

Does the New Wave in Moral Psychology Sink Kant?

This is the author's copy from *The Blackwell Handbook on Naturalism*. Kelly James Clark (ed.). (Wiley Blackwell, 2016), pp. 336-350. Please cite the published version.

[AN]VALERIE TIBERIUS

[T]The longstanding debate between Humeans and Kantians has been fueled by recent work in moral psychology that draws on empirical findings in psychology, neuroscience, and other fields. The empirical turn has not been kind to Kant. The psychologist Jonathan Haidt calls Kant's worship of reason a "rationalist delusion" (2012, 34), and one can easily get the impression that the Kantian project in ethics is as dead as the man who inspired it. Is this true? Has the new wave in moral psychology caused the Kantian boat to capsize? In this chapter, I distinguish some different Kantian claims that are under fire and I assess the evidence against them. Though I am not a Kantian, I argue that rumors of the death of Kantianism have been greatly exaggerated.

[A]Metaethics: The Empirical Critique of Rationalism

[T]The Kantian position is complex and nuanced, and there is no way to do it justice in a single chapter. Instead, this chapter will take for granted some familiarity with Kantianism and will assume a general definition of rationalism about moral judgment that will allow us to make sense of and evaluate the empirical attack on rationalism. Let's define rationalism about moral judgment as the view that the truth of a moral judgment is determined by rational principles. The Kantian rationalist also accepts that moral judgments give us reasons that motivate us insofar as we are rational (independently of our nonrational sentiments or desires). What does the empirical evidence tell us about this picture?

[B]*Emotions Influence and Cause Moral Judgments*

[T]We can start with evidence for sentimentalism, which is often taken to be evidence against rationalism. According to sentimentalism, emotions play an essential role in moral judgment. Sentimentalists differ over whether moral judgments are reports about or expressions of our sentiments, and over which sentiments under what conditions are crucial to moral judgment. Despite these disagreements, sentimentalists all agree that moral judgments are not made true by rational principles. For example, if moral judgments are reports about our emotions, then they are

made true by facts about our emotions (perhaps facts about the emotional response we would have in certain conditions). Or, if moral judgments are *expressions* of our emotions, then whatever story might be told about whether and in what sense moral judgments are truth apt, it couldn't be a story that appeals to rational principles (Blackburn 1984; Gibbard 1992). Sentimentalism of either form presents a picture of moral judgment that is at odds with the rationalist picture, according to which moral judgments do sometimes (when true) track the deliverances of reason.

[TI]The case for sentimentalism, particularly the case defended by Jesse Prinz (2007), often starts with evidence that emotions influence and cause moral judgment, of which there is a good deal.¹ To give one colorful example, here's an experiment that shows the effect of disgust on moral judgment. Psychologists asked participants to answer questions about the moral propriety of four different scenarios, two having to do with incest between first cousins, one having to do with the decision to drive rather than walk to work, and the last having to do with a studio's decision to release a morally controversial film (Schnall et al. 2008). The participants were divided into three different groups: no-stink, mild-stink, and strong-stink, distinguished by the amount of stink – in the form of “commercially available fart spray” sprayed into a nearby trash can – in the environment. The results were that feelings of disgust increase people's tendency to make harsh moral judgments. Other experiments have shown that anger makes people more punitive and harsh in their moral judgments about crimes against persons (Lerner, Goldberg, and Tetlock 1998; Seidel and Prinz 2013).

There is also evidence that emotions cause us to make moral judgments that we would not otherwise make. For example, Thalia Wheatley and Jonathan Haidt hypnotized half the participants in one study to feel disgust when they heard the word “often” and the other half to feel disgust when they heard the word “take.” All the participants then read some scenarios, one of which was this:

[DT]Dan is a student council representative at his school. This semester he is in charge of scheduling discussions about academic issues. He {tries to take/often picks} topics that appeal to both professors and students in order to stimulate discussion. (Wheatley and Haidt 2005)

[T]Participants were randomly assigned to a group that got one or the other of the phrases in the square brackets; no student saw both the phrases in the brackets.

¹ But see Jones (2006) for problems with Prinz's arguments.

[T]To those of us who have not been hypnotized, it doesn't seem like Dan has done anything wrong. Moreover, participants who read the scenario that did not contain their disgust-inducing word did not rate Dan's behavior as wrong. However, for the students who did feel disgust (because they read the scenario with the word that induced disgust in them), there was a tendency to rank Dan's actions as wrong. This is a case in which the people in question would not have made a judgment of moral wrongness at all were it not for the emotion of disgust they experienced.

The fact that emotions *influence* moral judgments does not establish that moral judgments *are* emotional responses, nor even that emotions are an essential part of moral judgment. (Nor does Prinz think it does – he offers this evidence as part of a larger case). This would only be an argument for sentimentalism if the sentimentalist understanding of moral judgment were the *only* way to explain the influence of emotions on moral judgment. Other explanations are possible; it could be that emotions influence moral judgment in the way that wearing rose-colored glasses can influence your judgment about the color of the sky: the glasses sway your judgment, but they're not part of the content of the judgment that the sky is pink. Evidence that emotions cause moral judgments is somewhat more difficult to explain away, but the person who rejects sentimentalism can still argue that when moral judgments are entirely caused by emotions, they are akin to manipulated illusions; after all, it has not been shown that *all* of our moral judgments are such that we would not make them were it not for our emotions.

Perhaps the sentimentalist would have a stronger argument if there were evidence that we simply *cannot* make moral judgments without emotions. We'll consider this possibility in the next section.

[B]Emotions are Necessary for Moral Judgment

[T]Some have thought that psychopaths provide evidence that we cannot make moral judgments without emotions, because (to oversimplify) psychopaths are amoral and they do not experience normal emotions like sympathy or compassion.² If we cannot make moral judgments without emotion, it might seem like emotions are essential to moral judgment in a way that supports the sentimentalist characterization.

² For brevity, I'll talk about "the psychopath," even though this is misleading. In reality, people called psychopaths are a rather varied group who score higher or lower on different diagnostic criteria for psychopathy, and who differ in terms of their capacity to understand moral norms (Aharoni, Sinnott-Armstrong, and Kiehl 2012).

[TI]Psychopathy is a personality disorder characterized by lack of empathy, impulsivity, egocentrism, and other traits. The disorder is often diagnosed by the Psychopathy Checklist, which asks a number of questions that cluster under the headings “aggressive narcissism” and “socially deviant lifestyle” (Hare and Vertommen 2003). Because psychopaths lack empathy, they are of interest to those who think emotions like empathy are essential to moral judgment. Shaun Nichols (2010), for example, thinks that the evidence from psychopathy counts against rationalism because psychopaths do not have defects of reasoning and yet do not seem to make moral judgments in the same way that the rest of us do.³ The basic argument goes this way:

[LN](1)Psychopaths do not make a distinction between moral wrongs and conventional wrongs.

(2) It is the defect to the emotional response system that is responsible for psychopaths’ decreased ability to distinguish moral wrongs from conventional wrongs.

(3) Therefore, a functioning emotional response system is essential to moral judgment.

[T]The conclusion of this argument is taken to be strong evidence for sentimentalism.

[TI]Let’s consider the steps of this argument in detail. The first thing to notice is the importance of the distinction between “moral” and “conventional.” Conventional norms, such as “You shouldn’t go outside in your pajamas,” are different from moral norms in a variety of ways. Moral norms are thought to be more serious and to have wider applicability than conventional norms. Conventional norms are thought to be contingent on an authority (such as a teacher or the law, or, in the case of the pajamas, a culture), and they receive a different kind of justification from moral norms, which are often justified in terms of harm or fairness (Nichols 2004). For example, young children will say that it would be wrong to pull another child’s hair, even if the teacher said it was o.k., because pulling hair hurts, whereas the wrongness of chewing gum in class depends on the teacher’s forbidding it.

It is widely believed that psychopaths don’t really understand this distinction (Blair 1995); that is, psychopaths think of what’s morally wrong as what’s prohibited by the local authority and do not see moral transgressions as more serious than other kinds of violation of rules. This claim is now considerably more controversial than it used to be, but even the latest research confirms that the “affective defect” part of psychopathy does predict poor performance in distinguishing moral from conventional wrongs (Aharoni, Sinnott-Armstrong, and Kiehl 2012). Because they don’t feel bad when others suffer, “they cannot acquire empathetic distress, remorse, or guilt. These emotional

³ See Roskies (2003) for an opposing viewpoint.

deficits seem to be the root cause in their patterns of antisocial behavior” (Prinz 2006). Further, these emotional deficits seem to be responsible for the fact that they don’t make the same kinds of moral judgments that the rest of us do.

Jeanette Kennett (2006) defends rationalism against this sentimentalist attack by arguing that psychopaths have *rational* defects that explain their inability to make real moral judgments or to be motivated by them. According to Kennett, the evidence suggests

[DT]that psychopaths have at best a weak capacity to stand back from and evaluate their desires, to estimate the consequences of their actions, to eschew immediate rewards in favour of longer term goals, to time order, to resolve conflicts among their desires, to find constitutive solutions. To these rational shortcomings we may add that psychopaths frequently choose grossly disproportionate means to their immediate ends or fail to adopt the necessary means to their proclaimed ends. (2006, 77)

[T]She argues that these rational defects play a crucial role in the psychopath’s moral defects. Further, Kennett argues, even if the relevant defect for the psychopath is an emotional defect, this need not embarrass the rationalist. After all, if the rationalist’s primary claim is about what makes moral judgments true, she can be agnostic about what mechanism is operating when we make moral judgments. In other words, the rationalist can accept that a properly tuned emotional system is necessary for making appropriate moral judgments, while maintaining that what is distinctive about moral judgment is that they are justified by rational principles.

[TI]The evidence from psychopathy, then, is not decisive against Kantianism. First of all, psychopathy is a tricky category that includes multiple defects, not all of which are emotional, and it is a complex matter to figure out what is responsible for what. Second, even if we agree that what is important is the ability to distinguish moral from conventional wrongs, and even if we agree that psychopaths with emotional defects are thereby hindered in their ability to make this distinction, we have not shown that reasoning plays no role in the making of moral judgments, nor that moral judgments are not to be assessed in terms of how well they track rational moral principles.

[B]Reasoning is Not Necessary for Moral Judgment

[T]We might take a further stab at defeating rationalism by looking at evidence that no reasoning is required for making moral judgments. Jonathan Haidt, famous for this attack on rationalism, contends that moral judgments cannot be rational judgments because we do not typically engage in reasoning when we make them.

[TI]Haidt (2001) argues that reasoning does not play the causal role in producing moral judgment that we once thought it did. He claims that moral judgments are typically made intuitively, on the

basis of sentiments (which he calls “intuitions”). While it is possible for us to reason about our moral judgments, according to Haidt, this happens rarely. He calls his theory of moral judgment “the social intuitionist model” (SIM), because moral judgments are quick and intuitive, and when reasoning is used, it is usually social reasoning that takes place as people talk and argue with each other to try to figure things out. SIM does allow that individual reasoning or “private reflection” occurs and can affect our judgments, but it is not the typical cause of moral judgment.

We can’t review all of Haidt’s evidence here, but one piece has attracted a great deal of discussion by philosophers and is therefore worth examining in some detail. This is the phenomenon of dumbfounding, which states that we can (and often do) construct justifications for intuitive judgments that were not made by reasoning. This creates the illusion of objective reasoning when what is really happening is *post hoc* rationalization.

Moral dumbfounding happens when a person cannot find any reasons for the moral judgment she makes and yet continues to make it anyway. In the study that introduced the phenomenon, subjects were presented with the following scenario:

[DT]Julie and Mark, who are brother and sister, are traveling together in France. They are both on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy it, but they decide not to do it again. They keep that night as a special secret between them, which makes them feel even closer to each other. So what do you think about this? Was it wrong for them to have sex? (Bjorklund, Haidt, and Murphy 2000)

[T]Most subjects say that the siblings’ behavior is wrong, and they offer reasons for their judgment. They say Mark and Julie may have a deformed child, or that it will ruin their relationship, or cause problems in their family, and so on. But because of the way the scenario is constructed, the interviewer can quickly dispel their reasons, which leads to the state of dumbfounding. According to Haidt, in an interview about his findings, dumbfounding only bothers certain people:

[DT]For some people it’s problematic. They’re clearly puzzled, they’re clearly reaching, and they seem a little bit flustered. But other people are in a state that Scott Murphy, the honors student who conducted the experiment, calls “comfortably dumbfounded.” They say with full poise: “I don’t know; I can’t explain it; it’s just wrong. Period.” (Sommers 2005)

[T]Dumbfounding, the argument goes, shows that most people don't make moral judgments for reasons. Rather, people offer *post hoc* rationalizations of their emotional convictions, and when these rationalizations are undermined, they stick with their convictions anyway.

[TI]Does Haidt's research on the *causes* of moral judgment present problems for Kantians? While Kantian rationalists don't make claims about the typical causes of moral judgment, perhaps they make assumptions that are undermined by Haidt's research. This is the question we will now explore.

First, let's consider whether Haidt and the Kantians mean the same thing by "reasoning." If they do not, then Haidt's challenge won't necessarily undermine rationalism. Haidt's picture of moral reasoning seems rather different from what Kantians take moral reasoning to be. Haidt concedes rare cases in which people "reason their way to a judgment by sheer force of logic" (Haidt and Bjorklund 2008, 819). But this is a caricature: moral rationalists do not typically think that we reason ourselves into moral positions by sheer force of logic. One tool of moral reasoning that Kantians think is particularly important is universalization: when we're unsure what to do, we should ask ourselves whether the intention of our action requires making a special exception for ourselves, or whether it is an intention that we think is acceptable for everyone to have. This sort of reasoning is not the sheer force of logic.

Still, there is a good case against the idea that we often engage in slow, deliberate conscious reasoning when we make moral judgments, and Kantians do seem to think that this sort of reasoning is important. But Kantians don't claim that we routinely engage in this kind of reasoning, nor do they claim that our moral judgments are typically caused by reasoning. The Kantian holds that some moral judgments (the correct ones) are backed up by rational principles and that we *could* – if we needed to – use our rational capacities (such as universalization) to justify them. This doesn't require that we always, or even typically, reason to our moral judgments. Indeed, it would be a waste of our precious cognitive resources to do this, since most of our moral judgments are fairly easy and uncontroversial. When you read in the paper that someone has stolen billions of dollars from a retirement fund, or that someone has sold 10-year-old children into sexual slavery, you find yourself making moral judgments about these people. There's no need for reasoning here; reasoning would be wasted effort, since the cases are so obvious. Reasoning, however, is needed in cases of conflict when we aren't sure what to do. For example, what should you do if you rear end someone's car in a parking lot when no one else is looking, causing a small amount of damage? Of if you discover your very good friend cheating on his taxes? Your automatic judgment might be to do

nothing (to drive away, to turn a blind eye for the sake of your friendship), but if you think about it, you might conclude that this wouldn't be right.

So, Kantians do not need to assume that our moral judgments are always caused by reasoning. What they do assume is that moral judgments can be justified and that reasoning – when it's done well – produces justification. On the Kantian view, then, while we could reason our way to a moral judgment if we needed to, it's not a problem if many of our moral judgments are fairly automatic. Haidt concedes that we sometimes arrive at judgments through private reasoning. He also thinks that we engage in social reasoning: reasoning with each other in the form of argument and gossip. Kantians do not need to assume that moral reasoning is always private. Indeed, reasoning with each other might help us overcome our biases, so that we can be more impartial and better universalizers.

[B] We Are Not Reflective Creatures

[T] Dumbfounding raises another problem, one concerning the Kantian assumption about the kinds of agents we human beings are. The assumption is that we are reflective creatures who aim to act for reasons that we can endorse as justifications of our actions. According to Christine Korsgaard (1996), ethics is a response to the practical problem that confronts us as reflective creatures, a problem that any normal human being will recognize: the problem of *what to do*. When confronted with a quandary about what to do, “[t]he reflective mind cannot settle for perception and desire, not just as such. It needs a *reason*. Otherwise, at least as long as it reflects, it cannot commit itself or go forward” (Korsgaard 1996, 93). When our desires or inclinations conflict with each other or with what we deem morally right, we need a conclusive reason to go one way rather than another, and that reason can't be another desire or inclination. Rather, reasons that answer this very practical problem must be underwritten by a rational principle or law.

[TI] Do we have reflective minds like this? Do we need a reason or a consideration that provides a justification for our action? And is it impossible for this reason to depend on a desire or a sentiment?

Dumbfounding seems to afford evidence against the first claim. A “comfortably dumbfounded” person seems perfectly happy without reasons for their moral judgments. If this is correct, it contradicts not just the view of a few Kantians, but also a general and very widely held assumption about moral judgments made by most philosophers (Kennett and Fine 2008). Introductory ethics textbooks and Introduction to Ethics courses often begin by distinguishing moral judgments from mere judgments of taste by pointing out that we have justifying reasons for the former but not for

the latter. “You morally ought not to eat kale” is importantly different from “Kale tastes awful,” for example, because the first implies that there are some reasons that make eating kale a bad idea, not just for someone who doesn’t like kale or is allergic, but for anyone. Sentimentalists and rationalists alike have agreed with this, and sophisticated sentimentalists have bent over backwards trying to accommodate the idea that we think that our moral judgments should be backed up with reasons (Blackburn 1998; Gibbard 1992). So, it’s worth thinking about what dumbfounding really shows. To cut to the chase, I don’t think the evidence shows that there’s anything wrong with the conception of a moral judgment as involving justifying reasons (Tiberius 2013).

To doubt that people care about the reasons for their moral judgments, we would have to see that people:

[LN](1) Do not think there are reasons for their judgments (as opposed to thinking that there are reasons, but not knowing what the reasons are).

[T]and:

(2) Are entirely unperturbed by this fact and have no inclination to reconsider their judgments.

[T]We do not know that (1) is true from the existing evidence. The claim “I don’t know why it’s wrong, it’s just wrong” is ambiguous between “It’s wrong for some reason that I don’t know” and “There’s no reason it’s wrong, it just is!” It is also possible that people think some things are self-evidently wrong and that they take this self-evidence to be a reason. We also do not know that (2) is true. The fact that people are unwilling to change their judgments in an artificial interview setting does not mean that they never feel pressure to change them. It may take time for shaken confidence to cause someone to change their judgments. We would need evidence of how people respond to the challenge to provide justification over the long term and in real-life settings. Further, even if a person were *never* to change her judgment when she discovered she had no reasons for it, this would not show that she didn’t care about reasons. If people don’t change their judgments when nothing is at stake or when they don’t feel like thinking about it at the time, this does not count against the claim that people care whether their judgments are justified when it counts.

[TI]The argument I’ve offered only establishes that *it hasn’t yet been shown* that people don’t care about reasons. But it is unlikely that studies will ever show that nobody has a conception of a moral judgment as one that is held for justifying reasons, unlike judgments of mere taste. Reflection on moral disagreement indicates that people often think of moral judgments as different from judgments of taste. People cast their votes for political candidates who agree with them about the

morality of abortion, they march in the streets to protest or support gay marriage, they donate money to charities that help stray animals or feed hungry people. It's hard to imagine that such morally committed people don't think there are reasons for the moral judgments upon which they are acting. Of course, such thoughts are not always correct, and people don't always care about justification. But the idea that moral judgments are different from judgments of taste (like "Kale tastes awful") with respect to their being supported by reasons is not just the crazy idea of a few philosophers: it is a pervasive idea in general. Indeed, many philosophers (Kantians and non-Kantians alike) would argue that those who deny that moral judgments require justification do not grasp what is distinctive about them.

We have been discussing the first of the Kantian claims mentioned earlier, namely, that the reflective mind needs a reason. The second claim is that the reason cannot be ultimately explained by a desire or a sentiment. The rationalist thinks that reasons (genuine reasons, ones that justify what we do) must be backed up by rational principles, not desires or sentiments. Their "authority" or justificatory weight cannot ultimately be grounded in personal preference, such as how we feel or what we care about. This argument seems to assume something about our psychology, namely, that we are reflective creatures of a particular kind: ones whose conviction that we have reasons to act rests fundamentally on the idea that there are universal principles that support these reasons. In other words, reflective creatures like us need to find something at the bottom of the pile of reasons that puts an end to our questions, something like a purely rational principle or moral law. I don't know of any empirical studies that bear directly on this claim, but it is relevant that there are many Humean philosophers who believe that all reasons are ultimately explained by desires and whose moral convictions do not seem to be undermined by this belief (e.g., Schroeder 2007). Once we distinguish the claim that moral judgments are supported by *some* justifying reasons from the claim that these reasons must be backed up by rational principles, it seems quite plausible to think that people could live with the absence of rational principles that make our moral judgments true. If the argument that there must be such principles if there are to be any reasons at all relies on a claim about our psychology, then the fact that we don't need rational principles would deflate it.

The Kantian, however, would deny that this argument has anything to do with our psychology. What is relevant, she'll say, is what we are like insofar as we are rational; it is only by thinking about what we are like insofar as we are rational beings that we can uncover the metaphysical nature of reasons. As Kant himself puts it, "Since moral laws must hold for every rational being as such, our principles must...be derived from the universal concept of a rational being as such" (Kant 1785/2002, 213/412). The Kantian argument is metaphysical, not empirical: the relevant premise

is that a consideration cannot count as a *reason* unless it would be sanctioned by a rational law, and that is just part of what it is for something to be a reason. Indeed, Kant thought it possible that we are not the kind of creature to whom the moral law applies, and on this view it is possible that there are moral reasons that could never be part of the explanation of human action.

True, the Kantian argument is an argument about what a reason is (as opposed to what we take reasons to be as a matter of our psychology). But if we cannot recognize ourselves (even our best, most rational selves) in the description of rational agents in this argument, then it's not clear why we should care about reasons as Kantians see them. If we are not at all like the reflective creatures Korsgaard and Kant discuss (creatures who seek principled reasons for action that can put an end to our questioning), then the view of reasons they are discussing is not relevant to us. This by itself would not prove that the Kantian conception of a reason is wrong, exactly, but it would make us question what the point of it is.

[B]The Threat to Rationalism: Conclusion

[T]Does the empirical evidence really provide a fundamental challenge to rationalism? What does seem to be threatened is a picture according to which we always arrive at our moral judgments by engaging in rational reflection and we are then motivated to act on these judgments by the sheer recognition of their rational status. It's unlikely that Kant held this extreme view. Whether he did or not, it seems to me that the most important Kantian assumptions about reasoning are compatible with much of the empirical research, because Kantians could be satisfied with a limited causal role for reasoning. Indeed, Kantians could even admit that emotions play an important role in producing moral judgments, because this is compatible with thinking that reasoning is how we *justify* our moral judgments and that rational principles are at the foundation of these justifications. As long as we can reject an emotionally caused but unjustified judgment, the Kantian view would not be endangered. Further, as long as reasoning can succeed in justifying our moral judgments, reasoning needn't *always* have this purpose.

[TI]Whether Kantian reasoning can justify our moral judgments depends on some deep issues in metaethics. In particular, it depends on whether there really are any rational principles that provide a foundation for our moral reasons. This is one of the fundamental philosophical disputes between the sentimentalist and the rationalist, and skepticism about the existence of principles of practical reason that have the authority to justify our moral judgments is my own reason for not being a Kantian. But this debate ultimately concerns philosophical questions about the nature of

practical reasons, not psychological facts about the causes of moral judgment. As far as the empirical challenge goes, the door for rationalism is still open.

[A]Normative Ethics: Empirical Challenges to Kantian Normative Positions

[T]Again, we will have to take for granted some familiarity with Kantian normative theory, whose core idea is that an action is morally permissible if its maxim (the underlying intention or subjective principle of an action) conforms to the categorical imperative. Two formulations of the categorical imperative have been particularly fruitful: the universal-law formulation, which tells us to act only on that maxim we could will to be a universal law, and the end-in-itself formulation, which enjoins us to treat people never merely as a means but always as ends in themselves.

[B]Reasoning Does Not Lead to Kantian Normative Conclusions

[T]Because Kantianism looks to the intentions underlying our actions to determine their moral status, it stands in contrast to consequentialist theories, according to which the amount of good or value produced by an action is what determines its rightness. This feature of Kantian normative theory entails that it is impermissible to promote the greater good if the maxim of the good-promoting action is one that cannot be universalized or one that fails to respect rational nature. It is this feature of Kantian normative theory that Joshua Greene, a neuroscientist and philosopher, attacks.

[TI]Greene argues that “characteristically deontological moral judgments” (such as the judgment that it is impermissible to sacrifice one person’s life for the sake of saving the lives of several others) are emotional rather than rational. Greene argues that different moral judgments about cases (intuitions) are caused in different ways, and that, together with some assumptions about when different mental processes are reliable and when not, we have good reason to discount some of our moral intuitions, in particular, the Kantian ones.

To understand Greene’s argument, we must confront the trolley (Foot 1967/2002; Thomson 1976). Consider the following cases:

- [LB]*Footbridge*: A trolley is hurtling down the tracks. There are five innocent people on the track ahead, and they will be killed if the trolley continues going straight. You are an innocent bystander (i.e., not an employee of the railroad, etc.) standing next to a large man on a footbridge

spanning the tracks. The only way to save the five people is to push this man off the footbridge and into the path of the trolley. What would you do?

- *Footbridge Switch*: A trolley is hurtling down the tracks. There are five innocent people on the track ahead, and they will be killed if the trolley continues going straight. You are an innocent bystander (i.e., not an employee of the railroad, etc.) standing next to a switch that controls a trap door, which opens on to the tracks. There is a large man on the trap door. The only way to save the five people is to pull the switch, thus dropping the large man into the path of the trolley. What would you do?

[T]The only difference between *Footbridge* and *Footbridge Switch* is that in the first case you have to push a man to his death, whereas in the second case you pull a switch that has the same result. Either way, the man is killed by the train: one will die and five will live (and if you don't act, five will die and one will live). Despite the similarity in the numbers, people tend to feel differently about these cases. Most people (63%) say that it's morally permissible to pull the switch in *Footbridge Switch*, but only 31% think that it is morally permissible to push the large man in *Footbridge* (Greene et al. 2009). Why this difference?

[TI]Greene thinks that the difference between *Footbridge* and *Footbridge Switch* can't be explained rationally. Think about it: in one case, you are right next to the man (close enough to touch him), and in the other case, you are a little farther away, but you can still make him fall into the train by pulling a switch. How could the tiny difference of physical distance make the difference between its being morally o.k. to kill him and its being morally wrong to kill him? Assuming that this tiny difference cannot make a real moral difference, instead of trying to explain our intuitions rationally, Greene offers a causal explanation. He claims that the different intuitions in *Footbridge Switch* and *Footbridge* are explained by the fact that we have two different cognitive systems in our brains. In short, one system, which is emotional and automatic, is engaged when thinking of physically touching the man, and it elicits the emotional judgment that we should not push the man into the train. The other, more reflective, system is engaged in response to reading the relatively cold *Footbridge Switch* case; since our emotions are not engaged, this more reflective system elicits the judgment that we should pull the switch in order to save more people. Let's consider this in a little more detail.

The theory that there are these two systems in the brain is called dual-process theory. System 1 is automatic, emotional, and quick. System 2 is controlled, deliberate, and slow. System 2 is what we normally think of as conscious reasoning or "thinking," but both systems produce judgments.

Greene analogizes the two cognitive systems to the automatic and manual modes on a camera. If your camera is on automatic, you can take pictures quickly, but you sacrifice quality. If your camera is on manual, you have more flexibility to cope with, say, different lighting conditions, but you sacrifice speed because of the conscious effort required to set things up (Greene 2014). Greene argues that the two processes in dual-process psychology tend toward different kinds of moral judgment: System 1 produces “characteristically deontological” judgments (concerning rules, rights, and duties), while System 2 produces “characteristically consequentialist” judgments (concerning the greatest benefit to the greatest number). The judgment that we should not sacrifice the one to save the lives of the others is taken to be characteristically deontological because it prioritizes the importance of not violating a man’s right to life by killing him over maximizing the number of saved lives; the judgment that we should sacrifice the large man for the sake of the others’ lives is taken to be characteristically consequentialist.

Using dual-process theory, Greene hypothesized an explanation for why people tend to make different judgments in *Footbridge Switch* and *Footbridge*. First, our automatic, emotional system of judgment will be triggered by the close and personal nature of the action in *Footbridge* (you have to touch the man to push him on to the tracks), and this system will cause us to judge that we should not push the man. Second, in the *Footbridge Switch* case, without any emotional trigger, our rational, calculative system will determine our judgment and we will consider the outcomes more rationally, thus leading us to say that it would be right to pull the switch. Some of the evidence in support of this hypothesis is neuroscientific: researchers can see from fMRI scans that the parts of the brain that are more active when people judge that it would be wrong to push the large man on to the tracks are the parts of the brain that are associated with emotional activity. And in many cases, patients with emotional deficits due to brain injuries are more likely to make consequentialist judgments. (This research gave rise to the fun title “Consequentialists Are Psychopaths” (Schwitzgebel 2011)). Further, when people are given more time to deliberate or are encouraged to reflect, they are more likely to make consequentialist judgments about a case (Greene 2014).

Suppose we accept the description of our psychology put forward by this research: we agree, for the sake of argument, that consequentialist judgments are associated with conscious/controlled reasoning processes, while deontological judgments are associated with automatic/emotional cognitive processes. What should we conclude? Selim Berker, a critic of Greene’s work, has argued that there is no bridge from the “is” of dual-process theory to the “ought” of ethics. He argues that the scientific evidence about the causes of our moral judgments is irrelevant to any claims about

which judgments are right or wrong (Berker 2009). According to Berker, in order to derive any conclusion about which judgment is trustworthy, we would have to rely on moral intuitions about what sorts of features of the world our judgments *ought* to be sensitive to. Without such assumptions, we couldn't argue that the cool and calculated judgment is *better* than the judgment that is influenced by the emotions caused by the personal nature of the action.

Berker is right. We must make normative assumptions (about what our judgments ought to be sensitive to) in order to derive a normative conclusion (about which judgments we can trust). But Greene does not deny this. Greene argues that the scientific evidence *together with* normative assumptions about what counts as good judgment support the conclusion that the consequentialist intuitions are better or more reliable. The argument in favor of trusting consequentialist intuitions depends on the assumption that our judgments *should* (normative term) be insensitive to "mere personal force" or "mere special proximity." Judgments that respond to these considerations alone – absent any other relevant considerations – are biased by irrelevant information. The scientific research supports the claim that our nonconsequentialist judgments are responding just to personal contact and proximity. The assumption that these features – the mere facts of proximity – are morally irrelevant is the sought-after extra, moral, premise.

Judith Jarvis Thomson agrees that there is an uncomfortable incongruity between our willingness to flip the switch and our unwillingness to push the large man. "Well, in *Footbridge Switch* I don't have to use both hands!" doesn't seem like much of a moral argument. But instead of concluding that we ought to count both pulling the switch and pushing the large man as the right things to do (as Greene does), Thomson argues that we ought to judge both actions wrong (Thomson 2008). She introduces another case in which *you* are on a side track that bypasses the five innocents stuck on the main track. In this case, you can pull a switch that would divert the train on to your side track, killing you and thereby saving the five innocent people. Sure, it would be nice if you did this, but, Thomson argues, you are not morally required to sacrifice yourself to save the five. Sacrificing yourself to save five others would be heroic or supererogatory (beyond the call of duty), but not required. Further, if you aren't required to sacrifice yourself, then the large man in *Footbridge Switch* isn't required to sacrifice himself to save five, either. By putting him on to the track, you make him do something that he isn't required to do, and something that you yourself would not do. This seems wrong, and it is plausible that it seems so on the grounds of a principle governing obligations of self-sacrifice (rather than merely because of a response to the anticipated personal contact). This new twist on the case makes us think that maybe we ought to make the same

judgment in *Footbridge Switch* and *Footbridge*: in neither case does morality require you to sacrifice a person to save five.

Thompson's conclusion is even more anticonsequentialist than the usual intuitions people have about these cases: we should pay even less attention to the cost-benefit analysis than we originally thought. Thomson's example shows that even if it is true that emotional reactions are driving the anticonsequentialist conclusion about *Footbridge*, this does not mean that there isn't a rational explanation for the anticonsequentialist verdict. Thomson's argument raises the stakes, because many will find it unintuitive to think that it is always impermissible to sacrifice one to save five. Nevertheless, some Kantians might want to accept that this is, in fact, the principled solution (or not: see Hill 1992).

[B]The Motive of Duty

[T]According to Kant, the only actions that are morally worthy are actions done from the motive of duty. The motive of duty is the motive of a good will, which is the kind of thing that is always good to have. It is, Kant thinks, unconditionally good. When we act from duty, we act with the intention – not just an inclination, but the “summoning of every means” – to do the right thing, because it is the right thing, no matter what we feel like doing (Kant 1785/2002, 196/394). A person who acts from the motive of duty does the right thing in virtue of her recognition that morality demands it and her background rational commitment to do what is morally required of her. The motive of duty is important not only because it is the only morally good motive, but also because morality applies only to beings who are capable of being motivated by duty (i.e., according to Kant, rational beings).

[TI]The claim that we could be motivated by a pure sense of duty, unconnected to our desires, has also been the subject of empirical critique. Schroeder, Roskies, and Nichols (2010) argue that given what we know about the causal pathways to voluntary behavior in the brain, it is difficult to sustain the view that we are motivated by our beliefs about what's morally right and wrong. They point out that there is one model for voluntary behavior that is brought about by higher cognitive centers independently of desire, but that this model is of the pathological behavior of Tourette syndrome.

This challenge assumes that to be motivated by duty is to be motivated by a purely cognitive state, such as a belief. Is the motive of duty such a state? Kant sometimes describes the motive of duty as respect for the moral law (Kant 1785/2002, 202/400), and respect is not obviously a belief. As T.M. Scanlon points out, if we understand desires very inclusively, as “pro-attitudes,” then there is still room for the motive of duty, because many pro-attitudes, such as “duty, loyalty, or pride, as well as an interest in pleasure or enjoyment,” can be brought about by reasoning (1998, 37). If respect is a

kind of pro-attitude, and if thinking about the requirements imposed on us by the moral law can cause this attitude of respect (in the way that recognizing beauty in nature can bring about the feeling of awe), then the motive of duty is a pro-attitude that can be brought about by reasoning. This suggestion does not by itself vindicate the motive of duty, but it does provide an avenue of defense for the Kantian on this point. Further defense of the idea would have to show that the recognition of moral requirements can indeed give rise to the right kind of motivational state.

[B]The Threat to Kantian Normative Ethics: Conclusion

[T]Some Kantian ideas about how we ought to behave are threatened by the empirical wave in moral psychology. It is a problem for a principle-based moral theory if there is nothing more to say about why we should behave differently in two similar cases except that we feel differently about the two cases. If we are goal-directed creatures for whom voluntary action is typically acting on our desires, this is a problem for the hypothesis that morally worthy action is motivated by a pure sense of duty that is entirely divorced from our desires. But, as we have seen, there may be more to say about trolley cases that is favorable to a nonconsequentialist picture. And there are other ways of thinking about the motive of duty than as a purely rational force that has no home in goal-directed psychology.

[TI]More importantly, however, there are many aspects of Kant's normative theory that are not targeted by the empirical wave. That we should treat each other as fundamentally valuable beings who have our own projects and plans for achieving them; that we should respect other people's judgments about their own lives; that recognizing the worth of other people requires that we spend some of our time helping them pursue their projects and making it possible for them to live autonomously: these are first-order ethical ideas worth taking seriously. One could reject Kant's metaphysical views about rationality, his draconian views about duty, and perhaps even his most dogmatic anticonsequentialist conclusions, but still think he has a lot of important ideas about how we ought to treat other people, morally speaking. It would be a shame if the current wave, which has produced so much valuable work in moral psychology, were to swamp these good ideas.

[A]Acknowledgments

[T]This chapter includes material from my book *Moral Psychology: A Contemporary Introduction* (Tiberius 2014). I thank the publisher for permission to include this material here. I would also like to thank Kelly James Clark, Josh Greene, and Connie Rosati for helpful comments and discussion.

[A]References

- [REF]Aharoni, E., Sinnott-Armstrong, W., and Kiehl, K.A. (2012). Can Psychopathic Offenders Discern Moral Wrongs? A New Look at the Moral/conventional Distinction. *Journal of Abnormal Psychology* 121(2): 484.
- Berker, S. (2009). The Normative Insignificance of Neuroscience. *Philosophy & Public Affairs* 37(4): 293–329.
- Bjorklund, F., Haidt, J., and Murphy, S. (2000). Moral Dumbfounding: When Intuition Finds No Reason. *Lund Psychological Reports* 2(1): 29.
- Blackburn, S. (1984). *Spreading the Word: Groundings in the Philosophy of Language*. Oxford: Clarendon Press.
- Blackburn, S. (1998). *Ruling Passions*. Oxford: Clarendon Press.
- Blair, R.J.R. (1995). A Cognitive Developmental Approach to Morality: Investigating the Psychopath. *Cognition* 57(1): 1–29.
- Foot, P. (1967/2002). The Problem of Abortion and the Doctrine of Double Effect. In *Virtues and Vices and Other Essays in Moral Philosophy*. Cambridge: Cambridge University Press.
- Gibbard, A. (1992). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Oxford: Oxford University Press.
- Greene, Joshua D., et al. (2009). "Pushing moral buttons: The interaction between personal force and intention in moral judgment." *Cognition* 111.3: 364-371.
- Greene, J. (2014). Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics. *Ethics* 124(4): 695–726.
- Haidt, J. (2001). The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment. *Psychological Review* 108(4): 814–834.
- Haidt, J. (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. London: Penguin.
- Haidt, J. and Bjorklund, F. (2008). Social Intuitionists Answer Six Questions about Morality. *Moral Psychology* 2: 181–217.
- Hare, R.D. and Vertommen, H. (2003). *The Hare Psychopathy Checklist-Revised*. North York: Multi-Health Systems.

- Hill, T. (1992). Making Exceptions without Abandoning the Principle: Or How a Kantian Might Think about Terrorism. In *Dignity and Practical Reason in Kant's Moral Theory*. Ithaca: Cornell University Press.
- Jones, K. (2006). Metaethics and Emotions Research: A Response to Prinz. *Philosophical Explorations* 9(1): 45–53.
- Kant, I. (1785/2002). *Groundwork for the Metaphysics of Morals*, translated by A. Zweig and T.E. Hill. Oxford: Oxford University Press.
- Kennett, J. (2006). Do Psychopaths Really Threaten Moral Rationalism? *Philosophical Explorations* 9(1): 69–82.
- Kennett, J. and Fine, C. (2008). Will the Real Moral Judgment Please Stand Up? *Ethical Theory and Moral Practice* 12(1): 77–96.
- Korsgaard, C.M. (1996). *The Sources of Normativity*. Cambridge: Cambridge University Press.
- Lerner, J.S., Goldberg, J.H., and Tetlock, P.E. (1998). Sober Second Thought: The Effects of Accountability, Anger, and Authoritarianism on Attributions of Responsibility. *Personality and Social Psychology Bulletin* 24(6): 563–574.
- Nichols, S. (2004). *Sentimental Rules: On the Natural Foundations of Moral Judgment*, vol. 13. Oxford: Oxford University Press.
- Nichols, S. (2010). How Psychopaths Threaten Moral Rationalism: Is It Irrational to be Amoral? In *Moral Psychology: Historical and Contemporary Readings*, edited by Thomas Nadelhoffer, Eddy Nahmias, and Shaun Nichols. Oxford: Blackwell.
- Prinz, J. (2006). The Emotional Basis of Moral Judgments. *Philosophical Explorations* 9(1): 29–43.
- Prinz, J.J. (2007). *The Emotional Construction of Morals*. Oxford: Oxford University Press.
- Roskies, A. (2003). Are Ethical Judgments Intrinsically Motivational? Lessons from ‘Acquired Sociopathy.’” *Philosophical Psychology* 16(1): 51–66.
- Scanlon, T.M. (1998). *What We Owe to Each Other*. Cambridge, MA: The Belknap Press of Harvard University Press.
- Schnall, S., Haidt, J., Clore, G.L., and Jordan, A.H. (2008). Disgust as Embodied Moral Judgment. *Personality and Social Psychology Bulletin* 34(8): 1096–1109.
- Schroeder, M.A. (2007). *Slaves of the Passions*. Oxford: Oxford University Press.

- Schroeder, T., Roskies, A., and Nichols, S. (2010). Moral Motivation. In *The Moral Psychology Handbook*, edited by J. Doris, pp. 72–110. Oxford: Oxford University Press.
- Schwitzgebel, E. (2011). Bartels and Pizarro: Consequentialists Are Psychopaths. *The Splintered Mind*. Available from: <http://schwitzsplinters.blogspot.com/2011/09/bartels-and-pizarro-consequentialists.html> (last accessed July 15, 2015).
- Seidel, A. and Prinz, J. (2013). Sound Morality: Irritating and Icky Noises Amplify Judgments in Divergent Moral Domains. *Cognition* 127(1): 1–5.
- Sommers, T. (2005). “Interview with Jonathan Haidt.” *The Believer*.
www.believermag.com/issues/200508/?read=interview_haidt.
- Thomson, J.J. (1976). Killing, Letting Die, and the Trolley Problem. *The Monist* 59(2): 204–217.
- Thomson, J.J. (2008). Turning the Trolley. *Philosophy & Public Affairs* 36(4): 359–374.
- Tiberius, V. (2013). In Defense of Reflection. *Philosophical Issues* 23(1): 223–243.
- Tiberius, V. (2014). *Moral Psychology: A Contemporary Introduction*. London: Routledge.
- Wheatley, T. and Haidt, J. (2005). Hypnotic Disgust Makes Moral Judgments More Severe. *Psychological Science* 16(10): 780–784.