistic systems. I postpone discussion of this version of the objection until Chapter Six.

drain all of the nonmoral value from life. We could also imagine a powerful egoist whose nonmoral values are so limited to his own interests and whose strength, both physical and social, is so great that he will best realize his nonmoral values through sheer domination of others. For him, the Hobbesian state of nature is perfectly acceptable, and our moral system would only interfere with, rather than support, his realization of his own values. For both of these characters, coherent evil seems the best choice.

There are three things that we might say in response to this form of the objection from coherent evil. First, such people are extremely rare, if they exist at all. Most of us *do* care about things that make our current moral system a better choice than coherent evil, and so for *us* there is no problem in saying that our system is better than theirs. Second, although this objection does undermine the possibility of giving a rational argument that should convince everyone, regardless of their interests or desires, to abide by a moral system like ours, that does not undermine the response to Ivan's insistence that "everything is permitted." There is a significant difference between admitting moral pluralism and allowing pure moral subjectivism. Third, every moral theory is open to some variation of this objection. Even a moral theory like Thomas Scanlon's, which yokes morality to rationality, cannot convince someone to be moral if he cares more for himself than for either morality or rationality. Thus, even though moral sense theory cannot show that such villainous characters rationally ought not to choose coherent evil, the theory remains useful in justifying our own moral decisions, given the concerns and desires that we actually have. Such cases persist, but they need not interfere with our ability to use the preceding methods of moral criticism to evaluate moral appearances. The proper response to such a person may not be argument; it may be coercion.

Setting such cases aside, then, appealing to nonmoral values as a way of deciding between competing moral systems provides a final method of moral criticism—one that

allows us to choose even between systems that are both internally consistent and consistent with all nonmoral beliefs. Even without putting faith in moral appearances, we need not accept the kind of nihilistic subjectivism according to which "everything is permitted." Appealing to nonmoral values provides an important part of the solution to the Karamazov Problem.

2. Practical coherence

The key to any good coherence theory of knowledge is to tie one's mental states to the world—to provide some external check on which systems one can adopt. In coherence theories of general knowledge, this check is provided by perception and action. Though we can categorize our perceptions in a myriad of ways and reject some as illusory, the world imposes some structure what we perceive. The world also determines what the consequences of our actions will be. Wedged between perception and action, our beliefs about the world have only so much room for variation.

The three preceding methods for criticizing moral systems related moral claims to non-evaluative premises in evaluative inferences, other moral claims, and nonmoral evaluative claims, respectively. But as of yet, we have not made explicit how these methods tie moral systems to the world—how the world imposes limits on which moral systems we can feasibly adopt.

The preceding methods can be unified and extended by introducing the notion of "practical coherence." Practical coherence is a feature of entire evaluative systems, including both moral and nonmoral "evaluations" (i.e., evaluative claims). Practical coherence has two aspects: "practical consonance" and "practical consilience." Both depend on the prescriptivist assumption asserted in Chapter One: all evaluations prescribe or proscribe various actions, though this prescriptive force may not exhaust

their meaning or use. To evaluate something positively is to recommend it; to evaluate something negatively is to discourage it.[6] It is therefore possible to connect any evaluation with one or more goals that are prescribed or proscribed by the evaluation. These goals figure prominently in the explication of practical coherence.

2.1. Practical consonance

Keeping in mind that each evaluation recommends a goal, we can understand practical consonance as a relation between evaluations. Two evaluations are practically consonant in a set of circumstances if, given the circumstances, achieving the goal recommended by one evaluation does not reduce the objective probability of achieving the goal recommended by the other. Contrapositively, if achieving the goal recommended by one evaluation reduces the objective probably of achieving the goal of the other, given the circumstances, then the two evaluations are "practically dissonant" in those circumstances. If the circumstances in question are our current ones, we may simply speak of evaluations being practically consonant or dissonant.

Consider, for instance, the following evaluations:

(A) Being concertmaster in the Berlin Philharmonic is good.

(B) Being a champion Olympic swimmer is good.

These two evaluations are highly practically dissonant. The first recommends a goal of being the concertmaster of the greatest symphony orchestra in the world. The second recommends a goal of becoming a world-class swimmer. Achieving either of these goals requires a tremendous amount of training and dedication from a very young

[6] See Chapter 1, §1.

age. Investing the time and effort to accomplish one dramatically reduces the likelihood of accomplishing the other.

Notice, however, that it is not *logically* impossible for one person to accomplish both goals. Although becoming concertmaster in Berlin makes it extremely unlikely that one will swim to Olympic gold, there is some minute possibility that someone could achieve both goals. Thus, unlike logical consistency but like empirical confirmation, practical consonance is a matter of degree. Achieving one goal can reduce the probability of achieving a second goal dramatically without reducing it to zero. Furthermore, it might reduce the probability of achieving a third goal only marginally. In that case, the first goal would be highly practically dissonant with the second, but only slightly practically dissonant with the third. In some cases, two claims might be *perfectly* practically dissonant: achieving one goal might make it impossible, given the circumstances, to achieve the other. We might call such claims "practically inconsistent." (Claims recommending logically inconsistent goals would, of course, fall into this category.)

Notice also that sufficient changes in circumstances might render otherwise practically dissonant evaluations consonant. We can imagine fantastic circumstances in which becoming concertmaster in Berlin does not reduce one's chances of becoming a champion Olympic swimmer: a child who has been genetically modified to maintain peak physical condition until the age of 50 might have time to train sufficiently both as an Olympic swimmer and as a violinist. But because the definition of practical consonance restricts the relevant circumstances to the present, such fantastic cases cannot show that evaluations like (A) and (B) are practically consonant.

Practical consonance must be assessed in terms of the current circumstances because the consonance of any given system of evaluations would be indeterminate if it

were not yoked to some set of circumstances, and only yoking it to current circumstances will yield a measure that is useful for current action. If moral epistemology is to help *us* decide what to do, then it must help us decide what to do *in the circumstances in which we find ourselves*, for those are the only circumstances in which we can act.

It is also worth emphasizing that our circumstances include the beliefs, values, dispositions, and actions of those around us. It is easy enough in modern America to keep body and soul together while espousing unorthodox religious views; but 'Being executed is bad' and 'Publicly denying the divinity of Jesus Christ is the right thing to do' were highly practically dissonant in Madrid at the height of the Inquisition. Their dissonance in those circumstances depends not on laws of nature, but on the beliefs, actions, and moral systems of early modern Spaniards.

Keeping these points in mind, we can also extend the notion of practical consonance to an entire evaluative system:

> (CONSONANCE) An evaluative system is practically consonant if its constituent evaluations are pairwise practically consonant—i.e., to the extent that achieving the goal recommended by any one evaluation does not reduce the objective likelihood of achieving the goal recommended by any another evaluation, given the current circumstances.

It may be that no one's evaluative system is perfectly practically consonant. But some are more practically consonant than others, and this is all we need to solve the Karamazov Problem.

2.3. Practical consilience

As many critics of coherence theories point out, however, consonance is cheap. A system of completely unrelated evaluations—that is, of evaluations that recommend goals that are completely orthogonal to one another—is perfectly consonant, as is a system of exactly one evaluation. The more interesting aspect of coherence is the positive side: in a coherent system of claims, the claims mutually support one another.[7] In the case of practical coherence, we might call this phenomenon "practical consilience."[8]

Two evaluations are practically consilient to the extent that, given the current circumstances, achieving the goal recommended by one increases the objective probability of increasing the goal recommended by the other. Consider, for instance, these evaluations:

(A) Being concertmaster in the Berlin Philharmonic is good.

(C) Practicing violin is good.

These two evaluations are practically consilient. Becoming concertmaster in Berlin provides both opportunity and motivation to practice violin; practicing makes a violinist more likely (though still not very likely) to become concertmaster in Berlin.

As with practical consonance, practical consilience is a matter of degree. Consider the following pair of evaluations:

---

[7] Laurence Bonjour, *The Structure of Empirical Knowledge* (Cambridge, MA: Harvard University Press, 1985).

[8] William Whewell coined the term "consilience" (literally, "jumping together") to describe the way in which scientific discoveries about one subject support and are supported by scientific discoveries about another subject. Practical consilience is named by analogy to Whewell's consilience, but it is not presumed to have all of the same characteristics. See: William Whewell, *Novum Organon Renovatum*, Third ed. (London: John W. Parker and Son, 1858), 70; Laura J Snyder, "Consilience, Confirmation, and Realism," in *Scientific Evidence*, ed. Peter Achinstein (Baltimore: Johns Hopkins University Press, 2005).

(C) Practicing violin is good.

(D) Enjoying practicing violin is good.

These two evaluations are significantly more practically consilient than (A) and (C). For although both being concertmaster in Berlin and enjoying practicing will increase the likelihood of practicing, practicing is far more likely to lead to enjoyment of practicing than to becoming concertmaster in Berlin. Furthermore, some evaluations will recommend logically equivalent or coextensive goals, so that achieving one involves achieving the other. Such evaluations are "perfectly practically consilient."

Again, as with practical consonance, we can imagine circumstances in which these two evaluations are not practically consilient. If positions in the Berlin Philharmonic were granted on a purely political basis, without regard to musical skill or interest, practicing would not make one more likely to win the principal chair. But these kinds of examples are as irrelevant to our present decisions as they were in the case of practical consonance. Practical consilience, like practical consonance, is relative to a specified set of circumstances (which include the beliefs, attitudes, and actions of other people).

Keeping these points in mind, we can also extend the notion of practical consilience to an entire evaluative system:

(CONSILIENCE) An evaluative system is practically consilient to the extent its constituent evaluations are pairwise practically consilient—i.e., to the extent that achieving the goal recommended by any one of its evaluations

> increases the objective likelihood of achieving the goals recommended by any other constituent evaluation, given the current circumstances.

As with practical consonance, it is unlikely that anyone's evaluative system is perfectly practically consilient. In fact, the need for an epistemology of evaluation arises precisely because our goals are not consilient. We must choose between competing ends when we cannot achieve both of them. But some evaluative systems are more practically consilient than others, and we will encounter less frustration, less disappointment, and more satisfaction as the consilience of our evaluative systems increases.

2.3. Practical coherence

Practical coherence is the combination of practical consilience and practical consonance. An evaluative system is practically coherent to the extent that its constituent evaluations are practically consilient and practically consonant.[9]

This is the point at which the world imposes itself on an otherwise coherentist theory of moral knowledge. The extent to which two evaluations are practically consilient or dissonant is not up to us. It depends on the way the world is. If humans cannot function without adequate food and sleep, then systems containing evaluations that too severely restrict our ability to eat and sleep will be practically dissonant. If human cooperation, which is necessary for the achievement of many of our goals, depends on trust, and as a matter of psychological fact, trust erodes quickly in the face of

[9] A perfectly practically consilient system is necessarily a practically consonant system, and so the latter condition may seem redundant. But in our actual evaluative systems, some evaluation, E_1, may be consilient with evaluation E_2 and dissonant with E_3. *Ceteris paribus*, such a system would be less practically coherent than a similar system in which E_1 is consilient with E_2 and irrelevant to E_3.

deceit or maleficence, then (*ceteris paribus*) systems praising deceit and maleficence will be more dissonant than those that do not. If our technological progress is such that we can satisfy our goals of material comfort only through hard work, then systems that value industriousness will be tend to be more consilient than those that do not. To the extent that there is some relatively stable "human nature," that nature will affect the practical coherence of our evaluative systems.[10] These objective features of the world impose a partial ordering on evaluative systems in terms of their practical coherence. Practical coherence thus provides a solution to the Karamazov Problem in which the world, not just our whim, determines which evaluative systems to prefer.

3. Practical coherence as a method of ethical inquiry

Practical coherence is worth adopting as a standard for evaluative systems only if it allows us to draw conclusions about which values to hold—only if it helps make ethical inquiry more productive. The details of such ethical inquiry, of course, need to be worked out—most likely through a long process of trial, error, and improvement. The scientific method did not spring, fully armed, from the head of some early modern thinker. It developed slowly, in fits and starts. There is no reason to expect any more from a method of ethical inquiry. We can, however, make some predictions about the use of practical coherence in ethical inquiry.

In principle, practical coherence constitutes a criterion by which to choose between evaluative systems from outside of those systems. If could develop a measure of practical coherence, it would provide an objective standard by which to choose one system over another. In practice, however, it is probably impossible to determine the practical coherence of entire evaluative systems with enough precision to compare them.

[10] For a plausible recent exposition of such a view, see Wong, *Natural Moralities*.

In practice, it will almost certainly be more fruitful to use practical coherence as a standard for evaluating proposed revisions to the evaluative systems that we now hold. Much as scientific experiments seek to confirm specific revisions to our view of the world, rather than attempting to justify an entire worldview all at once, we might consider proposed revisions to our evaluative system on a case by case basis. By providing an objective standard against which to measure alternative evaluative systems, practical coherence allows us to do applied ethics without relying on moral appearances.

It might be possible to develop helpful formal models of this process, akin to formal models of belief revision. In the dominant model of belief revision, known as the AGM model,[11] belief systems are represented as logically closed sets of sentences. The revision process is modeled through three kinds of operations: contraction, in which a belief is removed from the system; expansion, in which a belief is added to the system; and revision, in which a sentence is added and others are removed (for consistency's sake). Each operation requires changes to the belief system to accommodate the new beliefs. A formal model of evaluative revision would need to capture the effects of the removal, addition, or revision of evaluations on the probability of achieving the goals advocated by each evaluation in the revised system. As long as the model enables us to tell whether a proposed revision increases or decreases the practical coherence of the system, it would be useful in ethical inquiry.

3.1. Practical coherence and Dewey's method of ethical inquiry

The real power of practical coherence lies in the kind of pragmatist ethical inquiry advocated by John Dewey. Dewey suggested that the need for ethical inquiry arises in

[11] The AGM model was introduced in C. E. Alchourron, "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions," *The Journal of symbolic logic* 50, no. 2 (1985).

particular situations in which one's values conflict.[12] In other words, ethical inquiry begins in practical dissonance. The goal of ethical inquiry is to eliminate that practical dissonance. In some cases, this is a purely technical matter. Practical problems prevent one from achieving goals recommended by two conflicting evaluations, but if one can overcome these problems through cooperation, creativity, or better technology, one can eliminate practical dissonance without revising one's evaluations.

In other cases, however, one can overcome practical dissonance only through the revision of one's evaluative system. In such cases, the pressing question is which evaluations to change, and how to revise them.[13] Practical coherence offers a rubric by which to make those decisions. If two evaluations recommend incompatible goals, determine the effects of abandoning each goal on the likelihood of achieving the remaining goals recommended by one's evaluative system. Determine the effects of adopting slightly different goals by revising one or both of the conflicting evaluations. Identify the revisions that lead to the most practically coherent evaluative system and adopt them.

This process is undoubtedly difficult, and it cannot be done from the armchair, but it is not impossible. It is essentially the process of policy analysis—determining which goals will be met, which will not, and what side effects will occur, if one set of goals is adopted rather than another. Policy analysis requires a thorough command of relevant empirical subjects, sophisticated statistical techniques, and in especially complicated circumstances involving multiple actors, game theory. It is a complex, multidisciplinary undertaking.[14] If we understand ethical inquiry as the quest for

---

[12] Dewey and Tufts, *Ethics*, 207, John Dewey, *The Middle Works, 1899-1924*, vol. 20 (Carbondale: Southern Illinois University Press, 1982), 163–64.
[13] Dewey, *The Middle Works, 1899-1924*, 163–64.
[14] In their manifesto for "Duke naturalism," Owen Flanagan, David Wong, and Hagop Sarkissian propose a similar view of ethical inquiry, also inspired by Dewey. They liken

practical coherence, ethical inquiry becomes a multidisciplinary investigation into the effects of having the values that we do versus having some other values.

For example, consider an evaluative system that permits the consumption of meat in industrialized societies but discourages actions that exacerbate current environmental problems. The production of meat in the United States, at least, contributes to several major environmental problems. Cattle, in particular, emit large quantities of methane, which is a potent greenhouse gas. Under current agricultural practices, the use of synthetic fertilizer in growing corn, on which commercial livestock is largely fed, contributes to chemical pollution of waterways. Synthetic fertilizers depend on fossil fuels for their production. Thus, a system that condones normal patterns of meat eating but condemns environmental degradation is, in the current circumstances, practically dissonant.

There are a number of revisions that might resolve this dissonance. Becoming a vegetarian would greatly reduce one's support of animal husbandry. One could also abandon one's concern for environmental degradation. Other alternatives include significantly reducing one's meat consumption or restricting one's meat consumption to meat raised in specific ways. But all of these revisions in one's evaluative system have ramifications. Some are economic, some environmental, some nutritional, some social, and so forth. Will one individual's abstention from meat have any impact on environmental degradation? How would restricting oneself to meat raised in environmentally responsible ways affect environmental degradation? What are the social and economic consequences of that environmental degradation? Discovering the ramifications of each choice require serious, multidisciplinary study.

---

ethical inquiry to ecology, another multidisciplinary subject that deals with complex interactions within large systems. Both analogies are appropriate. See Owen J Flanagan, Hagop Sarkissian, and David Wong, "Naturalizing Ethics," in *Moral Psychology*, ed. Walter Sinnott-Armstrong (Cambridge, MA: The MIT Press, 2007).

The result of this inquiry will be a solution that is tied to a particular time and place, in which specific agricultural, economic, and social conditions prevail and particular environmental problems loom large. The inquiry will not yield a general rule, applicable to all people in all situations, about vegetarianism or environmentalism. But it will help people in those particular circumstances resolve a specific problem in their lives.

As the process of ethical inquiry is repeated in particular cases of practical dissonance, one's evaluative system will become increasingly practically coherent. Since a practically coherent evaluative system recommends goals such that achieving one increases the likelihood of achieving others, this means that more of one's actions will satisfy more of one's goals. The result is a more satisfying, flourishing life overall. This is precisely what Dewey advocated as the purpose of ethical inquiry.

3.2. Methodological practical coherentism

This view of ethical inquiry can be encapsulated in a thesis that I call "methodological practical coherentism":

> (MPC) *Ceteris paribus*, given two evaluative systems, one should adopt the more practically coherent system.

It is important to recognize that (MPC) is a purely methodological claim about the epistemology of ethics. It is not an attempt to define the moral or ethical in terms of practical coherence; it is a recommendation about how to revise our evaluative schemes. Just as the scientific method recommends a strategy for revising beliefs about the world

without defining Truth in terms of hypotheses and confirmation, so (MPC) recommends an epistemological strategy without defining Goodness in terms of the method.

The *ceteris paribus* clause serves primarily to capture one important limitation on (MPC). Some evaluations are all but beyond our power to reject. For instance, few people could sincerely hold that pain is usually good or that sleep or social interaction is bad. These stubborn evaluations are to an evaluative system as Quine's "recalcitrant experiences" are to the web of belief.[15] Furthermore, these evaluations have significant implications for the other evaluations in one's system. In order to maintain a practically coherent evaluative system, the person who values social interaction will need to value a constellation of other behaviors, character traits, and states of affairs that facilitate positive social interaction. Thus, while it may be possible to construct a highly practically coherent evaluative system by eliminating most evaluations from it, this would require eliminating our positive evaluations of basic human needs. Because we cannot, in practice, give up those evaluations, such a system is not preferable to the system we have now.

4. Some consequences of practical coherentism

Conceiving of ethics as the pursuit of practical coherence has two important, though seemingly contradictory, consequences. On the one hand, it almost certainly results in some degree of relativism. On the other hand, it leads to a limited kind of moral realism, understood as the claim that ethical truths are partly "stance independent."

It seems very unlikely (though it is logically possible) that there is exactly one maximally practically coherent evaluative system. Given any set of starting points, there

[15] W. V. Quine, "Two Dogmas of Empiricism," *The Philosophical Review* 60, no. 2 (1951): 40.

are probably multiple, distinct, complete evaluative systems that are equally practically coherent. And given that people actually have very diverse values, and thus a diverse set of starting points, the number of maximally coherent evaluative systems that people could justifiably endorse increases dramatically. Furthermore, given that people rarely, if ever, formulate a *complete* evaluative system—that is, a system that contains an evaluation of every situation—there are likely to be an even greater number of *partial* evaluative systems that are equally practically coherent.[16] With no way to choose between these systems, the theory entails that the same thing may be wrong for one person but right for another—and that what *makes* it wrong for the first but right for the second is, ultimately, the fact that each person values different things. In this respect, the theory probably entails relativism, even after maximizing practical coherence. This will surely provoke a range of objections, from semantic and logical objections to ethical ones. I will address these objections in the next chapter.

Surprisingly, the theory also yields a limited form of moral realism. For any particular person and any particular evaluation, whether that claim is true depends largely on the relations between the evaluation, other evaluations the person endorses, and the way the world is. These relations do not depend on what the person thinks about them. This means that a person could very easily be wrong in endorsing a particular evaluation—and not because she misunderstands the nonmoral facts, but because her evaluation does not practically cohere with the rest of her evaluative system. Thus, the person who values security and contentment but thinks that recreational violence is just fine has made a serious evaluative error, regardless of whether she recognizes the practical incoherence among her values. Furthermore, given the relative difficulty of

---

[16] *Cf.* Allan Gibbard's notion of a "hyperplan" or Catherine Wilson's notion of "theory-level relativism." See Allan Gibbard, *Thinking How to Live* (Cambridge, MA: Harvard University Press, 2003), 54; Catherine Wilson, *Moral Animals: Ideals and Constraints in Moral Theory* (Oxford: Oxford University Press, 2004), 62–64.

changing one's preferences for security and contentment, the most effective way to eliminate that practical incoherence is to abandon the positive evaluation of violence.[17] Recreational violence is wrong for her, even if she sincerely disagrees. Thus, although what is valuable for a person is a function of what she values, it is not determined solely by what she values. Evaluative truths, on this view, therefore meet at least a weak condition for objectivity: something can be good without anyone valuing it, and there could be something that everyone values that is not valuable.[18]

5. Some puzzles about practical coherence

There is, of course, a great deal still to be said about practical coherence. There are a number of objections that ethicists are likely to raise. I will address those in the next chapter. But there are four puzzles about practice coherence that can be addressed more easily, if only by pleading ignorance.

First, one might wonder how one's current goals relate to one's future goals. If achieving one's present goals makes it harder to achieve future goals, which goals should one abandon? The answer, it seems, will depend significantly on what the goals are. But in general, it would probably be wise to strive for consilience between current and future goals, and to strike a balance between them when this is impossible. People who, in their youth, value goals that both fulfill current desires and pave the way for future success have, over the course of their lives, more consilient evaluative systems than either those who devote their youths to pleasurable though unedifying pursuits or those who spend their youth working so hard for future goals that they miss the pleasures of youth

---

[17] It may always be possible, of course, to "save the appearance" that violence is okay by making sufficiently drastic alterations elsewhere in one's evaluative system.
[18] *Cf.* Robert Brandom, *Making It Explicit: Reasoning, Representing, and Discursive Commitment* (Cambridge, MA: Harvard University Press, 1994), 601–07.

altogether. Human beings typically lead relatively long, varied lives. What we do in our youth—and what others in our society do over the many decades that we live—will significantly affect what we can do later in life. It is not easy to know how exactly this longevity and variation impacts decisions about how to live. Thus, developing a more detailed response to this puzzle requires serious ethical inquiry.

Second, one might wonder how the relative utility of achieving various goals figures into decisions about which goals to adopt. It seems, *prima facie*, that a system prescribing many consilient but unrewarding goals would be more coherent overall than one that recommended goals that were all highly rewarding but difficult to achieve at once. For instance, achieving success in a rewarding career, maintaining an active social life, having engaging hobbies, and being part of a caring family are all rewarding activities for most people; but achieving any one of them—much of less all of them—can be difficult. It may be more practically consilient to aim for a tolerable job, a few decent friends, and a family that causes little heartache. To twist Mill's phrase, it may be more consilient to be a fool satisfied than Socrates dissastisfied. Whether this is true, however, is a serious ethical question. It is a central tenet of Buddhism that the best life is one that is free from desire—and so necessarily free from the satisfaction of desire. Though Western philosophy rarely acknowledges the possibility, perhaps the simple life truly is the better life. Practical coherence countenances the revision of any evaluation—even the claim that pleasure (understood as a subjective feeling) is good.

That said, so long as we value things like pleasure, evaluative systems recommending goals that bring great pleasure will be more consilient than those that do not. Thus, *ceteris paribus*, when evaluations conflict, our evaluative systems will be more consilient if we discard the evaluation recommending the less pleasurable goal.

Comparative utility, then, does figure prominently into practical coherence as long as we value pleasure.

Third, a puzzle arises from the use of objective probabilities, rather than subjective probabilities, in characterizing practical coherence. Because we sometimes err in judging objective probabilities, we will sometimes judge two evaluations to be consilient or dissonant when they are not. Suppose that, in such a case, we wrongly judge some action, *A*, to be right. There is one sense in which we ought to do *A*, since to do otherwise would be to do something that we believed to be wrong, but another sense in which we ought not to do *A*, since it would diminish the likelihood of achieving our other goals. This kind of tension between "subjective duty" and "objective duty" has been much discussed by ethicists.[19] Nothing about (MPC) requires that one type of duty take priority over the other; it is enough to note that (MPC) entails that a divergence between subjective and objective probabilities would give rise to conflicting subjective and objective duties.

Finally, one might wonder what the relation is between practical coherence and the psychological mechanisms by which we form moral appearances. A well-founded answer to this question must await better knowledge about what kinds of evaluative systems are practically coherent. But it seems reasonable to suppose that evaluative systems will typically contain at least some traditional moral values—values like honesty, integrity, generosity, and courage. It might be that we evolved the psychological mechanisms we did precisely because those mechanisms instill attraction to practically coherent values and aversion to practically dissonant values. Such evolved attractions and aversions would prove useful in guiding our behavior even when we do not understand how particular actions conduce to the overall satisfaction of our desires.

---

[19] See, e.g., W. D. Ross, *Foundations of Ethics* (Oxford: Clarendon Press, 1939); H. A. Prichard, *Moral Obligation* (Oxford: Clarendon Press, 1949).

Moral appearances may, in some cases, lead us to practical coherence. And if we could verify that they do, we could rely on them as signs of practical coherence and therefore use them to justify moral claims. But their epistemic credentials would be underwritten by antecedent moral knowledge, which must initially be derived without their help.

6. Conclusion

After hearing for the first time of Ivan Karamazov's theory that everything is permitted if there is no afterlife, the divinity student Rakitin tells Ivan's brother that the theory is absurd. "Humanity," he says, "will find in itself the power to live for virtue even without believing in immortality. In the love of freedom, of equality and the brotherhood of man, it will find it."[20] In other words, we should look for the basis of morality not in some external reality independent of human interests, but in our pursuit of our own values.

When we come to doubt moral appearances, it may seem that we have undermined any possibility of moral epistemology, just as it seems to Ivan that denying an afterlife undermines the justification of moral claims. It may seem that without moral appearances, we have no way to judge one moral claim or moral system any better than any other.

By introducing the notion of practical coherence, we tie our moral systems to one another, to our nonmoral evaluations, and to the relations that the world imposes on those evaluations. An evaluative system is practically coherent to the extent that its constituent evaluations are practically consilient and practically consonant—i.e., to the extent that achieving the goals prescribed by each constituent evaluation increases, or at least does not decrease, the likelihood of achieving the goals prescribed by the other constituent evaluations, given the current circumstances. This yields a stance-

[20] Dostoyevsky, *The Brothers Karamazov*, 105.

independent, useable criterion by which to judge some evaluative systems as better than others, which is encapsulated in the principle of methodological practical coherentism. The evaluative systems endorsed by (MPC) are better than their rivals in the sense that adopting one rather than another leads to a more unified, satisfying, flourishing human life, as defined from within that system itself.

This pragmatist conception of ethical inquiry as the pursuit of practical coherence conflicts with some of the deepest convictions of Western ethical thought. It will therefore provoke a number of objections, to which I now turn.

6 Practical Coherence and Relativism

Medieval morality plays had stock villains. Lust, Greed, Sloth, and the rest of the Seven Deadly Sins featured prominently. Moral philosophy has its stock villains, too. The Amoralist, the Fool, the Akratic, and various others appear repeatedly. Few, however, attract more hostility and disdain than the Relativist. The Relativist, like the Vice character in morality plays, is an all-purpose villain. His sins are not only ethical, but also logical and semantic. Not only is he thought to endorse the most vile actions, but he is said to be inconsistent or even nonsensical in maintaining that one act can be right for one person and wrong for another.

Endorsing methodological practical coherentism (MPC), as defined in the previous chapter, almost certainly commits one to ethical relativism. To conceive of ethics as the search for greater practical coherence is therefore to invite a host of objections. The most important objections to the view can all be seen as versions of common objections to relativism. Fortunately, once we clarify the kind of relativism entailed by (MPC), these objections can all be met.

1. Methodological practical coherentism and ethical relativism

Methodological practice coherentism provides a method for revising evaluative systems in order to make them better, according to a nonmoral and thus non-question-begging criterion. It places no *a priori* restrictions on the content of the evaluations that may be admitted in a system, and it allows that different systems will contain different evaluations. It is entirely possible, if not probable, that (MPC) will therefore fail to yield a

single maximally coherent evaluative system. This means that (MPC) is likely to lead to a kind of ethical relativism. But in order to understand how (MPC) handles the objections usually raised against relativism, it is important to understand in more detail what kind of relativism it entails and how that entailment works.

1.1. How methodological practical coherentism leads to relativism

The primary reason to think that (MPC) will lead to relativism is the phenomenon of "evaluative ambivalence," or the underdetermination of value by the demands of practical coherence. Evaluative ambivalence arises when we discover practical dissonance within any evaluative system but recognize that there are mutually exclusive, though equally effective, options for eliminating that practical dissonance. That is, we may find that eliminating any one of several evaluations improves a system's practical coherence to the same extent. (MPC) provides no guidance in choosing between those options. From the perspective of practical coherence, we face an existential choice. If we are to improve our evaluative system, we must choose one of those options, but in a situation of genuine evaluative ambivalence, there are no grounds for choosing one rather than another.

Evaluative ambivalence is a more general form of the "moral ambivalence" that underwrites David Wong's theory of moral relativism.[1] Moral ambivalence, according to Wong, arises when we discover tensions between two compelling moral values but recognize that "reasonable people" might resolve this tension in different ways.[2] Consider, for instance, the tension between individual autonomy and the common good.

---

[1] Wong, *Natural Moralities* (New York: Oxford University Press, 2006), xiv, 20–28.
[2] Wong appeals readily to "reasonable people," "human nature," and "basic moral values." Though I try to avoid these ideas in my account of evaluative ambivalence, my account nonetheless parallels and is indebted to his account of moral ambivalence.

Improving the lot of the worst off or defending common goods sometimes demands restrictions on personal liberty. Different people, or different cultures, might reasonably balance these two demands in somewhat different ways. Philippa Foot and Thomas Scanlon, as well as Wong, endorse just this ambivalence about the precise balance between individual autonomy and the common good.[3]

Wong argues that moral ambivalence leads to moral relativism because groups or individuals that resolve these moral tensions in different ways will end up with different moral systems. If each point of disagreement between two moral systems can be traced to an issue of moral ambivalence, and each system's resolution of that ambivalence is reasonable, then there are no rational grounds for taking one of those systems to be superior to the other. This does not mean that "anything goes," however. If some disagreement between the two systems can be traced to a mistake in fact or inference in one system, or to an resolution of moral ambivalence that sacrifices far more of the common good than is necessary to secure certain personal liberties, then the system infected by that nonmoral error or unnecessarily drastic revision can be judged inferior (at least in that respect) to the other.

Similarly, evaluative ambivalence leads to multiple, equally defensible evaluative systems. This can be illustrated by the case of career choices. A law student choosing between corporate and public interest law might find the two career tracks equally attractive, though for very different reasons. The former might demand perseverance, but the latter might require greater patience and moral courage; the former would be more lucrative, while the latter might be more rewarding in other ways; the former might allow her to donate far more money to worthy causes, though the latter might enable her

---

[3] T. M. Scanlon, "Fear of Relativism," in *Virtues and Reasons*, ed. Rosalind Hursthouse, Gavin Lawrence, and Warren Quinn (Oxford: Oxford University Press, 1995); Philippa Foot, "Moral Relativism," Lindley Lecture (University of Kansas, 1978).

to work directly for those worthy causes; etc. If the lawyer's life would be equally practically coherent in both cases, either choice would be justified.

The same reasoning probably applies to cases of practical dissonance arising from values that we now consider "moral" values. There may be no difference in practical coherence between several different ways of balancing our concern for human suffering with our concern for individual autonomy, given that, in our current circumstance, redistributing wealth drastically enough to end extreme poverty requires trampling individuals' autonomous control of their wealth. If there are multiple solutions to the problem of poverty that are equally practically coherent, then (MPC) entails that there no grounds for choosing one of those over the others—and thus no grounds for criticizing those who choose one of the other equally good solutions.

The result of evaluative ambivalence, given (MPC), is ethical relativism: For any given case of evaluative ambivalence, different people may improve the practical coherence of their evaluative systems in different ways. And those who choose one way will therefore be justified in making different evaluative claims that those who choose a different way, though both sides may have to admit that the other side's views are as well justified as their own.

2. Objections based on ethical relativism

The objections prompted by (MPC)'s relativist implications can be divided into two kinds. Some objections are ethical objections, concerned with ethical problems created by relativism. The rest are technical objections, involving logical or semantic problems that are thought to infect relativist theories of ethics.

It will simplify the discussion of these objections to use words like 'moral', 'nonmoral', 'immoral', 'good', 'evil', 'right', 'wrong', etc. to denote values, actions,

character traits or other things that are typically considered "moral," "good," etc. In other words, throughout this chapter, phrases like "morally good actions" should be read as "actions that would typically be considered morally good." Thus, saying that it is "wrong" to torture kittens is not to endorse the claim that it is wrong to torture kittens, but only that kitten-torture is typically thought to be wrong. (Of course, this is not to deny that kitten-torture is wrong, either.)

2.1. The objection from moral "perversion"

According to Wong, opponents of ethical relativism sometimes suggest that the theory is "but a few slippery steps short of collaboration with the Nazis."[4] Relativism is often thought to lead to morally "perverse" views—that is, to views that permit morally abhorrent behaviors like recreational killing, rape, slavery, and so forth. This is similar to the "objection from coherent evil" discussed in the last chapter,[5] but we can now state it more precisely, as it relates to (MPC):

(1) (MPC) leads to a form of relativism on which morally perverse acts (such as recreational killing, rape, slavery, etc.) are permissible for some people.
(2) Morally perverse acts are morally impermissible for all people.
∴ (3) We should reject (MPC), since it leads us to false evaluative claims.

Setting aside the concern that the second premise may beg the question against (MPC) and relativism in general, the objection fails because the first premise is almost

[4] Wong, *Natural Moralities*, xi.
[5] See Chapter 5, §1.3.

certainly false. Contrary to initial appearances, the form of relativism entailed by (MPC) does not permit acts like those named in the objection.

It is understandable that critics would think that (MPC) would permit such "perverse" evaluative systems. After all, traditional moral values are often conceived as limits on our pursuit of nonmoral values. It sometimes seems that we could achieve other goals more effectively if only we would abandon moral constraints. So, it seems plausible that an evaluative system that rejected traditional moral stances on topics like killing, coercion, manipulation, and reciprocity would more practically coherent than one that retained those traditional stances.

However, the history of ethics features several attempts to show that the best way to realize nonmoral values is through the realization of moral values. These arguments are, in effect, rule-consequentialist arguments, according to which the pursuit of traditional moral values is necessary to the realization of traditionally nonmoral values.[6] If the realization of nonmoral values requires the pursuit of moral values, then a morally perverse evaluative system, which prizes nonmoral values but not moral values, will be less practically coherent than one that prizes both. Thus, these rule-consequentialist arguments imply that (MPC) will not license systems that sacrifice moral values to nonmoral values. All of these attempts, however, suffer from a similar problem and therefore need to be supplemented with more contemporary arguments.

The most famous such attempt is Hobbes', though other social contract theorists follow roughly the same approach. Hobbes is the most explicit about the nonmoral importance of morality. He maintains that morality rests on a set of rules, discoverable by reason, whose purpose is to avoid the wretched "war of all against all" that prevails in

---

[6] Not all of the philosophers who attempt to ground moral values in nonmoral values technically count as rule-consequentialists, but the label captures an important aspect of the particular arguments to which I refer.

the absence of these values. From the basic precept that people ought to abide by whatever maxims are necessary to secure civil tranquility, Hobbes derives a number of paradigmatic moral values, including a prohibition against "robberies and violence," a demand for equal treatment of persons, etc. Unless such values are widely adopted, Hobbes argues, there will be no industry, no commerce, no material comforts, no intellectual activity, and no personal security.[7]

Utilitarians are also explicit in grounding moral values in the nonmoral value of pleasure. Many offer rule-utilitarian arguments to the effect that various moral values are essential to maximizing pleasure. Hume, for instance, argues that justice—particularly in the sense of security in property rights—provides "infinite advantages" by ensuring that "every one knows what he may safely possess; and the passions are restrain'd in their partial and contradictory motions."[8] Mill observes that the fleeting nonmoral benefit of telling a lie is dwarfed by the "transcendent expediency" of honesty.[9] Thus, like the social contract theorists, the utilitarians show that prioritizing moral values over nonmoral values is generally conducive to practical coherence

However, those who make such arguments are generally aware of the gulf between a proof that morality conduces to the general happiness and an argument that being moral conduces to one's own happiness. As Sidgwick puts it, "the inseparable connexion between Utilitarian Duty and the greatest [nonmoral] happiness of the *individual* who conforms to it cannot be satisfactorily demonstrated."[10] That is, while it is fairly easy to see how moral rules conduce to the general happiness, it is much harder—perhaps impossible—to show that each individual benefits by following moral rules.

---

[7] Thomas Hobbes, *Leviathan* (1651), I.xiii–xv.
[8] Hume, *A Treatise of Human Nature*, 3.2.2.9.
[9] Hudson, *Ethical Intuitionism*, Ch. 2.
[10] Sidgwick, *The Methods of Ethics*, Book IV, Concluding Chapter, §4, emphasis added.

Hobbes famously observes that some "Fools" deny that his arguments rationally bind them to pursue traditional moral values under all circumstances. But The Fool recognizes that he can sometimes break the rules without incurring any cost. Hobbes replies that the Fool is irrational to count on his peers' inability to discover and punish his wrongdoing, but many commentators are rightly skeptical of this response. Some individuals can and do break traditional moral rules on occasion without suffering any punishment. Since it seems that these individuals can realize their nonmoral values without fully realizing their moral values, they can safely jettison moral values from their evaluative systems. While they may still observe moral rules in most cases, this is for purely prudential reasons, not because they prize moral values themselves.

The Hobbesian Fool therefore represents the strongest version of the objection from moral perversion. The Fool, in his most dangerous form, is a person who generally acts in accordance with moral rules, but will break them when opportunities arise. Usually, such transgressions will be relatively minor, such as telling a lie to gain some small advantage, or stealing small quantities of stationery from an employer for personal use. These, after all, are the opportunities that arise most frequently. The chance for more serious transgressions, such as grand larceny or murder, arises much less often and carries far more substantial risks. The prudent Fool will break such stringent rules only in very special circumstances.

Hobbes' response to this kind of "prudent Fool" is inadequate. Note, however, that Hobbes' response seems to turn entirely on the question of whether the Fool himself can reasonably expect to escape punishment. The Fool assumes that either no one will pin his crime on him or no one will punish him for it. This is the assumption that Hobbes unconvincingly attacks. But there is another, less defensible, assumption behind the Fool's position. The Fool assumes that his actions will not indirectly interfere with

the goods that he gains from others' widespread observation of moral rules, and recent psychological research suggests that this may be false.

Contemporary social psychological research reveals that people are extremely sensitive to general levels of compliance with social norms, even if particular violators of those norms cannot be identified. That is, when people are aware that others are violating a norm, they become significantly less likely to follow those norms. It is much easier to recognize that someone is violating a norm than it is to know who violated it, since there is usually evidence of the violation. Thus, it is possible for someone's anonymous norm violations to cause a decline in general observance of that norm, even if the original transgressor is never identified.

The most important research on this topic is Cristina Bicchieri's.[11] For instance, Bicchieri and her colleague Erte Xiao studied participants engaged in the "Dictator game." The game involves two participants: a "dictator" and a "receiver." The experimenters give the dictator $10, and the dictator decides how much of that money, if any, to give to the receiver. Then the game is over, and each player keeps whatever money they have. For example, if the dictator decides the give $4 to the receiver, the receiver walks away with $4 and the dictator keeps $6. Bicchieri and Xiao found that by artificially manipulating participants' beliefs about how players in another study had acted and what other people thought that dictators ought to do, they could manipulate how much money dictators gave to receivers. In particular, dictators were significantly more likely to offer $4 or $5 to receivers if they thought that most other dictators behaved that way than if they thought that most dictators gave only $1 or $2.[12] More generally, Bicchieri

---

[11] C. Bicchieri, *The Grammar of Society: The Nature and Dynamics of Social Norms* (Cambridge: Cambridge University Press, 2006).
[12] Cristina Bicchieri and Erte Xiao, "Do the Right Thing: But Only If Others Do." *Journal of Behavioral Decision Making* 22, no. 2 (2007): 191–208.

argues that there is "plenty of evidence that most people are *conditional cooperators:* They cooperate when they expect others to cooperate," but not otherwise.[13]

This "conditional cooperation" has important implications for the prudent Fool. Even if the Fool is never personally punished for violating one of the moral rules required for his society to flourish, the fact that someone violated an important norm will usually be noticed, and this violation will contribute to the erosion of the norm itself. Since even the Fool depends on general compliance with the norm, this indirectly interferes with the Fool's realization of other nonmoral values. And since we can rarely be confident that our own transgression in some particular circumstance will not be part of a downward spiral into general noncompliance within our community, it is almost always better to observe important moral values than not. The results of violating moral rules may not be catastrophic. The Fool may not be cast into the Hobbesian wilderness, but he may damage the peace or prosperity of his community. Given that humans are relatively long-lived creatures, and that even slight changes in rates of material progress make big differences when compounded over years, the harm that the Fool does to himself will still be significant.

There are, presumably, two kinds of cases in which the Fool may still have reason to violate the norms that he hopes others will follow. There may be cases in which his transgression will not even be noticed, much less linked to him. If no one notices that a norm has been violated, the violation cannot precipitate a downward spiral of noncompliance. Perhaps (MPC) cannot show that one ought not to violate norms in such cases. But such cases are necessarily minor, and so it seems a minor failing, if it is a failing at all, that (MPC) cannot prohibit them.

---

[13] Bicchieri, *The Grammar of Society*, 140.

Of greater concern, though, are cases in which the Fool is sufficiently confident that others will continue to follow a norm, even if he violates it. In seems that an evaluative system that takes an opportunistic view of such situations will be more practically coherent than one that does not. This is a serious problem for (MPC), for which no solution is readily apparent. However, it is worth noting that the inability to convince everyone that they should always be moral is a lesser incapacity than the inability to show anyone that they should ever be moral. It is also an inability shared by every existing moral theory, and though many other theories claim that the Fool "should" be convinced that he ought always to be moral, the difference in practice is negligible. No theory presents an argument that the Fool cannot ignore.

Once the empirical facts about conditional norm-following in humans are taken into account, rule-consequentialist arguments from the social contract tradition show that evaluative systems that permit morally perverse acts, which include but are not limited to robbery, most violence, slavery, etc., are less practically coherent than evaluative systems that prohibit them. Even if (MPC) does lead to relativism, then, it will not lead to the endorsement of morally perverse evaluative systems. It does leave rational grounds for criticizing those who would endorse those acts that are widely regarded as morally abominable.

2.2. The objection from "ethical decay"

Another objection to relativism is that it allegedly undermines our commitment to our ethical values—that it leads to a kind of ethical decay in which we cease to care about ethics, where "caring about ethics" means taking ethical values seriously in deciding what to do. There are two ways of interpreting this objection. It might mean that as a matter of psychological fact, accepting ethical relativism causes people to care less about

ethics. More interestingly, it might mean that once someone accepts ethical relativism, she has less *reason* to care about ethics.[14] Both versions, however, are philosophically toothless, for neither gives us an epistemic reason to deny ethical relativism.

We can formulate the objection as follows:

(1) Accepting ethical relativism causes (or rationally should cause) people to care less about ethics.
(2) People should not care less about ethics.
(3) People should not do things that will lead them to do other things that they should not do.

∴ (4) people should not accept ethical relativism.

This argument—if it is sound—may provide an *ethical* reason to reject relativism, assuming that one thinks it is important for people to care about ethics, but the argument does not provide an *epistemic* reason to do so. The fact that widespread belief in a proposition would have bad consequences (or would have bad consequences if everyone were rational) is no evidence against the proposition. The most intellectually honest position to take, for someone genuinely convinced of the soundness of this objection, is that of the Bishop of Worcester's wife toward the theory of evolution: "Let us hope that it is not true, but if it is, let us pray that it will not become generally known."

It is worth pointing out, however, that the objection is probably not sound on either interpretation. There is no evidence that ethical relativism causes people to care less about ethics. More importantly, there is no reason that ethical relativism *should* cause people to care less about ethics. The fear, it seems, is that relativism unmasks our

[14] Betsy C. Postow, "Dishonest Relativism," *Analysis* 39, no. 1 (1979); Scanlon, "Fear of Relativism," 230.

ethical commitments as arbitrary by showing that we could just have as well have chosen a different set of ethical commitments. But many of our deepest commitments are like this. We might think, in a reflective moment, that there are other spouses, other friends, and other careers that we might reasonably have chosen, but this does not seem to constitute a reason to care less about the spouse, friends, or career that we actually chose. As Scanlon says, citing Nietzsche, "Why…do people think that they can have an ideal only if it is *the ideal*?"[15]

2.3. The objection from incoherence

Among the more technical criticisms of relativism, one of the most common is the objection that relativism is incoherent because it allows contradictory ethical claims to be true simultaneously. For instance, suppose that Mary and Elizabeth hold different views about whether divorce is ethically permissible, with Mary insisting, "Divorce is wrong," and Elizabeth saying, "Divorce is not wrong." On the surface, it appears that Mary and Elizabeth are asserting inconsistent claims with the logical form W*d* and ~W*d*, respectively, where 'W' stands for 'is wrong' and '*d*' for 'divorce'. But if Mary's and Elizabeth's evaluative systems were equally practically coherent, it might seem that (MPC) is committed to saying that divorce is and is not wrong—that W*d* & ~W*d*—which is incoherent.[16]

If Mary and Elizabeth's claims were non-normative in nature—if, say, Mary were claiming that Christopher had converted to Catholicism and Elizabeth were claiming that Christopher had not converted to Catholicism—we would normally believe that one of them must be wrong. It could not be the case that Christopher both had and had not

[15] Scanlon, "Fear of Relativism," 231.
[16] For an extended discussion of this objection from incoherence, see D. Lyons, "Ethical Relativism and the Problem of Incoherence," *Ethics* 86, no. 2 (1976).

converted to Catholicism. Ordinarily, the fact that some theory permits a contradiction is grounds for rejecting that theory. Thus, if we are going to adopt (MPC), we must deal with the fact that it appears to permits contradictions.

Assuming that there are equally practically coherent evaluative systems that contain apparent contradictions, there are two ways to deal with this objection from incoherence.[17] We could hide behind the notion of justification, claiming that Elizabeth and Mary are equally justified in making their respective claims, though at most one of their claims is true. Alternatively, we could adopt an alternative logical regimentation of evaluations, on which evaluative claims are always implicitly relativized to an evaluative system.[18]

2.3.1. Relying on justification rather than truth

Methodological practical coherentism is explicitly formulated as an epistemic principle about when to accept one evaluative system over another. It is therefore a principle about justification, rather than truth. Just as the principle 'Accept any claim that has a probability greater than .999' may justify a soon-to-be lottery winner in believing falsely that he will not win the lottery, so (MPC) may justify Mary in believing falsely that divorce is wrong. The kind of relativism that (MPC) entails, therefore, would only be a relativism of justification, not of truth.

---

[17] Some ethical contradictions could simply be accepted, if we are willing to accept dialetheism. This may solve some of the thornier technical problems with (MPC), but ethical relativism entails a larger number of ethical contradictions than dialetheists are likely to accept. Thus, it is not an adequate solution to the problem as a whole. See G. Priest, "What Is So Bad About Contradictions?" *The Journal of Philosophy* 95, no. 8 (1998).

[18] These strategies are more specific versions of the methods that Lyons considers—viz., showing that relativism need not take both judgments to be true and showing that the two judgments are not really contradictory. See Lyons, "Ethical Relativism and the Problem of Incoherence."

We might adopt this solution if we thought that there were evaluative truths, but that they were forever beyond our reach. We would have an ethical analogue of an instrumentalist view of science. Our methods may not get us to the truth, but because they yield "better" evaluations (in some sense of 'better'), we are justified in using them to decide what to value.

Alternatively, we could deny that ethical claims have truth values, but allow that some are better justified than others, then the fact that E*x* and ~E*x* are equally justified does not entail a contradiction. Since Ex and ~E*x* have no truth values, neither does E*x* & ~E*x*. And since none of those propositions have truth values, they cannot stand in relations of logical entailment to one another.

Neither of these options is particularly appealing. If we insist that there are unknowable ethical truths, we need an account of what it is that makes those ethical truths true—and if the account is to avoid the apparent contradictions implied by ethical relativism, it probably needs to be independent of our actual evaluations. If we insist that ethical claims have no truth values, it is unclear in what sense they can be justified—and even less clear why we should care which are justified and which are not. On the one hand, there seems to be little practical difference between these two options—that is, between holding an instrumentalist view of ethical inquiry and holding that some ethical claims are better justified than others though none have truth values. It might therefore be reasonable simply to set the issue aside. But on the other hand, ethical theorists have recently been moving away from the strictly noncognitivist view that ethical claims have no truth value. Thus, many theorists will prefer a response on which we can comfortably take our evaluative claims to be true or false.

2.3.2. An alternative logical regimentation of evaluations

Mary and Elizabeth's assertions—'Divorce is wrong' and 'Divorce is not wrong'—certainly seem like contradictions when they are assimilated to claims like 'Christopher converted to Catholicism' and 'Christopher did not convert to Catholicism'. But they seem much less problematic when assimilated to claims like 'I am hungry' or 'Edinburgh is nearby'. These claims are context-sensitive; they must be understood as relativized to a particular speaker, time, place, etc. If Mary says, "I am hungry," and Elizabeth says, "I am not hungry," no one cries, "Contradiction!" because their utterances express the propositions H*m* and ~H*e*, respectively, where $m \neq e$. The logical regimentation of those claims—that is, the way we represent those claims in logical notation—differs from the regimentation suggested by their "surface grammar." Adopting an alternative logical regimentation of evaluations, whether indexical or otherwise, could therefore provide a better response to the objection from incoherence.

If we regiment evaluations in such a way that each evaluative claim asserts a relation between some subject and an evaluative system, then the apparent contradiction between Mary's and Elizabeth's statements disappears. Since, by hypothesis, they are using different evaluative systems—call them s_1 and s_2—the logical forms of their claims are Wds_1 and ~Wds_2. Since $s_1 \neq s_2$, there is no problem in allowing that these claims are both true. All ethical predicates, on this view, have an extra, hidden argument, which specifies the evaluative system with respect which the predicate applies to its other arguments.

This suggestion is not new, but it is not widely accepted, either. Gilbert Harman argues that certain types of ethical claims, at least, are relativized either to cultures or to

an agent's motivating attitudes.[19] David Wong holds that ethical claims are relativized to a culture.[20] Jesse Prinz claims that ethical claims are indexical claims, relativized to an individual's "moral sentiments."[21]

Much of the recent discussion of alternative logical regimentation of ethical claims focuses on indexical interpretation of evaluations.[22] The arguments against the indexical interpretation of ethical terms rest on speakers' divergent intuitions about indexical and ethical statements, respectively. Competent language users do not normally treat ethical terms as indexical, whereas they reliably treat terms like 'I' or 'here' as indexical. Most people would treat Mary's claim about divorce as contradicting Elizabeth's claim, whereas no one would treat the statement 'I am in Scotland', uttered by Mary, as contradicting the statement 'I am not in Scotland', uttered by Elizabeth. Furthermore, the indexical view of ethical terms does not sit well with speakers' intuitions about the correctness of certain statements.

There are two strategies for refuting these arguments. The first is to meet them on their own ground and show that speakers' intuitions do not actually show that ethical terms are not indexical. Jesse Prinz takes this strategy in arguing that ethical terms fail standard tests for indexicality primarily for pragmatic reasons. For instance, consider this sentence:

> (TORTURE) When Dick said, "It's morally permissible to torture prisoners," he spoke truly, but it's wrong to torture prisoners.

---

[19] Gilbert Harman, "Moral Relativism Defended," *The Philosophical Review* 84, no. 1 (1975); Harman, *The Nature of Morality*.
[20] Wong, *Natural Moralities*.
[21] Prinz, *The Emotional Construction of Morals*.
[22] Prinz provides the best discussion of this debate, with special attention to criticisms raised by Ernest Lepore and Herman Cappelan. See ibid., 199–205.

This sentence probably sounds anomalous to most people, since it seems to endorse and reject the torture of prisoners. The following sentence, however, is acceptable in most contexts:

> (PRISONER) When John said, "I was once held prisoner," he spoke truly, but I was never held prisoner.

If ethical terms like 'morally permissible' and 'wrong' were indexical, as 'I' is, then (TORTURE) should be as acceptable as (PRISONER). Most people would presumably categorize (TORTURE) as anomalous, which counts against an indexical theory of moral claims, such as Prinz's.

This apparent failure, Prinz argues, is the result of pragmatic expectations. Just as one would only say, "It's raining, but it's not raining here," if the first clause were indexed to some location other than "here," so one would only utter (TORTURE) if the first clause were indexed to some evaluative system other than one's own. Normally, however, when we express an evaluation, we index it to our own evaluative system. Thus, the natural interpretation of (TORTURE) would be the self-contradictory claim, "When Dick said, 'It's morally permissible to torture prisoners', he spoke truly (according to my value system), but it is wrong to torture prisoners (according to my value system)." If, however, we interpret (TORTURE) so that 'morally permissible' is indexed to Dick's evaluative system and 'wrong' is indexed to a different evaluative system, it might be true.

Haggling over which sentences are anomalous, however, is not the only way to show that the logical regimentation of ethical claims differs from their surface grammar. Nor is it the best, for speaker's intuitions, on which judgments of anomaly are based, are not the only determinant of logical regimentation.

We can sometimes overturn speakers' intuitions about the logical regimentation of an expression. Consider the term 'simultaneous'. Intuitively, '*x* was simultaneous with *y*' should be regimented as a two-place relation: S*xy*. Neither its content nor its truth seems to vary with the context in which it is uttered, and so it seems that no third argument is needed in the logical regimentation. But this is false. The theory of special relativity implies that when *x* and *y* are spatially-separated events, '*x* is simultaneous with *y*' is relative to a frame of reference. Two events that are simultaneous for Mary may are not simultaneous for Leia, if Mary and Leia are moving relative to one another. 'Simultaneous' must therefore be regimented as a three-place relation—S*xyf*, where *f* is a frame of reference. We discovered this only after we recognized that there is no objective, God's-eye perspective from which to assess claims about simultaneity, which recognition was forced upon by the discovery that the speed of light in invariant. The recognition that '*x* is moving' must be regimented as a two-place relation—viz., as M*xy*, where *y* is that with respect to which *x* is moving—also represented a discovery. This discovery derived from the recognition that there is no objective, God's-eye perspective from which to detect movement. Likewise, the recognition that ethical claims must always be assessed from within some evaluative system or other warrants the surprising claim that ethical terms must be regimented as a two-place relation: E*xs*, where *x* is the subject of evaluation and *s* is an evaluative system. Surface grammar and speaker's intuitions make no more difference in this case than they do in the cases of simultaneity or motion.

The analogy with 'simultaneous' or 'is moving' has the added advantage of motivating an alternative logical regimentation of ethical terms independently of the need to respond to the objection from incoherence. Einstein and Galileo provided us with new ways of understanding seemingly ordinary phenomena. These new ways of understanding the world would have led to apparently incoherent statements like 'When

Salviati said, “The ball is moving,” and Simplico said, “The ball is not moving,” both spoke truly’. We found that these inconsistencies could be happily resolved by admitting that ‘moving’ and ‘simultaneous’ were relative to a frame of reference, despite the fact that speakers’ intuitions did not support such a logical regimentation of those terms. Likewise, ethical relativism leads to seemingly incoherent ethical statements. Following Einstein and Galileo, we can eliminate the alleged incoherence by adopting a new logical regimentation of the terms under consideration. Thus, when Mary says, “Divorce is wrong,” and Elizabeth says, “Divorce is not wrong,” they are actually saying, “Divorce is wrong according to evaluative system s_1,” and “Divorce is not wrong according to evaluative system s_2,” respectively. These are not contradictory statements. There is no incoherence here.

2.4. The objection from moral disagreement

This strategy for avoiding the objection from incoherence may seem to lead the relativist out of the frying pan and into the fire. If Mary can say that divorce is wrong without contradicting Elizabeth’s claim that divorce is not wrong, then it seems that Mary and Elizabeth cannot really disagree about whether divorce is wrong. We could regiment Mary’s claim as Wds_1 and Elizabeth would be claiming Wds_2, where $s_1 \neq s_2$. Thus, their claims are compatible. Furthermore, since few people will have identical evaluative systems, and most evaluations are relativized to the speaker’s evaluative system, people will rarely disagree ethically. But the claim that we can do not have ethical disagreements contradicts our ordinary experience and practice. People certainly act as if they have frequent, genuine ethical disagreements. We can state this objection from disagreement more precisely:

(1) Ethical relativism implies that people rarely disagree ethically.

(2) People frequently disagree ethically.

∴ (3) Ethical relativism is false.

Most relativists rebut this objection by denying (2). After all, the only evidence for (2) is that people *appear* to have ethical disagreements. That is, they *act* as if they have ethical disagreements and they believe that they have ethical disagreements. When Henry says, "Dog fighting is wrong," and Michael says, "Dog fighting is not wrong," Henry and Michael each takes himself to be disagreeing with the other. They may argue about it as if they disagreed, each trying to change the other's mind. Properly understood, then, the objection from disagreement is:

(1) If ethical relativism is true, then people rarely disagree ethically.

(A1) People frequently appear to disagree ethically.

(A2) The best explanation of the frequent appearance of ethical disagreements is that people do disagree ethically.

∴ (2) People do disagree ethically.

∴ (3) Ethical relativism is false.

Relativists who deny (2) typically deny (A2), offering three complementary explanations of putative ethical disagreement. First, many people mistakenly believe that ethical terms are not relative, and so believe, for instance, that 'Dog fighting is wrong' contradicts 'Dog fighting is not wrong' no matter who says it. These people take themselves to disagree only because they are confused.[23] Second, even putative

[23] Ibid., 120; Wong, *Natural Moralities*, 77.

disagreement may provoke negative emotions, which could motivate the same behaviors as genuine disagreement, such as reproach, criticism, and debate—regardless of how rational such behaviors are. Thus, people will frequently *appear* to disagree, even if they do not.[24] Third, even if people are not disagreeing about ethical claims, they may still disagree about what to do. That is, they may make ethical claims that prescribe mutually exclusive actions. For even ethical claims that, strictly speaking, have different meanings or truth conditions may nonetheless prescribe or proscribe the same type of action. A putative ethical disagreement may therefore be a genuine pragmatic disagreement.[25] As Philippa Foot points out, this kind of disagreement can arise even when there is no question of disagreement about the truth. If two people issue incompatible imperatives, each will hope to see his own imperative obeyed.[26] Since these explanations can account for the appearance of ethical disagreement without positing actual ethical disagreement—and therefore without making whatever claims of ethical theory would be necessary to ground ethical disagreement—the relativist can reject the objection from disagreement.

Though this is technically an adequate response to the objection from disagreement, many critics of relativism will be unimpressed. This response seems to leave intact the unsettling idea that we cannot meaningfully disagree with someone like Hitler, Stalin, or Mao about whether to order the deaths of millions of people. To call our disagreement with them a pragmatic difference belies the feeling that these men have made a terrible mistake when they come to think their actions permissible. It suggests that ethical debate between people with different evaluative systems is rarely rationally justified. Some hard-nosed relativists, like Harman, might accept this, and it

---

[24] Prinz, *The Emotional Construction of Morals*, 120–21.
[25] Ibid., 121; Wong, *Natural Moralities*, 77–78.
[26] Foot, "Moral Relativism."

is not clear what argument critics will give, beyond appealing to moral appearances, to show that the hard-nosed relativist is wrong. But relativists can explain why ethical debate is often a reasonable response to putative ethical disagreement.

As David Wong emphasizes, few people live by values that are totally alien to us.[27] Most differences in evaluative systems arise from differences in the priority attached to different values or differences in the application of evaluations to concrete situations. For instance, some people might value the good of the community as a whole over the autonomy of the individual. Two people who both condemn taking an innocent human life might disagree about whether that evaluation is relevant to decisions about first trimester abortions. In fact, the values of those with whom with have ethical disagreements are rarely completely alien to us. Our differences with the mobster are primarily about when violence is appropriate—but not, for most us, whether violence is ever appropriate, and certainly not about whether loyalty, family, and reciprocity are good things. Our differences with the Inquisitor are about whether securing religious conversions is worth the infliction of great suffering—but not whether suffering is, *ceteris paribus*, bad. Though the distance between what we are prepared to do and what they are prepared to do may be great, the distance between what we value and what they value may be smaller than it appears.

There are, however, exceptions. Prinz documents a wide range of allegedly fundamental disagreements over values. The Aztecs endorsed cannibalism. The Romans adored gladiators. The Chinese long permitted concubinage. The Saudis allow ploygyny, the Tibetans polyandry. Different societies hold widely different views on premarital sex, marital infidelity, homosexuality, and even bestiality. The Inuit openly

---

[27] Wong, *Natural Moralities*.

practiced infanticide until quite recently, and the Chinese still practice it, though discreetly.[28]

Some, but not all, of these differences can be explained as differences in prioritization of shared values. Prinz insists that the Romans' glorification of gladiators depended on a peculiarly Roman interest in physical courage and valor. But most contemporary Westerners, at least, value courage and valor, too. We place a greater value on other things, like human dignity, however, which lead us to cringe at the thought of humans fighting each other to the death for spectators' entertainment. Some people argue that Inuit infanticide is merely an adaptation to a harsh environment that cannot support many people, and that we, too, might practice infanticide in such conditions. But as Prinz notes, most of us would consider infanticide only as a last resort in terrible conditions—as a horrific, if necessary, evil—whereas the Inuit did not seem to view it that way. It is not clear that they valued the lives of such infants at all. Likewise, many differences in sexual mores do not seem to be explicable in terms of differently prioritized values. In such cases, there may be little for either party to gain from ethical discussion.

Still, given the extensive overlap in evaluations between the evaluative systems of the parties to an ethical disagreement, we would sometimes be wise to take others' disagreement as an invitation to reconsider the practical coherence of our own system. The tension between the long-term gains of cooperation and the short-term of taking advantage of others' cooperation yields practical dissonance. Eastern societies, in general, have resolved this tension differently than Western societies have, placing a greater emphasis on community than most Westerners do. Given that Eastern evaluative systems arose as a response to the same practical dissonance that pushed us

---

[28] Prinz, *The Emotional Construction of Morals*, 187.

Westerners into ours, perhaps we ought to examine Eastern solutions—to listen to the reasons that they offer in ethical debates—and see if any of their solutions would improve the practical coherence of our own evaluative systems.

In other words, ethical debate is worthwhile because the pragmatic differences that arise from incompatible evaluations point to alternative solutions to ethical problems. And since we cannot know in advance whether adopting a different solution will improve the practical coherence of our own evaluative system, it is often worth discussing others' reasons for holding their divergent ethical views. We or they may profit from such discussion, even if, technically, we do not disagree about any ethical claims.

So far, this response to the objection from disagreement has been two-pronged. On the one hand, relativists can argue that the objection from disagreement is unsound. It rests on the assumption that the best explanation of the frequent appearance of ethical disagreement is that people do frequently disagree about ethical claims, and this is false. But on the other hand, relativists can admit that ethical debate can be fruitful for both parties, even though they may not, strictly speaking, disagree about anything.

It is worth noting, however, that it is possible, on the proposed logical regimentation of evaluations, for two people who hold distinct evaluative systems to disagree. Even if Simplicio and Salvatore are moving relative to one another, and therefore not in the same frame of reference, they can meaningfully disagree about whether something is moving *relative to Simplicio*. Likewise, Mary and Elizabeth might disagree about whether killing Elizabeth would be wrong *relative to Mary's evaluative system*. There is no reason that Elizabeth cannot say that Mary's evaluative system would be more practically coherent if Mary took a dimmer view of murder. The two could intelligibly debate the effects on specific revisions to Mary's evaluative system on

the likelihood that Mary would achieve her various goals. Thus, not only is ethical discussion often fruitful, but it is entirely possible for it to be a debate about the truth of a single proposition, relativized to a single evaluative system.

It is also possible for putative evaluative disagreements to serve as a façade for genuine disagreements about which evaluative system to adopt. We might interpret Mary and Elizabeth not as making assertions about the place of divorce within their respective evaluative systems, but as advocating that everyone should adopt an evaluative system that condemns or permits divorce. When their evaluative systems overlap considerably, arguing about divorce itself may be a productive way to reach agreement. When their systems diverge substantially, they may need to resort to an explicit discussion of the practical coherence of their respective systems to make any progress.

A similar point holds about the notion of moral progress, which might be construed as disagreement across time. When contemporary ethicists insist that some feature of modern evaluative systems, such as the abolition of slavery, constitutes moral progress, they can be interpreted as claiming that our current is more practically coherent (at least in that respect) than some previous systems. The current system is, in that sense, better than previous systems, and thus constitutes progress.

3. Conclusion

If methodological practical coherentism appears unsettling at first, thinking carefully about its implications should put some of those fears to rest. Though (MPC) probably leads to relativism, the kind of relativism it entails does not suffer from the objections typically raised against relativism. It will not lead to moral perversion or ethical decay. It leaves room for ethical disagreement and explains the value of ethical discussion even

in the absence of genuine disagreement. It does not lapse into incoherence, though it does require an alternative logical regimentation of evaluative claims.

Conclusion

"Philosophy recovers itself," suggests John Dewey, "when it ceases to be a device for dealing with the problems of philosophers and becomes a device, cultivated by philosophers, for dealing with the problems of men."[1] Non-philosophers would be excused for thinking that moral philosophy needs no such recovery—that it is and always has been "a device for dealing with the problems of men." Those with more than passing acquaintance with moral philosophy, however, know otherwise. Much of what moral philosophers do has little bearing on the problems we face in life. Especially in the twentieth century, many moral philosophers have set aside the original Socratic question—"How should one live?"—in favor of more abstract questions about the meaning of moral language, the nature of moral thought, the metaphysical status of "moral facts," and so forth.

This dissertation is, in some ways, one more contribution to those theoretical questions about ethics. I cannot pretend that I have said anything enlightening about how one should live. But if this dissertation seems to address only "problems of philosophers," I hope it is a step toward solving an important, if abstract, "problem of men." The problem I have in mind is one that H. Standish Thayer identifies as "the main problem of modern philosophy." It consists in "finding some way of integrating in one conceptual scheme a sound interpretation of natural knowledge and of moral values."[2]

This problem is a problem for all of us, not just for philosophers. Over the last several centuries, our understanding of the natural world has far surpassed our understanding of value. The scientific view of the world sometimes threatens to expel

---

[1] John Dewey, *The Middle Works, 1899-1924*, vol. 10 (Carbondale: Southern Illinois University Press, 1980), 46.
[2] H. S. Thayer, *Meaning and Action: A Critical History of Pragmatism* (Indianapolis: Bobbs-Merrill, 1968), 454.

value from it altogether. Until we develop a conception of value suitable to our contemporary understanding of the natural world, our moral knowledge will wither in the shadow of our scientific knowledge. Only when we can place our moral knowledge on a par with our scientific knowledge—only when we can integrate them into one conceptual scheme—will we make the kind of progress in ethics that will enable us to build a better world for ourselves.

I have argued that our existing conception of moral knowledge is tied firmly to what I call "moral appearances." Moral appearances are noninferentially derived token mental states the intentional content of which, if expressed, would amount to a moral claim (as opposed to a nonmoral claim). Though I doubt that it is possible to give a theoretically adequate definition of "moral claim," some paradigmatic moral claims include 'Torture is wrong', 'Stalin was evil', and 'It is unfair to force someone to atone for their ancestor's misdeeds'. Paradigmatic nonmoral claims include non-evaluative claims, such as 'Manhattan is an island', and nonmoral evaluative claims like 'Mushrooms are disgusting'. Thus, if someone sees Eddie Adams' famous photograph of General Nguyen Ngoc Loan executing a Viet Cong prisoner during the Vietnam War, and immediately thinks, "It was morally wrong of him to execute that prisoner," that person has had a moral appearance—even if, in the end, the person concludes that it was permissible for the general to do what he did.

Moral appearances, in the guise of "moral intuitions," play an important role in most moral epistemology because they provide the basic premises in many moral arguments. They also provide the data from which many moral theories are built and against which moral theories are tested. They play this role because they provide the only obvious entrance into the realm of normative discourse—into claims about what ought to be, as opposed to what is. Since we cannot draw conclusions about what ought

to be from premises concerning only what is, the premises supported by moral appearances seem indispensable to moral epistemology.

Despite their widespread use in moral philosophy, however, moral appearances are not yet well understood. Where do they come from? Why do we have the moral appearances that we do? Science is beginning to discover answers to these questions. Recent work in functional neuropsychology, cognitive science, neural pathology and psychopathology provides a wealth of new information on this front. Much of this research points in a sentimentalist direction, but several lines of evidence point to a more complicated picture. Synthesizing two of the main lines of research in the field, I proposed a new, "developmental sentimentalist" model of moral psychology, which I call the "Mencian creature model." On this model, our moral appearances arise from "moral sentiments," and these sentiments develop through a process of emotional conditioning mediated by unconscious action-categorizing processes.

Given this naturalistic account of moral psychology, we can explain our moral appearances without supposing that their intentional content is true. I argued that this provides a *prima facie* reason to discount moral appearances in deciding which moral claims to endorse. I also argued that common responses to this problem, including sensibility theory and rationalism, fail to vindicate moral appearances. Sensibility theory, as it is typically construed, is caught between abandoning normativity and defining moral properties in circular terms. More plausible versions of sensibility theory, which relativize moral properties to individuals or groups, fail to provide a reason to take our moral appearances seriously as providing premises about how one ought to behave.

Rationalism about moral appearances fails because it must resort, ultimately, to rational intuitionism about some moral claims. Using Robert Audi's reformulation of Rossian intuitionism as a stalking horse, I argued that rational intuitionism is

inadequate as a moral epistemology. Not only is the rationalist view of moral appearances psychologically problematic, but the "soft" self-evidence preferred by contemporary intuitionists is especially vulnerable to epistemic problems involving "peer disagreement." These two independent problems render rationalism unfit as a vindication of moral appearances.

A new problem arises from this doubt about moral appearances. If we cannot use moral appearances to justify moral claims, it may seem that the only option left for moral epistemology is coherence theory; but it is unclear how we could show that one coherent moral system is any better than any other. To solve this problem, I introduced the notion of "practical coherence." A system of evaluative claims, containing both moral and nonmoral evaluations, is practically coherent to the extent that, given current circumstances, achieving the goal(s) recommended by any one evaluative claim in the system increases, or at least does not reduce, the probability of achieving the goals recommended by other claims in the system. Because the practical relations between different goals are determined by the world, not by what we think, practical coherence ties evaluative systems to the world. This dependence on both the values that we hold and the attitude-independent relations among various goals yields an unusual combination of limited ethical relativism and moderate moral realism.

I argued that the kind of ethical relativism engendered by reliance on practical coherence survives the usual objections to ethical relativism. It does not lead to seemingly objectionable moral systems or undermine our commitment to morality. It is not incoherent, and it does not make moral disagreement impossible or moral discussion pointless.

What I take to be most important in this dissertation, however, is the recommendation of what I called "methodological practical coherentism," which is the

thesis that, *ceteris paribus*, given two evaluative systems, one should adopt the more practically coherent system. I suggested that this principle, when put into practice, yields a view of ethical inquiry much like John Dewey's. When we identify defects in our evaluative systems—aspects of our systems that are not as practically coherent as they could be—we should investigate the effects of various revisions on the practical coherence of those systems. This investigation would require the kind of multidisciplinary, empirical research involved in fields like policy analysis or ecology. By reuniting the methods we use to investigate value with the methods we use to investigate the natural world, this vision of ethical inquiry could integrate our knowledge of values with our knowledge of the world.

It is, however, one thing to suggest that it could be done. It is another to do it. While this dissertation points the way to what I hope will prove a fruitful method of ethical inquiry, there is still a great deal of work to be done. The most productive way of refining this proposed methodology, I suspect, is to put it into practice—to use it to solve specific "problems of men," refining it as needed. The way to solve these problems is not to develop a unified theory of ethics based on practical coherence, and then apply that theory to problems, but rather to begin with a specific problem, identify the practical dissonance underlying that problem, and propose revisions to our evaluative system that improve the system's practical coherence.[3] The way forward, then, lies not in doing more theory, but in doing ethics in accordance with the theory developed here.

My hope is that pursuing ethics in this way will not only solve the specific problems under investigation, but also illuminate new ways of arguing for ethical conclusions and provide the basis for new theories about what is most valuable. If these

---

[3] *Cf.* Bryan Norton's distinction between "applied" and "practical" philosophy in Bryan G. Norton, "Integration or Reduction: Two Approaches to Environmental Values," in *Environmental Pragmatism*, ed. Andrew Light and Eric Katz (London: Routledge, 1995), 107–08.

things are worth doing, however, it will be because they enable us to cope more effectively with our changing world and thus to live happier, more satisfying lives.

Bibliography

Alchourron, C. E. "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions." *The Journal of symbolic logic* 50, no. 2 (1985): 510.

Annas, J. "Comments on John Doris's Lack of Character." *Philosophy and Phenomenological Research* 71, no. 3 (2005): 636.

Anscombe, G. E. M. "Modern Moral Philosophy." *Philosophy* 33, no. 124 (1958).

Aronson, Elliot. *The Social Animal*. 5th ed. New York: Freeman, 1988.

Arpaly, N. "Comments on Lack of Character by John Doris." *Philosophy and Phenomenological Research* 71, no. 3 (2005): 643.

Audi, Robert. *The Good in the Right: A Theory of Intuition and Intrinsic Value*. Princeton, NJ: Princeton University Press, 2004.

———. "Prospects for a Value-Based Intuitionism." In *Ethical Intuitionism: Re-Evaluations*, edited by Philip Stratton-lake, 2002.

———. "Self-Evidence." *Philosophical Perspectives* 13 (1999): 205–49.

Bicchieri, Cristina. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press, 2006.

Bicchieri, Cristina, and Erte Xiao. "Do the Right Thing: But Only If Others Do." *Journal of Behavioral Decision Making* 22, no. 2 (2007): 191–208.

Bjorklund, Fredrik, and Jonathan Haidt. "Vivid Violations Are Morally Worse." Lund, Sweden: University of Lund, in preparation.

Blackburn, Simon. *Ruling Passions*. New York: Oxford University Press, 1998.

Blair, James, A. A. Marsh, E. Finger, K. S. Blair, and J. Luo. "Neuro-Cognitive Systems Involved in Morality." *Philosophical Explorations* 9, no. 1 (2006): 13–27.

Blair, R. J. R. "A Cognitive Developmental Approach to Morality: Investigating the Psychopath." *Cognition* 57 (1995): 1–29.

———. "The Emergence of Psychopathy: Implications for the Neuropsychological Approach to Developmental Disorders." *Cognition* 10 (2006): 414–42.

———. "Moral Reasoning in the Child with Psychopathic Tendencies." *Personality and Individual Differences* 22 (1997): 731–39.

Blair, R. J. R., L. Jones, F. Clark, and M. Smith. "Is the Psychopath "Morally Insane"?" *Personality and Individual Differences* 19 (1995): 741–52.

Blair, R. J. R., J. Monson, and N. Frederickson. "Moral Reasoning and Conduct Problems in Children with Emotional and Behavioral Difficulties." *Personality and Individual Differences* 31 (2001): 799–811.

Blair, R. J. R., C. Sellars, I. Strickland, F. Clark, A. O. Williams, and M. Smith. "Emotion Attributions in the Psychopath." *Personality and Individual Differences* 19 (1995): 431–437.

Bonjour, Laurence. *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press, 1985.

Bono. "Foreword." In *The End of Poverty*, edited by Jeffrey D. Sachs. New York: Penguin, 2005.

Borg, Jana Schaich, Catherine Hynes, John Van Horn, Scott Grafton, and Walter Sinnott-Armstrong. "Consequences, Action, and Intention as Factors in Moral Judgments: An fMRI Investigation." *Journal of Cognitive Neuroscience* 18, no. 5 (2006): 803–17.

Brandom, Robert. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press, 1994.

Ciaramelli, Elisa, Michela Muccioli, Elisabetta Làdavas, and Giuseppe di Pellegrino. "Selective Deficit in Personal Moral Judgment Following Damage to Ventromedial Prefrontal Cortex." *SCAN* 2 (2007): 84–92.

Clarke, Samuel. *A Discourse Concerning the Being and Attributes of God, the Obligations of Natural Religion, and the Truth and Certainty of the Christian Revelation*. Glasgow: Richard Griffin, 1823.

Crisp, Roger. "Sidgwick and the Boundaries of Intuitionism." In *Intuitionism: Re-Evaluations*, edited by Philip Stratton-Lake, 56–75. Oxford: Oxford University Press, 2002.

Cushman, Fiery, Liane Young, and Marc Hauser. "The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm." *Psychological Science* 17, no. 12 (2006): 1082–89.

D'Arms, Justin, and Daniel Jacobson. "Sensibility Theory and Projectivism." In *The Oxford Handbook of Ethical Theory*, edited by David Copp, 665. New York: Oxford University Press, 2006.

Damasio, A. R. "Individuals with Sociopathic Behavior Caused by Frontal Damage Fail to Respond Autonomically to Social Stimuli." *Behavioural Brain Research* 41, no. 2 (1990): 81.

Dancy, Jonathan. *Ethics without Principles*. Oxford: Oxford University Press, 2004.

DePaul, Michael R. "Intuitions in Moral Inquiry." In *The Oxford Handbook of Ethical Theory*, edited by David Copp. Oxford: Oxford University Press, 2006.

Dewey, John. *The Middle Works, 1899-1924*, vol. 10. Carbondale: Southern Illinois University Press, 1980.

———. *The Middle Works, 1899-1924*, vol. 20. Carbondale: Southern Illinois University Press, 1982.

Dewey, John, and James H. Tufts. *Ethics*. New York: Henry Holt, 1908.

Doris, John M. *Lack of Character: Personality and Moral Behavior*. Cambridge: Camrbidge University Press, 2002.

———. "Précis of Lack of Character." *Philosophy and Phenomenological Research* 71, no. 3 (2005): 632.

Dostoyevsky, Fyodor. *The Brothers Karamazov*. Translated by Andrew R. MacAndrew. New York: Random House, Bantam Dell, 1970.

Dworkin, Ronald. "Truth and Objectivity: You'd Better Believe It." *Philosophy and Public Affairs* 25, no. 2 (1996): 87–139.

Elga, Adam. "Reflection and Disagreement." *Noûs* 41, no. 3 (2007): 478–502.

Feldman, Richard. "Epistemological Puzzles About Disagreement." In *Epistemology Futures*, edited by Stephen Hetherington. Oxford: Oxford University Press, 2006.

———. *Epistemology*. Upper Saddle River, NJ: Prentice Hall, 2003.

Flanagan, Owen J., Hagop Sarkissian, and David Wong. "Naturalizing Ethics." In *Moral Psychology*, edited by Walter Sinnott-Armstrong. Cambridge, MA: The MIT Press, 2007.

Foot, Philippa. "Killing and Letting Die." In *Abortion: Legal and Moral Perspectives*, edited by Jay L. Garfield and Patricia Hennessey. Amherst: University of Massachusetts Press, 1984.

———. "Moral Relativism." Lindley Lecture. University of Kansas, 1978.

Frankena, William. *Ethics*. Englewood Cliffs, NJ: Prentice-Hall, 1973.

Gansberg, Martin. "37 Who Saw Murder Didn't Call the Police; Apathy at Stabbing of Queens Woman Shocks Inspector." *New York Times*, March 27, 1964.

Gibbard, Allan. *Thinking How to Live*. Cambridge, MA: Harvard University Press, 2003.

Gigerenzer, Gerd. "Moral Intuition = Fast and Frugal Heuristics?" In *Moral Psychology*, edited by Walter Sinnott-Armstrong. Cambridge, MA: MIT Press, 2007.

Grafman, Jordan. "The Structured Event Complex and the Human Prefrontal Cortex." In *Principles of Frontal Lobe Function*, edited by Donald T. Stuss and Robert T. Knight. New York: Oxford University Press, 2000.

Greene, J. D. "Why Are VMPFC Patients More Utilitarian? A Dual-Process Theory of Moral Judgment Explains." *Trends in Neurosciences* 11, no. 8 (2007): 322.

Greene, Joshua D., and Jonathan Haidt. "How (and Where) Does Moral Judgment Work?" *Trends in Cognitive Sciences* 6, no. 12 (2002): 517–23.

Greene, Joshua D., L. E. Nystrom, A. D. Engell, J. M. Darley, and J. D. Cohen. "The Neural Bases of Cognitive Conflict and Control in Moral Judgment." *Neuron* 44 (2004): 389–400.

Greene, Joshua D., R. B. Sommerville, L. E. Nystrom, J. M. Darley, and J. D. Cohen. "An fMRI Investigation of Emotional Engagement in Moral Judgment." *Science* 293 (2001a): 2105–08.

———. "Supplementary Material." Science, http://www.sciencemag.org/cgi/content/full/293/5537/2105/DC1.

Haidt, Jonathan. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108, no. 4 (2001): 21.

———. "The Emotional Dog Gets Mistaken for a Possum." *Review of General Psychology* 8 (2004): 283–90.

———. "The New Synthesis in Moral Psychology." *Science* (2007).

Haidt, Jonathan, and Fredrik Bjorklund. "Social Intuitionists Answer Six Questions About Moral Psychology." In *Moral Psychology*, edited by Walter Sinnott-Armstrong. Cambridge, MA: MIT Press, 2008.

Haidt, Jonathan, Fredrik Bjorklund, and Scott Murphy. "Moral Dumbfounding: When Intuition Finds No Reason." Charlottesville, VA: University of Virginia, 2004.

Haidt, Jonathan, and Matthew A Hersh. "Sexual Morality: The Cultures and Emotions of Conservatives and Liberals." *Journal of Applied Social Psychology* 31, no. 1 (2001): 191–221.

Haidt, Jonathan, S. Koller, and M. Dias. "Affect, Culture, and Morality, or Is It Wrong to Eat Your Dog?" *Journal of Personality and Social Psychology* 65 (1993): 613–28.

Hare, R. M. *Essays in Ethical Theory*. Oxford: Oxford University Press, 1989.

———. *The Language of Morals*. Oxford: Clarendon Press, 1952.

Harman, Gilbert. "Moral Relativism Defended." *The Philosophical Review* 84, no. 1 (1975).

———. "The Nonexistence of Character Traits." *Proceedings of the Aristotelian Society* 100, no. 2 (2000).

———. *The Nature of Morality*. New York: Oxford University Press, 1977.

Hauser, Marc, Fiery Cushman, Liane Young, R. Kang-Xing Jin, and John Mikhail. "A Dissociation between Moral Judgments and Justifications." *Mind & Language* 22, no. 1 (2007): 1–21.

Hauser, Marc D. *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong*. New York: ecco, 2006.

Hauser, Marc, Liane Young, and Fiery Cushman. "Reviving Rawls' Linguistic Analogy: Operative Principles and the Causal Structure of Moral Actions." In *Moral Psychology*, edited by Walter Sinnott-Armstrong. Cambridge, MA: MIT Press, 2008.

Heekeren, Hauke R., Isabell Wartenburger, Helge Schmidt, Kristin Prehn, Hans-Peter Schwintowski, and Arno Villringer. "Influence of Bodily Harm on Neural Correlates of Semantic and Moral Decision-Making." *NeuroImage* 24 (2005): 887–97.

———. "An fMRI Study of Simple Ethical Decision-Making." *NeuroReport* 14, no. 9 (2003): 1215–19.

Hobbes, Thomas. *Leviathan*. 1651.

Hudson, William Donald. *Ethical Intuitionism*. London: Macmillan, 1967.

Hume, David. *An Enquiry Concerning the Principles of Morals*. 1751. Reprint, Indianapolis, IN: Hackett, 1983.

———. *A Treatise of Human Nature,* 1740.

Johns, J. H., and H. C. Quay. "The Effect of Social Reward on Verbal Condition in Psychopathic and Neurotic Military Officers." *Journal of Consulting & Clinical Psychology* 26 (1962): 217–20.

Kahane, Guy, and Nicholas Shackel. "Do Abnormal Responses Show Utilitarian Bias?" *Nature* 452, no. 7185 (2008).

Kappel, Klemens. "Challenges to Audi's *Ethical Intuitionism*." *Ethical Theory and Moral Practice* 5, no. 4 (2002): 391–413.

Kelly, Thomas. "The Epistemic Significance of Disagreement." In *Oxford Studies in Epistemology*, edited by Tamar Szabo Gendler and John Hawthorne, 167–196. Oxford: Oxford University Press, 2005.

———. "Peer Disagreement and Higher Order Evidence." In *Disagreement*, edited by Richard Feldman and Ted Warfield. New York: Oxford University Press, forthcoming.

Kennett, Jeanette, and Cordelia Fine. "Internalism and the Evidence from Psychopaths and 'Acquired Sociopaths'." In *Moral Psychology*, edited by Walter Sinnott-Armstrong. Cambridge, MA: MIT Press, 2008.

Kiehl, Kent A. "Without Morals: The Cognitive Neuroscience of Criminal Psychopaths." In *Moral Psychology*, edited by Walter Sinnott-Armstrong. Cambridge, MA: MIT Press, 2008.

Koenigs, M. "Koenigs et al. Reply." *Nature* 452, no. 7185 (2008).

Koenigs, Michael, Liane Young, Ralph Adolphs, Daniel Tranel, Fiery Cushman, Marc Hauser, and Antonio R. Damasio. "Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgments." *Nature* 446 (2007): 908–11.

Lau, D. C. "Introduction." In *Mencius*. London: Penguin Books, 1970.

Luo, Qian, Marina Nakic, Thalia Wheatley, Rebecca Richell, Alex Martin, and R. James R. Blair. "The Neural Basis of Implicit Moral Attitude—an IAT Study Using Event-Related fMRI." *NeuroImage* 30 (2006): 1449–57.

Lyons, D. "Ethical Relativism and the Problem of Incoherence." *Ethics* 86, no. 2 (1976): 107.

MacIntyre, Alisdair. *A Short History of Ethics*. 2nd ed. Notre Dame, IN: Notre Dame University Press, 1996.

Mandelbaum, Maurice. *The Phenomenology of Moral Experience*. Glencoe, IL: The Free Press, 1955.

McDowell, John. "Values and Secondary Qualities." In *Mind, Value, and Reality*. Cambridge: Harvard University Press, 1985/1998.

Mencius. *Mencius*. Translated by D. C. Lau. London: Penguin Books, 1970.

Mill, J. S. *Autobiography*. New York: Penguin, 1990.

Milton, John. *Paradise Lost*. 1667.

Moll, Jorge. "Moral Judgments, Emotions and the Utilitarian Brain." *Trends in Neurosciences* 11, no. 8 (2007).

Moll, Jorge, Ricardo de Oliveira-Souza, Ivanei E. Bramati, and Jordan Grafman. "Functional Networks in Emotional Moral and Nonmoral Social Judgments." *NeuroImage* 16 (2002): 696–703.

Moll, Jorge, Ricardo de Oliveira-Souza, and Paul J. Eslinger. "Morals and the Human Brain: A Working Model." *NeuroReport* 14, no. 3 (2003): 299–305.

Moll, Jorge, Ricardo de Oliveira-Souza, Paul J. Eslinger, Ivanei E. Bramati, Janaína Mourão-Miranda, Pedro Angelo Andreiuolo, and Luiz Pessoa. "The Neural Correlates of Moral Sensitivity: A Functional Magentic Resonance Imaging Investigation of Basic and Moral Emotions." *The Journal of Neuroscience* 22, no. 7 (2002).

Moll, Jorge, Ricardo de Oliveira-Souza, Fernanda Tovar Moll, Fatima Azevedo Ignacio, Ivanei E. Bramati, Egas Caparelli-Daquer, and Paul J. Eslinger. "The Moral Affiliations of Disgust: A Functional MRI Study." *Cognitive and Behavioral Neurology* 18, no. 1 (2005): 68–78.

Moll, Jorge, Ricardo de Oliveira-Souza, Roland Zahn, and Jordan Grafman. "The Cognitive Neuroscience of Moral Emotions." In *Moral Psychology*, edited by Walter Sinnott-Armstrong. Cambridge, MA: MIT Press, 2008.

Moll, Jorge, Paul J. Eslinger, and Ricardo de Oliveira-Souza. "Frontopolar and Anterior Temporal Cortex Activation in a Moral Judgment Task." *Arq. Neuro-Psiquiatr* 59, no. 3B (2001).

Nichols, Shaun. *Sentimental Rules: On the Natural Foundations of Moral Judgment*. New York: Oxford University Press, 2004.

Norton, Bryan G. "Integration or Reduction: Two Approaches to Environmental Values." In *Environmental Pragmatism*, edited by Andrew Light and Eric Katz. London: Routledge, 1995.

Nowell-Smith, P. H. *Ethics*. Harmondsworth: Penguin, 1954.

Paley, William. *The Principles of Morality and Political Philosophy*. n.p.: P. Byrne, L. White, and W. M'Kenzie, 1788.

Peirce, Charles S. "Some Consequences of Four Incapacities Claimed for Man." *Journal of Speculative Philosophy* 2 (1868).

Pinker, Steven. "The Moral Instinct." *The New York Times Magazine*, January 13, 2008 2008.

Postow, Betsy C. "Dishonest Relativism." *Analysis* 39, no. 1 (1979).

Prichard, H. A. *Moral Obligation*. Oxford: Clarendon Press, 1949.

Priest, G. "What Is So Bad About Contradictions?" *The Journal of Philosophy* 95, no. 8 (1998).

Prinz, Jesse. *The Emotional Construction of Morals*. New York: Oxford University Press, 2007.

Quine, W. V. "Two Dogmas of Empiricism." *The Philosophical Review* 60, no. 2 (1951).

Rachels, James. "Active and Passive Euthanasia." *New England Journal of Medicine* 292, (1975): 78–80.

Raphael, D. D. *British Moralists: 1650–1800*. 2nd Revised ed. 2 vols. Indianapolis, IN: Hackett, 1990.

Rawls, John. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1971.

Roskies, A. "Are Ethical Judgments Intrinsically Motivational? Lessons From "Acquired Sociopathy"." *Philosophical Psychology* 16, no. 1 (2003): 51–66.

Ross, W. D. *Foundations of Ethics*. Oxford: Clarendon Press, 1939.

Scanlon, T. M. "Fear of Relativism." In *Virtues and Reasons*, edited by Rosalind Hursthouse, Gavin Lawrence and Warren Quinn. Oxford: Oxford University Press, 1995.

Schnall, S. H., Jonathan Haidt, and G. L. Clore. "Irrelevant Disgust Makes Moral Judgment More Severe, for Those Who Listen to Their Bodies ." Charlottesville, VA: University of Virginia, 2005.

Selby-Bigge, L. A. *British Moralists*. 2 vols. Whitefish, MT: Kessinger, 2007.

Sextus, Empiricus. *Outlines of Skepticism*. Translated by Julia Annas and Jonathan Barnes. Cambridge: Cambridge University Press, 1994.

Shafer-Landau, Russ. *Moral Realism: A Defence*. Oxford: Clarendon Press, 2003.

Sidgwick, Henry. *The Methods of Ethics*. 7th ed. Indianapolis: Hackett, 1981.

Skaperdas, Stergios. "The Political Economy of Organized Crime: Providing Protection When the State Does Not." In *Conflict and Governance*, edited by Amihai Glazer and Kai Andreas Konrad. Springer, 2003.

Smetana, J. "Understanding of Social Rules." In *The Development of Social Cognition: The Child as Psychologist*, edited by M. Bennett. New York: Guilford Press, 1993.

Snyder, Laura J. "Consilience, Confirmation, and Realism." In *Scientific Evidence*, edited by Peter Achinstein. Baltimore: Johns Hopkins University Press, 2005.

Solomon, R. C. "What's Character Got to Do with It?" *Philosophy and Phenomenological Research* 71, no. 3 (2005).

Stevenson, Charles Leslie. "The Emotive Meaning of Ethical Terms." *Mind* 46, no. 181 (1937): 14–31.

Stillman, Tyler F., Roy F. Baumeister, and C. Nathan DeWall. "What's So Funny About Not Having Money? The Effects of Power on Laughter." *Personality and Social Psychology Bulletin* 33, no. 11 (2007): 1547–58.

Stratton-Lake, Philip, ed. *Ethical Intuitionism: Re-Evaluations*. Oxford: Oxford University Press, 2002.

———. "Introduction." In *Ethical Intuitionism: Re-Evaluations*, edited by Philip Stratton-Lake. Oxford: Oxford University Press, 2002.

———. "Pleasure and Reflection in Ross's Intuitionism." In *Ethical Intuitionism: Re-Evaluations*, edited by Philip Stratton-Lake. Oxford: Oxford University Press, 2002.

Sturgeon, Nicholas. "Moral Explanations." In *Morality, Reason, and Truth: New Essays on the Foundations of Ethics*, edited by David Copp and David Zimmerman. Totowa, NJ: Rowman & Allanheld, 1985.

Thayer, H. S. *Meaning and Action: A Critical History of Pragmatism*. Indianapolis: Bobbs-Merrill, 1968.

Thomson, Judith Jarvis. "A Defense of Abortion." *Philosophy and Public Affairs* 1, no. 1 (1971): 47–66.

———. "Killing, Letting Die, and the Trolley Problem." *Monist: An International Quarterly of General Philosophical Inquiry* 59, no. 2 (1976): 204–17.

Tisak, M. "Domains of Social Reasoning and Beyond." In *Annals of Child Development*, edited by R Vasta. London: Jessica Kingsley, 1995.

Valdesolo, Piercarlo, and David DeSteno. "Manipulations of Emotional Context Shape Moral Judgment." *Psychological Science* 17, no. 6 (2006): 476–77.

Vonnegut, Kurt. *Slaughterhouse-Five*. New York: Delacorte Press, 1969.

Wheatley, Thalia, and Jonathan Haidt. "Hypnotic Disgust Makes Moral Judgments More Severe." *Psychological Science* 16, no. 10 (2005): 780–84.

Whewell, William. *Novum Organon Renovatum*. Third ed. London: John W. Parker and Son, 1858.

Wiggins, David. *Needs, Values, Truth: Essays in the Philosophy of Value*. Oxford: Clarendon Press, 1998.

Williams, Bernard. *Ethics and the Limits of Philosophy*. Cambridge, MA: Harvard University Press, 1985.

Williams, Michael. *Unnatural Doubts*. Princeton: Princeton University Press, 1996.

Wilson, Catherine. *Moral Animals: Ideals and Constraints in Moral Theory*. Oxford: Oxford University Press, 2004.

Wong, David B. *Natural Moralities*. New York: Oxford University Press, 2006.

Wood, Jacqueline N., and Jordan Grafman. "Human Prefrontal Cortex: Processing and Representational Perspectives." *Nature Reviews Neuroscience* 4 (2003): 139–47.

Wright, Crispin. "Moral Values, Projection and Secondary Qualities." *Proceedings of the Aristotelian Society* Supplemental Vol. LXII (1988): 1–26.