SELIM BERKER

The Normative Insignificance of Neuroscience

I. INTRODUCTION

No doubt when historians of science look back on the first decade of the twenty-first century, they will dub it "The Age of the fMRI." Functional magnetic resonance imaging has revolutionized the empirical study of the human mind, leading to far-reaching changes in the research paradigms of psychology, economics, and (especially) the burgeoning field of neuroscience; by one estimate, an average of eight peer-reviewed articles employing fMRI were published per day in 2007.[1] So perhaps it was inevitable that empirically minded philosophers would take some of these fMRI studies to have profound implications for philosophy. Indeed, it has recently been argued that the groundbreaking research by psychologist Joshua D. Greene and colleagues into the neural bases for

Earlier versions of this article were presented at the Arché Philosophical Research Centre in St. Andrews, at the Harvard Kennedy School's Safra Center for Ethics, at the Harvard Humanities Center's Cognitive Theory and the Arts Seminar, and at the 2009 Rocky Mountain Ethics Congress in Boulder, Colorado (where my commentator was Daniel Demetriou). Joshua D. Greene was kind enough to attend both sessions at Harvard and to offer clarifications and replies. For written comments on earlier drafts, I am indebted to Jacob Beck, Carol Berker, Tyler Doggett, Frances Kamm, Christine Korsgaard, Seana Shiffrin, Judith Jarvis Thomson, and Hasko Vonkriegstein (on behalf of the moral psychology reading group at Toronto University). For helpful comments and discussion, many thanks as well to Arthur Applbaum, Sharon Berry, Tim Button, Yuri Cath, Colin Chamberlain, Norman Daniels, Daniel Demetriou, Tom Dougherty, Matti Eklund, Nir Eyal, Michael Frazer, Johann Frick, Micha Glaeser, Ned Hall, Ulrike Heuer, Jonathan Ichikawa, Ole Koksvik, Arnon Levy, Louis Menand, Oded Na'aman, Dilip Ninan, François Recanati, Simon Rippon, T. M. Scanlon, Susanna Siegel, Alison Simmons, Leo Ungar, Manuel Vargas, Alex Voorhoeve, Glen Weyl, Liane Young, and Elia Zardini. Finally, I am grateful to the Editors of *Philosophy & Public Affairs* for a number of extremely helpful suggestions.

1. Jonah Lehrer, "Picture Our Thoughts: We're Looking for Too Much in Brain Scans," *The Boston Globe* (August 17, 2008).

our moral intuitions should lead us to change our opinions about the trustworthiness of those intuitions. Crudely put, Greene and his colleagues think there are two warring subsystems underlying our moral intuitions: the first makes use of emotional neural processes and generates the sorts of judgments typically associated with deontological positions in ethics; the second makes use of more cognitive neural processes and generates the sorts of judgments typically associated with utilitarian/consequentialist positions in ethics; and the two subsystems duke it out for one's overall moral verdict about a given case.[2] By itself, this claim is merely an empirical hypothesis about what, as a matter of fact, causes us to make the moral judgments that we do make. However, Peter Singer and Greene himself have argued that this empirical hypothesis, if true, would also yield conclusions about the sorts of moral judgments that we *should* make. In particular, Singer and Greene think that the truth of Greene's empirical hypothesis would give us good grounds to discount our deontological intuitions about cases, but *not* to discount our utilitarian/consequentialist intuitions about cases.[3]

In this article I wish to scrutinize this last claim. More specifically, I will argue that once we separate the bad arguments for why Greene et al.'s empirical research has normative implications from the better arguments for that conclusion, we can see that the neuroscientific results are actually doing no work in those better arguments. Or to put my central contention most provocatively: either attempts to derive normative implications from these neuroscientific results rely on a shoddy inference, or they appeal to substantive normative intuitions (usually about what sorts of features are or are not morally relevant) that render the neuroscientific results irrelevant to the overall argument. However,

2. Joshua D. Greene, R. Brian Sommerville, Leigh E. Nystrom, John M. Darley, and Jonathan D. Cohen, "An fMRI Investigation of Emotional Engagement in Moral Judgment," *Science* 293 (2001): 2105–8; Joshua D. Greene, Leigh E. Nystrom, Andrew D. Engell, John M. Darley, and Jonathan D. Cohen, "The Neural Bases of Cognitive Conflict and Control in Moral Judgment," *Neuron* 44 (2004): 389–400; and Joshua D. Greene, Sylvia A. Morelli, Kelly Lowenberg, Leigh E. Nystrom, and Jonathan D. Cohen, "Cognitive Load Selectively Interferes with Utilitarian Moral Judgment," *Cognition* 107 (2008): 1144–54.

3. Peter Singer, "Ethics and Intuitions," *The Journal of Ethics* 9 (2005): 331–52; Joshua D. Greene, "From Neural 'Is' to Moral 'Ought': What Are the Moral Implications of Neuroscientific Moral Psychology?" *Nature Reviews Neuroscience* 4 (2003): 847–50; and Joshua D. Greene, "The Secret Joke of Kant's Soul," in *Moral Psychology, Vol. 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, ed. Walter Sinnott-Armstrong (Cambridge, Mass.: MIT Press, 2008), pp. 35–79.

my conclusions here are not entirely negative: although I am skeptical about the prospects for deriving normative implications from neural facts about how we happen to reach moral verdicts, in the article's final section I sketch a way in which neuroscience could play a more indirect role in sculpting our normative conclusions.

It should be clear that much is at stake in this debate. Obviously if Greene's and Singer's arguments for why we should privilege our consequentialist intuitions over our deontological ones were sound, there would be far-reaching implications for contemporary debates in first-order ethics. But the implications are even wider than that. So far the only sorts of philosophical intuitions that have been systematically studied using brain-imaging technology have been moral intuitions about cases, and even then moral intuitions about only a small class of cases. However, it is only a matter of time before fMRI-based studies of other varieties of philosophical intuitions are conducted—only a matter of time before someone, somewhere, studies what parts of the brain light up when the typical person has intuitions about general phenomena such as knowledge or free will, or about specific puzzles such as Newcomb's problem or the sorites paradox. One can almost see how this research will go. First, no doubt, someone will hypothesize that there are two separate systems vying for one's overall verdict about whether a given hypothetical scenario counts as an instance of knowledge, or causation, or free action. Then, no doubt, someone (possibly the same person) will conclude that the empirical evidence for this hypothesis gives us good reason to discount the verdicts of one of those systems but not the other. So if the sorts of arguments offered by Greene and Singer are successful, they have the potential to radically alter how we go about adjudicating whether philosophical intuitions of a given sort are reliable—and, by extension, to radically alter the methodology with which we go about arguing for first-order philosophical claims.

Enough speculation, though, about what sorts of empirical research may or may not be conducted in the future, and about what sorts of philosophical arguments may or may not be offered on the basis of that research. My task here will be to focus on the neuroscientific research that has been conducted into the physiological basis for our moral intuitions about hypothetical cases, and on what the normative implications of that research might be. Thus the first order of business will be to summarize the essential details of Greene et al.'s research.

II. NEUROSCIENTIFIC RESULTS

Greene and his colleagues chose to focus their empirical studies on our moral intuitions about a certain class of cases made famous by Philippa Foot and Judith Jarvis Thomson.[4] Consider the following scenario (here I use the exact wording employed by Greene et al. in their studies):

> *Trolley Driver Dilemma*: "You are at the wheel of a runaway trolley quickly approaching a fork in the tracks. On the tracks extending to the left is a group of five railway workmen. On the tracks extending to the right is a single railway workman.
>
> If you do nothing the trolley will proceed to the left, causing the deaths of the five workmen. The only way to avoid the deaths of these workmen is to hit a switch on your dashboard that will cause the trolley to proceed to the right, causing the death of the single workman.
>
> Is it appropriate for you to hit the switch in order to avoid the deaths of the five workmen?"[5]

Assuming that our topic is *moral* appropriateness, most people judge that it *is* appropriate to hit the switch. However, contrast that case with the following (again, the wording is Greene et al.'s):

> *Footbridge Dilemma*: "A runaway trolley is heading down the tracks toward five workmen who will be killed if the trolley proceeds on its present course. You are on a footbridge over the tracks, in between the approaching trolley and the five workmen. Next to you on this footbridge is a stranger who happens to be very large.
>
> The only way to save the lives of the five workmen is to push this stranger off the bridge and onto the tracks below where his large body will stop the trolley. The stranger will die if you do this, but the five workmen will be saved.

4. See Philippa Foot, "The Problem of Abortion and the Doctrine of Double Effect," *Oxford Review* 5 (1967): 5–15; Judith Jarvis Thomson, "Killing, Letting Die, and the Trolley Problem," *The Monist* 59 (1976): 204–17; and Judith Jarvis Thomson, "The Trolley Problem," *Yale Law Journal* 94 (1985): 1395–415. All three articles are reprinted in *Ethics: Problems and Principles*, ed. John Martin Fischer and Mark Ravizza (Fort Worth, Tex.: Harcourt Brace Jovanovich, 1992).

5. Greene et al., "An fMRI Investigation," supplementary material (available at ⟨http://www.sciencemag.org/cgi/content/full/sci;293/5537/2105/DC1/⟩).

> Is it appropriate for you to push the stranger onto the tracks in order to save the five workmen?"[6]

Most people judge that it *is not* appropriate to push the large stranger. What explains this difference in our moral judgments? On the one hand, it might seem puzzling that a majority of people judge differently about these two cases, since in each what is at stake is the life of five people versus the life of one. But on the other hand, there are myriad differences between these two scenarios that could (it might be thought) explain why most of us make a moral distinction between them.

The task of trying to fix on a morally relevant feature of these two scenarios that explains why we are justified in giving differing verdicts about them has come to be known as *the trolley problem*.[7] Usually the presupposition of this literature is that our moral intuitions about these cases (and others of their ilk) are largely accurate, the goal being to find a plausible moral principle that both agrees with and explains our intuitive verdicts about the cases in question. But what makes the trolley

6. Ibid.

7. Actually, this isn't entirely correct. The trolley problem is usually taken to be, not the problem of explaining our differing verdicts about the footbridge and trolley driver dilemmas, but rather the problem of explaining our differing verdicts about the footbridge dilemma and a variant of the trolley driver dilemma in which you are a bystander who sees the runaway trolley and can hit a switch that will divert the trolley onto the sidetrack containing the one person. Indeed, Thomson (who introduced the term "trolley problem" into the philosophical lexicon) thinks there is no problem explaining the difference in our intuitive reactions to the trolley driver and footbridge dilemmas; for Thomson (and for others following her), the real problem is explaining what grounds our different judgments about the bystander and footbridge dilemmas. And though Singer's summary of Greene et al.'s research suggests that it was the bystander dilemma that was tested (see Singer, "Ethics and Intuitions," p. 339), and though Greene himself, when describing his research, almost always summarizes the trolley driver dilemma in a way that is ambiguous between the driver and bystander variants (see Greene et al., "An fMRI Investigation," p. 2105; Greene et al., "Neural Bases," p. 389; and Greene, "Secret Joke," pp. 41–42), it is worth pointing out that in all of the published studies I discuss in this article, it was only the driver, not the bystander, version of the standard trolley dilemma that was studied. Perhaps it is being assumed that our judgments about the driver and bystander cases (and their neural correlates) will be the same; however, many philosophers mark a distinction between these two cases, and in her most recent discussion of the trolley problem ("Turning the Trolley," *Philosophy & Public Affairs* 36 [2008]: 359–74), Thomson argues that although it is permissible to divert the trolley if one is the driver, it is impermissible to divert the trolley if one is a bystander. (Thus, on Thomson's current way of seeing things, there actually is no trolley problem, since the very formulation of that problem contains a false presupposition that there is a morally relevant difference between the bystander and footbridge cases, but no morally relevant difference between the bystander and driver cases.)

problem so hard—indeed, what has led some to despair of our ever finding a solution to it—is that for nearly every principle that has been proposed to explain our intuitions about trolley cases, some ingenious person has devised a variant of the classic trolley scenario for which that principle yields counterintuitive results. Thus as with the Gettier literature in epistemology and the causation and personal identity literatures in metaphysics, increasingly baroque proposals have given way to increasingly complex counterexamples, and though some have continued to struggle with the trolley problem, many others have simply given up and moved on to other topics.[8]

Rather than deal with the *normative task* of proposing principles that *justify* our responses to trolleylike cases, Greene and his colleagues decided to pursue the *descriptive task* of investigating the physiological processes that *underlie* our responses to these sorts of cases. Their central empirical hypothesis requires making two distinctions: first a distinction between two different classes of moral judgments, and second a distinction between two different classes of psychological processes. The distinction between classes of moral judgments is as follows. Notice that a judgment that it is morally permissible to hit the switch in the trolley driver dilemma is precisely the sort of verdict predicted by a utilitarian or, more generally, consequentialist moral framework: since one's hitting the switch presumably results in a state of affairs with greater aggregate well-being than the state of affairs that would result were one not to hit the switch, according to most forms of consequentialism one is morally required—and hence permitted—to hit the switch. Following Greene, let us call particular-case moral judgments that, like the judgment that it is morally permissible to hit the switch in the trolley driver dilemma, are "easily justified in terms of the most basic consequentialist principles" *characteristically consequentialist judgments*.[9] One judgment that is not characteristically consequentialist is the judgment that it is morally impermissible to push the overweight individual

8. For a survey of the early classics of the trolley problem literature, see the papers collected in Fischer and Ravizza (eds.), *Ethics: Problems and Principles*. More recent classics not included in that anthology are Judith Jarvis Thomson, *The Realm of Rights* (Cambridge, Mass.: Harvard University Press, 1990), chap. 7; F. M. Kamm, *Morality, Mortality, Vol. II: Rights, Duties, and Status* (Oxford: Oxford University Press, 1996), chaps. 6–7; and F. M. Kamm, *Intricate Ethics* (Oxford: Oxford University Press, 2007), chaps. 1–6.

9. Greene, "Secret Joke," p. 39.

to his demise in the footbridge dilemma: precisely what sets deontological moral theories apart from consequentialist ones is that deontological theories tend to yield the result that it is impermissible to kill another person in this way, even for the sake of "the greater good." So, following Greene again, let us call particular-case moral judgments that, like the judgment that it is morally impermissible to push the obese man in the footbridge dilemma, are "in favor of characteristically deontological conclusions" *characteristically deontological judgments*.[10] This gives us a twofold distinction between types of particular-case moral judgments, which is usually taken to correspond to a twofold distinction between types of particular-case moral intuitions.[11]

10. Ibid. Note that these two definitions are not parallel. As Greene uses the expressions, "characteristically consequentialist judgment" means "judgment supported by the sort of moral principle that typically distinguishes consequentialist theories from deontological ones," whereas "characteristically deontological judgment" means "judgment in favor of the sort of verdict that typically distinguishes deontological theories from consequentialist ones." (The contrast is at the level of supporting principles in the one case, at the level of particular judgments in the other.) Thus even though nearly all deontologists judge that it is permissible to divert the trolley in the trolley driver dilemma, such a judgment counts as characteristically consequentialist but does not count as characteristically deontological.

11. In philosophical discussions of the metaphysics and epistemology of intuitions, there is an ongoing debate over whether intuitions just are judgments arrived at in a particular way (for example, not as a result of explicit reasoning, testimony, and so on), or whether intuitions are a separate class of mental entities that stand to intuitive judgments as perceptual experiences stand to perceptual judgments. For the former view, see Alvin Goldman and Joel Pust, "Philosophical Theory and Intuitional Evidence," in *Rethinking Intuition*, ed. Michael R. DePaul and William Ramsey (Lanham, Md.: Rowman & Littlefield, 1998), pp. 179–97; for the latter, see George Bealer, "Intuition and the Autonomy of Philosophy," in the same volume, pp. 201–39. We need not take a stand on this debate here, since even if moral intuitions are separate entities over and above the moral judgments formed on their basis, there will usually be an intuitive moral judgment corresponding to each moral intuition. Thus either we can say (if we identify intuitions with intuitive judgments) that the experiments in question directly study our moral intuitions, or we can say (if we distinguish intuitions from intuitive judgments) that the experiments indirectly study our moral intuitions by collecting data on our moral judgments, which are taken to be tightly correlated with our moral intuitions. In what follows I will generally be fairly lax in sliding back and forth between talk of judgments and talk of intuitions.

(That said, I am not using "intuition" as that term is used in much of the psychology literature, where it refers to any sort of automatic, spontaneous "gut feeling" that one might have. See most of the studies cited in David G. Myers, *Intuition: Its Powers and Perils* [New Haven, Conn.: Yale University Press, 2002] for this sort of usage, which has little to do with what philosophers mean when they talk about intuitions.)

There are a number of reasons to be worried about this distinction between characteristically consequentialist and characteristically deontological moral judgments, but I want to put them aside for the time being so that we can get Greene et al.'s empirical hypothesis on the table. The second distinction upon which that hypothesis depends is a distinction between two kinds of psychological processes: *emotional processes* and *"cognitive" processes*. (Following Greene et al.'s useful convention,[12] I use scare quotes when I mean "cognitive" to refer to specifically nonemotional information processing, in contrast to the also widespread use of "cognitive" to refer to information processing in general, as it does in the phrase "cognitive science.") Exactly how to flesh out the emotional versus "cognitive" process distinction is a contentious matter, so it is worth noting that Greene et al. use "emotional processing" to refer to information processing that involves behaviorally valenced representations that trigger automatic effects and hence "have direct motivational force,"[13] and they use " 'cognitive' processing" to refer to information processing that involves "inherently neutral representations . . . that do not automatically trigger particular behavioral responses or dispositions."[14] Emotional processes tend to be *fast and frugal* (providing quick responses on the basis of a limited amount of information), and they tend to be *domain-specific* (responding to particular subject matters, rather than any subject matter in general). By contrast, "cognitive" processes tend to be *slow but flexible* and *domain-neutral*, and it has been found that, at least in nonmoral cases, "cognitive" processes are recruited for such things as abstract reasoning, problem solving, working memory, self-control, and higher executive functions more generally.[15] Regions of the brain that have been associated with emotional processing include the following: the medial prefrontal cortex; the posterior cingulate/precuneus; the posterior superior temporal sulcus/inferior parietal lobe; the orbitofrontal/ventromedial prefrontal cortex; and the amygdala. Regions of the brain that have been associated with "cognitive" processing include the dorsolateral prefrontal cortex and the parietal lobe.

12. See Greene et al., "Neural Bases," p. 389; and Greene, "Secret Joke," p. 40.

13. Greene et al., "Neural Bases," p. 397.

14. Greene, "Secret Joke," p. 40.

15. The qualification "at least in nonmoral cases" is crucial here: to say at the outset that "cognitive" processes handle abstract reasoning and problem solving in all domains (including the moral) is question begging.

Thus we have a twofold distinction between types of moral judgments (characteristically consequentialist ones versus characteristically deontological ones) and a twofold distinction between types of psychological processes (emotional ones versus "cognitive" ones). The natural question to ask is: which sort (or sorts) of processes underlie each sort of judgment? The proposal put forward by Greene and his colleagues is as follows:

> *Greene et al.'s Dual-Process Hypothesis*: Characteristically deontological judgments are driven by emotional processes, whereas characteristically consequentialist judgments are driven by "cognitive" processes, and these processes compete for one's overall moral verdict about a given case.[16]

In one way, this is an extremely old picture of how the moral mind works. The idea that reason and passion struggle for one's overall moral stance was already commonplace by the time Hume wrote the *Treatise*. "Nothing is more usual in philosophy, and even in common life, than to talk of the combat of passion and reason," Hume tells us (*Treatise* 2.3.3), before going on to argue against this Combat Model of the soul (as Christine Korsgaard usefully dubs it).[17] So Greene et al.'s embracing of the Combat Model is not new.[18] What is new, though, is the surprising twist that they give to that model. Whereas deontology is usually associated with reason and consequentialism with the sentiments, Greene and his colleagues claim that in fact the opposite is true. On their picture, when we contemplate hitting the switch in the trolley driver dilemma, our more "cognitive" brain processes perform a cool

16. See Greene et al., "Neural Bases," p. 398; Greene et al., "Cognitive Load," p. 1145; and Greene, "Secret Joke," pp. 40–41.

17. See Christine Korsgaard, "Self-Constitution in the Ethics of Plato and Kant," *Journal of Ethics* 3 (1999): 1–29. Korsgaard contrasts the Combat Model of the soul with an alternate Constitution Model that she finds in Plato and Kant.

18. Though perhaps it is new to empirical psychology: as Greene et al. tell the story ("Neural Bases," pp. 397–98; "Cognitive Load," p. 1145), empirical psychology went through a long period where, under the influence of Lawrence Kohlberg, it was widely believed that processes of reasoning underwrite the moral judgments of mature adults, followed by a more recent period in which Jonathan Haidt and others have proposed that moral judgments are primarily driven by automatic, emotional processes. See Lawrence Kohlberg, "Stage and Sequence: The Cognitive-Developmental Approach to Socialization," in *Handbook of Socialization Theory and Research*, ed. David A. Goslin (Chicago: Rand McNally, 1969), pp. 347–480; and Jonathan Haidt, "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* 108 (2001): 814–34.

and detached cost-benefit analysis, yielding the verdict that it is permissible to hit the switch, but when we contemplate pushing the heavy fellow in the footbridge dilemma, our more emotional brain processes kick in and scream at us, "Don't do that!" thus overriding the cost-benefit analysis that would have deemed it permissible to push the man.

Greene et al.'s dual-process hypothesis is an empirical hypothesis, and as such it yields a number of empirical predictions. Here I mention just two. First, the dual-process hypothesis predicts that contemplation of cases like the footbridge dilemma should produce increased neural activity in regions of the brain associated with emotional processes, whereas contemplation of cases like the trolley driver dilemma should produce increased neural activity in regions of the brain associated with "cognitive" processes. Second, the dual-process hypothesis predicts that people who reach a nonstandard verdict about cases like the footbridge dilemma should take longer to reach their verdict than those who reach a standard verdict (since they are overriding an emotional response in order to come to their final verdict), whereas people who reach a nonstandard verdict about cases like the trolley driver dilemma should take approximately as long to reach a verdict as those who reach a standard verdict (since no emotional response is being overridden).

Because the results from fMRI machines are statistically noisy, Greene and his colleagues could not test these predictions simply by comparing people's reactions to the footbridge and trolley driver dilemmas. Instead, they needed to compare the neural-activity and response-time results when people responded to a large number of cases that are "like the footbridge dilemma" (and hence give rise to deontological judgments) and a large number of cases that are "like the trolley driver dilemma" (and hence give rise to consequentialist judgments). But this leads to a problem: how do we sort the former class of dilemmas from the latter? Greene et al. settled on the following "purely descriptive"[19] way of doing so: cases that are like the footbridge dilemma were deemed to involve *harm that is brought about in an "up close and personal" way*, whereas cases that are like the trolley driver dilemma were deemed to involve *harm that is brought about in an impersonal way*. What, though, does it

19. Greene, "Secret Joke," p. 43.

mean for harm to be brought about in an "up close and personal" way? Here is Greene et al.'s "first cut" proposal: a moral dilemma counts as personal if and only if "the action in question (a) could reasonably be expected to lead to serious bodily harm (b) to a particular person or a member or members of a particular group of people (c) where this harm is not the result of deflecting an existing threat onto a different party."[20] The basic idea can be helpfully summarized in the slogan "ME HURT YOU":

> The "HURT" criterion [= (a)] picks out the most primitive kinds of harmful violations (e.g., assault rather than insider trading) while the "YOU" criterion [= (b)] ensures that the victim be vividly represented as an individual. Finally, the "ME" criterion [= (c)] captures a notion of "agency," requiring that the action spring in a direct way from the agent's will, that it be "authored" rather than merely "edited" by the agent.[21]

Moral dilemmas that were deemed by a set of independent coders to meet conditions (a), (b), and (c) were classified as *personal moral dilemmas*, and all other moral dilemmas were classified as *impersonal moral dilemmas*. Then, working under the assumption that the personal versus impersonal moral dilemma distinction tracks the dilemma-giving-rise-to-a-deontological-judgment versus dilemma-giving-rise-to-a-consequentialist-judgment distinction, Greene et al. used this way of divvying up moral dilemmas into two piles to test their neural-activity and response-time predictions.

And what Greene and his colleagues found was truly remarkable. They had a number of subjects respond to approximately twenty personal moral dilemmas, approximately twenty impersonal moral dilemmas, and approximately twenty nonmoral dilemmas while inside fMRI machines.[22] These machines track the magnetic signature of oxygenated

20. Greene et al., "An fMRI Investigation," p. 2107, n. 9.

21. Greene et al., "Neural Bases," p. 389.

22. In experiment 1 in Greene et al., "An fMRI Investigation," nine subjects responded to fourteen personal moral dilemmas, nineteen impersonal moral dilemmas, and twenty nonmoral dilemmas, presented in random order, while inside fMRI machines. (For the exact wording of the dilemmas that were used, see the supplementary material at ⟨http://www.sciencemag.org/cgi/content/full/sci;293/5537/2105/DC1/⟩.) Each dilemma was presented as three screens of text, after which the subject was required to give a verdict about

blood, which is widely taken to be a fairly accurate way of measuring the level of neural activity in different portions of the brain. Thus Greene et al. were able to test their predictions about neural activity. Moreover, they found that their predictions were largely borne out. When responding to personal moral dilemmas, subjects exhibited increased activity in the following brain areas associated with emotional processes: the medial prefrontal cortex, the posterior cingulate/precuneus, the posterior superior temporal sulcus/inferior parietal lobe, and the amygdala.[23] When responding to impersonal moral dilemmas, subjects exhibited increased activity in two classically "cognitive" brain regions, the dorsolateral prefrontal cortex and the parietal lobe.[24] Furthermore, during several trials Greene et al. measured their subjects' response time to each question, and they reported that their response-time prediction was also confirmed: although subjects who gave emotionally incongruent answers to personal moral dilemmas took almost two seconds longer, on average, to respond than those who gave emotionally congruent responses, there was no comparable effect for impersonal moral

---

the appropriateness of a proposed action by pressing one of two buttons ("appropriate" or "inappropriate"). Subjects were allowed to advance to each subsequent screen of text at their own rate, though they were given a maximum of 46 seconds to read through all three screens and respond. In experiment 2 in that same article, nine different subjects responded to twenty-two personal moral dilemmas, nineteen impersonal moral dilemmas, and twenty nonmoral dilemmas, using the same protocol. (The set of personal moral dilemmas was altered in experiment 2 to remove a possible confound in the experimental design: see p. 2108, n. 24.)

In Greene et al., "Neural Bases," thirty-two new subjects responded to twenty-two personal moral dilemmas, eighteen impersonal moral dilemmas, and twenty nonmoral dilemmas, using the same protocol. (These were the same dilemmas used in experiment 2 of "An fMRI Investigation," except one of the impersonal moral dilemmas was dropped.) The data from these new subjects together with the data from the nine subjects from experiment 2 in "An fMRI Investigation" were then analyzed together as a whole.

23. See Greene et al., "An fMRI Investigation," p. 2106, fig. 1, and p. 2107; and Greene et al., "Neural Bases," p. 391 and p. 392, table 1. The superior temporal sulcus was originally labeled "angular gyrus" in the first study. Activity in the amygdala was not detected in the first study but was detected in the larger second study. Due to a "magnetic susceptibility artifact," neither study was able to image the orbitofrontal cortex, another brain area that has been associated with emotional processing (see Greene et al., "Neural Bases," p. 2108, n. 21).

24. See Greene et al., "An fMRI Investigation," p. 2106, fig. 1, and p. 2107; and Greene et al., "Neural Bases," p. 391 and p. 392, table 1.

dilemmas.[25] All told, Greene et al.'s empirical results present an impressive case for their dual-process hypothesis.[26]

III. METHODOLOGICAL WORRIES

In general it is dangerous (and perhaps futile) for philosophers to resist empirically based challenges by calling into question the methodology of the relevant experiments, or the interpretation of their results. Not only are philosophers often not well trained at evaluating scientific studies, but also they need to be extremely careful that the (alleged) design flaws to which they point are not ones that could easily be overcome in future research, or ones that in the end are irrelevant to the main philosophical

25. See Greene et al., "An fMRI Investigation," p. 2107, fig. 3.

26. In the body of this article I have focused on the neuroimaging and response-time findings, since these results are particularly vivid and tend to capture the public's imagination. However, there have been a number of follow-up studies which have been taken to lend further support to Greene et al.'s dual-process hypothesis, including the following:

(1) In an additional experiment in "Neural Bases," Greene et al. found that when subjects contemplated "difficult" personal moral dilemmas (where degree of difficulty was measured by response time), the anterior cingulate cortex (a brain region associated with conflict monitoring) and the dorsolateral prefrontal cortex (a brain region associated with abstract reasoning) exhibited increased activity, in addition to regions associated with emotion. They also found that the level of activity in the dorsolateral prefrontal cortex was positively correlated with consequentialist responses to these "difficult" personal moral dilemmas.

(2) In a follow-up study, Greene and colleagues found that consequentialist responses to "difficult" personal moral dilemmas took longer when subjects were required to perform a cognitively intensive task at the same time as responding to the dilemmas, but deontological responses did not exhibit this effect. See Greene et al., "Cognitive Load."

(3) Michael Koenigs, Liane Young, and colleagues found that patients with damage to their ventromedial prefrontal cortex (a brain region associated with the emotions) gave a greater percentage of consequentialist responses to the personal moral dilemmas from Greene et al.'s "An fMRI Investigation" than control subjects did. See Michael Koenigs, Liane Young, Ralph Adolphs, Daniel Tranel, Fiery Cushman, Marc Hauser, and Antonio Damasio, "Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgments," *Nature* 446 (2007): 908–11.

(4) Piercarlo Valdesolo and David DeSteno found that respondents who had watched a funny clip from *Saturday Night Live* were more likely to give a consequentialist response to the footbridge dilemma than those who had watched a clip from a dull documentary beforehand, but there was no comparable effect for the trolley driver dilemma. See Valdesolo and DeSteno, "Manipulations of Emotional Context Shape Moral Judgment," *Psychological Science* 17 (2006): 476–77.

issues at stake.[27] Nevertheless, I think it is worth bringing up three empirical issues about Greene et al.'s research before turning to the more important question of what, in principle, the normative implications of this sort of research could be. These three issues don't entirely

27. One potential design worry about Greene et al.'s research that falls into the former category (i.e., worries that could easily be overcome in future research) is as follows. In order to have a comparison class for their neural-activity data, Greene and his colleagues had their subjects respond to a number of nonmoral dilemmas, such as:

> *Turnips Dilemma*: "You are a farm worker driving a turnip-harvesting machine. You are approaching two diverging paths.
>
> By choosing the path on the left you will harvest ten bushels of turnips. By choosing the path on the right you will harvest twenty bushels of turnips. If you do nothing your turnip-harvesting machine will turn to the left.
>
> Is it appropriate for you to turn your turnip-picking machine to the right in order to harvest twenty bushels of turnips instead of ten?" (Greene et al., "An fMRI Investigation," supplementary material)

The trigger question here was formulated in terms of appropriateness to make it as parallel as possible to the trigger questions for the moral dilemmas tested, but one might worry that it sounds very odd to ask whether it is "appropriate" to turn a machine one way rather than the other. (Should we interpret this as some sort of prudential appropriateness? Is there even such a notion?) Moreover, the answer to this so-called dilemma is completely obvious, and all told I estimate that of the twenty nonmoral dilemmas used in Greene et al.'s studies, twelve have completely obvious answers, six are somewhat less obvious, and only two are genuine dilemmas; thus one might worry that too many of these nonmoral "dilemmas" have readily evident answers for them to serve as an accurate comparison class. However, both of these worries could easily be avoided in future research: the set of nonmoral dilemmas could be altered to include a greater number of difficult dilemmas, and the trigger question for the dilemmas (both moral and nonmoral) could be phrased in a less awkward way. (Indeed, in their "Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgments," Koenigs et al. used Greene et al.'s dilemmas but rephrased the trigger questions so as to ask whether the subjects "would" perform the action in question, rather than asking whether they deem it "appropriate" to perform the action, perhaps for this very reason. However, this rewording introduces new problems, since one's judgments about *what one would do* and one's judgments about *what it is morally permissible to do* might pull apart.)

One potential design worry about Greene et al.'s research that falls into the latter category (i.e., worries that ultimately are not relevant to the philosophical issues at stake) is as follows. Because of the limitations of fMRI technology, Greene and his colleagues were only able to study neural activity when subjects contemplated hypothetical scenarios. Thus their dual-process hypothesis has, at least so far, not been tested with regard to moral judgments about actual scenarios (whether currently perceived or merely remembered). However, these limited results are enough for the philosophical questions at issue. Even if we could only conclude from Greene et al.'s research that deontological judgments about hypothetical cases are not trustworthy but could not make a parallel claim with regard to deontological judgments about actual cases, that would still be an extremely significant result, given the ubiquity of appeals to hypothetical cases in first-order moral theorizing.

undermine the empirical findings, but they do cast them in a different light. Moreover, at least one of these issues will be crucial to our discussion of the normative implications of Greene et al.'s research.

First Empirical Issue. Neural activity in at least one brain region associated with emotion was found to be correlated with consequentialist judgment. In particular, Greene and his colleagues found that activity in the posterior cingulate, a portion of the brain known to be recruited for emotional processes, predicts characteristically consequentialist responses to personal moral dilemmas (for example, a response that it is appropriate to push the portly gentleman in the footbridge dilemma).[28] Greene et al. concede that these results cast doubt on the simplest version of the dual-process hypothesis, according to which consequentialist judgments are wholly tied to "cognitive" processes and deontological judgments wholly tied to emotional processes. They write, "Like David Hume . . . we suspect that all action, whether driven by 'cognitive' judgment or not, must have some affective basis."[29] However, what is at stake here is whether *all moral judgment*, not *all action*, has an affective basis. Also, Hume is a dangerous ally to call on at this point: Hume's contention was not just that all moral judgments "have some emotional component," as Greene at one point contends Hume's view to be,[30] but moreover that all moral judgments are *entirely driven* by the passions. On Hume's picture, the struggle in our souls is not between reason and passion, with a few passions along for the ride on reason's side; rather, the fundamental struggle is between different passions, with reason the underling carrying out each passion's whims. "Reason is, and ought only to be slave of the passions, and can never pretend to any other office than to serve and obey them" (*Treatise* 2.3.3) is not a slight amendment of the dual-process hypothesis; it is a complete subverting of it.

In his own writings Greene tries to reinstate a contrast between the processes underlying deontological judgments and the processes underlying consequentialist judgments by proposing that the emotions that drive deontological judgments are "alarmlike," whereas those that are present during consequentialist judgments are "more like a currency."[31]

28. See Greene et al., "Neural Bases," p. 395, table 4, and p. 397.
29. Ibid., p. 397.
30. Greene, "Secret Joke," p. 41.
31. Ibid.; see also pp. 64–65.

However, this alleged contrast appears to have no empirical backing, short of an appeal by Greene to his own phenomenology when he considers cases like the trolley driver and footbridge dilemmas. So in the end I think the best option for Greene and his colleagues is not to slide into a full-blown Humean sentimentalism, nor to posit a speculative (and potentially question-begging) phenomenological distinction between "alarmlike" and "currencylike" emotions, but rather to concede that at least one region of the brain traditionally associated with the emotions appears to be recruited for characteristically consequentialist judgments and then to insist that exactly how crucial a role this region plays in such judgments is a topic for future research. Their finding with regard to the posterior cingulate muddies the waters a bit, but it need not mean that the dual-process hypothesis is completely off target.

Second Empirical Issue. When interpreted correctly, the response-time data collected by Greene et al. do not, in fact, confirm their prediction about comparative response times. Each of Greene et al.'s personal moral dilemmas was arranged so that a response in which the subject deems it "appropriate" to perform the proposed action (for example, appropriate to shove the hefty stranger in the footbridge dilemma) goes against the hypothesized emotional response, and a response in which the subject deems it "inappropriate" to perform the proposed action (for example, inappropriate to shove the hefty stranger in the footbridge dilemma) is consonant with the hypothesized emotional response. In their original paper in *Science*, Greene et al. reported that responses of "appropriate" to their various personal moral dilemmas took on average 6.8 seconds, whereas responses of "inappropriate" took on average 5.0 seconds,[32] and they took this to confirm their response-time prediction.[33] However, *this is a statistically invalid way of presenting these results*. A more accurate way of presenting the results would be to calculate the average difference in response time between the "appropriate" responses and the "inappropriate" responses *for each question*, and then to calculate the average of these average differences. Such a way of calculating things prevents differences from question to question in the average response time and differences from question to question in the ratio of "appropriate" to "inappropriate" responses (two factors that

32. Greene et al., "An fMRI Investigation," p. 2107, fig. 3.
33. Ibid., p. 2107.

should be irrelevant to the truth of the hypothesis being tested) from skewing the final results.[34]

As it turns out, this statistical point ends up mattering. The group of personal moral dilemmas that Greene and his colleagues tested were a motley bunch, including not only the footbridge dilemma, but also cases such as the following:

> *Architect Dilemma*: "You are a young architect visiting one of your construction sites with your boss. Your boss is a despicable individual who makes everyone around him miserable including you.
>
> It occurs to you that if you were to push him off the building you are inspecting he would fall to his death and everyone would think it was an accident.
>
> Is it appropriate for you to push your boss off of the building in order to get him out of your life?"[35]
>
> *Hired Rapist Dilemma*: "You have been dissatisfied with your marriage for several years. It is your distinct impression that your wife no longer appreciates you. You remember how she appreciated you years ago when you took care of her after she was mugged. You devise the following plan to regain your wife's affection.
>
> You will hire a man to break into your house while you are away. This man will tie up your wife and rape her. You, upon hearing the horrible news, will return swiftly to her side, to take care of her and comfort her, and she will once again appreciate you.

34. To see how these factors could skew the results on Greene et al.'s way of calculating things, consider this. Suppose we took the personal moral dilemma that produced the greatest percentage of "appropriate" answers, and then added a large amount of filler text to the dilemma so as to increase by some set amount the response time for any answer to it. Then—assuming that this change does not affect the ratio of "appropriate" to "inappropriate" responses for that dilemma—our change will have raised *the average of all answers of "appropriate" to personal moral dilemmas* (Greene et al.'s proposed metric) quite a bit more than it raises *the average of all answers of "inappropriate" to personal moral dilemmas*. However, *the average of the average differences in response time between "appropriate" and "inappropriate" response for each personal moral dilemma* (my proposed metric) would be unaffected by such a change.

35. Greene et al., "An fMRI Investigation," supplementary material.

> Is it appropriate for you to hire a man to rape your wife so that she will appreciate you as you comfort her?"[36]

So there is a worry that the difference between the average response times for all answers of "appropriate" and the average response times for all answers of "inappropriate" to personal moral dilemmas was largely due to the fact that almost all respondents very quickly answered "inappropriate" to cases like the architect and hired rapist cases, which are hardly deserving of the title "dilemma." And this was indeed what happened: in a later paper, Greene et al. admit that when their response-time data are analyzed with cases like the architect and hired rapist cases thrown out, they reveal "no reliable differences in RT [response time]" between those who gave a response of "appropriate" and those who gave a response of "inappropriate" to personal moral dilemmas.[37]

I bring up this issue because in presenting the empirical case for the dual-process hypothesis, Singer and Greene lean quite heavily on the response-time data.[38] Perhaps future research with larger sample sizes will confirm some version of the response-time prediction; however, it is important to keep in mind that at this point in time the response-time

36. Ibid.

37. Greene et al., "Cognitive Load," p. 1146, n. 5. Actually, when they concede this, Greene et al. are worried about a slightly different issue: here they are worried that cases such as the architect and hired rapist dilemmas should not be included in the analysis, since an answer of "appropriate" to such dilemmas does not correspond (or does not obviously correspond) to a consequentialist judgment about such a case. Though I share this worry (see my third empirical issue, below), my point here is somewhat different. Even if we toss out the response-time data from the architect and hired rapist dilemmas, it is still statistically invalid to compare the average response time of every answer of "appropriate" to the average response time of every answer of "inappropriate," rather than averaging the average differences in response time between answers of "appropriate" and "inappropriate" for each question.

It is not clear to me whether Greene et al. now realize this point. On the one hand, in a more recent study in which they compare subjects' response times when responding to moral dilemmas while performing a cognitively intensive task (see n. 26), Greene et al. continue to present their response-time data in the statistically invalid manner (see Greene et al., "Cognitive Load," p. 1149, fig. 1, and p. 1150, fig. 2). On the other hand, they write in that same study, "This general pattern also held when item, rather than participant, was modeled as a random effect, though the results in this analysis were not as strong" (ibid., p. 1149).

38. See Singer, "Ethics and Intuitions," pp. 341–42; and Greene, "Secret Joke," p. 44. See also Joshua D. Greene, "Reply to Mikhail and Timmons," in *Moral Psychology, Vol. 3*, pp. 105–17, at p. 109.

prediction has not been borne out, which in fact is an empirical strike *against* the dual-process hypothesis.[39]

Third Empirical Issue. Greene et al.'s tentative criteria for sorting personal from impersonal moral dilemmas are an inadequate way of tracking the dilemma-giving-rise-to-a-deontological-moral-judgment versus dilemma-giving-rise-to-a-consequentialist-moral-judgment distinction. To claim that characteristically deontological judgments only concern bodily harms is nothing short of preposterous; after all, the stock in trade of deontology is supposed to involve not just prohibitions on murder and mayhem, but also requirements against lying, promise-breaking, coercion, and the like.[40] But even within the realm of bodily harms, there is an abundance of clear counterexamples to Greene et al.'s proposal. To mention just one: Frances Kamm's famous Lazy Susan Case is a dilemma giving rise to a consequentialist moral judgment, but it is deemed a personal moral dilemma by Greene et al.'s tripartite "ME HURT YOU" criteria. In this case, a runaway trolley is heading toward five innocent people who are seated on a giant lazy Susan. The only way to save the five people is to push the lazy Susan so that it swings the five out of the way; however, doing so will cause the lazy Susan to ram into an innocent bystander, killing him.[41] Kamm's intuition about this case is characteristically consequentialist: she thinks it is permissible to push the lazy Susan, thereby killing the one to save the five. However, in doing so one would initiate a new threat (ME) that causes serious bodily harm (HURT) to a person (YOU), so this case counts as a personal moral dilemma according to Greene et al.'s criteria. Thus Greene et al.'s crucial assumption that we can establish a claim about the psychological processes underlying deontological and consequentialist judgments by

39. Recently, three psychologists and one philosopher reanalyzed Greene et al.'s data from "An fMRI Investigation" and definitively established that Greene et al.'s response-time prediction was, in fact, disconfirmed by that data. See Jonathan McGuire, Robyn Langdon, Max Coltheart, and Catriona Mackenzie, "A Reanalysis of the Personal/Impersonal Distinction in Moral Psychology Research," *Journal of Experimental Social Psychology* 45 (2009): 577–80.

40. Of course, exactly how we cash out these prohibitions/requirements will vary from one deontological theory to the next. (That said, it is important to keep in mind that a Ten-Commandments-style "Never, ever kill!", "Never, ever lie!", et cetera, version of deontology is not the only option; indeed, such a picture is rarely, if ever, defended outside of introductory ethics courses.)

41. Kamm, *Morality, Mortality, Vol. II: Rights, Duties, and Status*, p. 154.

testing the differing processes utilized to think about personal versus impersonal moral dilemmas is seriously called into question.

Actually, matters are even worse than that. What we really have are three distinctions: (i) the distinction between moral dilemmas that typically elicit a characteristically deontological reaction and those that typically elicit a characteristically consequentialist reaction; (ii) the distinction between moral dilemmas that intuitively involve harm brought about in an "up close and personal" way and those that do not; and (iii) the distinction between moral dilemmas that satisfy Greene et al.'s "ME HURT YOU" criteria and those that do not. The problem is that *none* of these distinctions matches up with the others. We have already seen how Kamm's Lazy Susan Case shows that distinction (i) is not distinction (iii). Kamm's case also shows why distinction (i) is not distinction (ii): killing someone by ramming a giant lazy Susan tray into him presumably counts as harming that person in an "up close and personal" manner, yet this case is one that gives rise to a characteristically consequentialist judgment. Moreover, there are a variety of cases that show that distinction (ii) is not distinction (iii). Most famously, a variant of the footbridge case in which there is a trapdoor under the fat man that you can trigger from afar intuitively counts as a case in which someone is not harmed in an "up close and personal" way, yet such a case is deemed to be a personal moral dilemma by the "ME HURT YOU" criteria, since triggering the trapdoor initiates a new threat (ME) that causes serious bodily harm (HURT) to a specific individual (YOU).[42] So all three of these distinctions pull apart.[43]

42. If you doubt that the ME criterion holds in this case, keep in mind that it is usually a standard feature of the footbridge case and its variants that *the fall* is what kills the fat man, whose body then serves as a weight to slow down the runaway trolley.

43. In the body of this article I have appealed to Kamm's Lazy Susan Case in order to argue that the personal dilemma versus impersonal dilemma distinction (whether construed intuitively or in terms of the "ME HURT YOU" criteria) is not the dilemma-typically-giving-rise-to-a-deontological-judgment versus dilemma-typically-giving-rise-to-a-consequentialist-judgment distinction. I chose to use Kamm's case since it is familiar from the trolley problem literature. However, a slightly cleaner version of the same case that would equally well serve my purposes is as follows: instead of being on a lazy Susan, the five innocent people are on a dolly (i.e., a wheeled platform) that you can push out of the way of the oncoming trolley, but doing so will cause the dolly to roll down a hill and smash an innocent bystander to death. (Note that in order for either of these cases to be a counterexample to Greene et al.'s proposal, it is not necessary that *everyone* makes a characteristically consequentialist judgment about such a case; all we need is for a characteristically consequentialist judgment to be the standard reply.)

Greene et al. were never under any delusions that their initial proposal for sorting cases that are like the footbridge dilemma from cases that are like the standard trolley dilemma was fully adequate; after all, they explicitly called it a "first cut" proposal.[44] And Greene himself now admits that their proposal "does not work" and "is clearly wrong."[45] Greene and his colleagues consider it a task for future research to determine the proper way of characterizing the distinction that they tried to capture with the "ME HURT YOU" criteria.[46]

So where does this leave us? Even if some emotional processes are tied to consequentialist judgment, and even if Greene et al.'s response-time prediction does not hold up, and even if their way of mapping the deontological versus consequentialist judgment distinction onto the personal versus impersonal dilemma distinction is in need of revision, nonetheless Greene et al.'s neural-activity results strongly suggest that something like the dual-process hypothesis may well be true, though perhaps in a modified form.[47] The question to which I would like to now turn is: what follows from these findings?

Kamm notes that lazy Susan cases raise difficulties for Greene et al.'s personal versus impersonal distinction in her *Intricate Ethics*, pp. 142–43 and p. 180, n. 34; Greene concedes the point in his "Reply to Mikhail and Timmons," p. 108. See also p. 43, n. 37, and p. 418 of *Intricate Ethics*, where Kamm discusses a second sort of case that poses problems for that distinction.

44. Greene et al., "An fMRI Investigation," p. 2107.

45. Greene, "Reply to Mikhail and Timmons," pp. 107, 114.

46. Further evidence of the inadequacy of Greene et al.'s "ME HURT YOU" criteria is provided by Guy Kahane and Nicholas Shackel, "Do Abnormal Responses Show Utilitarian Bias?" *Nature* 452:7185 (2008): E5–E6. Kahane and Shackel had five professional moral philosophers categorize the moral dilemmas from Greene et al., "An fMRI Investigation," and they found that only five out of the nineteen personal moral dilemmas used in both experiments (26 percent) and only ten out of the twenty-two impersonal moral dilemmas used in experiment 2 (45 percent) were deemed by a majority of these philosophers to involve a choice in which a deontological option is contrasted with a consequentialist option. The data from Kahane and Shackel's study can be found at ⟨http://ethics-etc.com/wp-content/uploads/2008/03/dilemma-classification.pdf⟩.

47. What about the additional studies that have been taken to lend further support to the dual-process hypothesis (see n. 26)? Here, too, I think the empirical upshot is far from certain. Some worries about those studies:

(1) In "Neural Bases," Greene et al. took activity in the anterior cingulate cortex during contemplation of "difficult" personal moral dilemmas to provide evidence that there was a conflict between two subsystems in the brain. However, as they themselves note (p. 395), the exact function of the anterior cingulate cortex is not

IV. NORMATIVE IMPLICATIONS: THREE BAD ARGUMENTS

Greene and Singer think that quite a lot follows from Greene et al.'s experimental findings.[48] In particular, they think that these findings give

currently known, and the hypothesis that it is devoted to conflict monitoring is just one among several.

(2) While it is true that in "Cognitive Load" Greene et al. found that consequentialist responses to "difficult" personal moral dilemmas took longer when the subjects were performing a cognitively intensive task at the same time (as the dual-process hypothesis predicts), they did *not* find that subjects gave a lower percentage of consequentialist responses when performing the cognitively intensive task (as the dual-process hypothesis would also predict). Greene et al. try to explain away this troubling piece of counterevidence by speculating that the subjects were "trying to push through" the interference caused by the cognitively intensive task (Greene et al., "Cognitive Load," p. 1151), but as Adina Roskies and Walter Sinnott-Armstrong note, "this story makes sense only if subjects knew in advance that they wanted to reach a utilitarian judgment." See Roskies and Sinnott-Armstrong, "Between a Rock and a Hard Place: Thinking about Morality," *Scientific American Mind* 19 (2008), ⟨http://www.sciam.com/article.cfm?id=thinking-about-morality⟩.

(3) Jorge Moll and Ricardo de Oliveira-Souza raise some doubts as to whether Koenigs et al.'s data in their "Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgments" on patients with damage to the ventromedial prefrontal cortex really support Greene et al.'s dual-process hypothesis. Among other worries, Moll and de Oliveira-Souza point out that these patients also had damage to the anterior dorsolateral prefrontal cortex, a more "cognitive" portion of the brain that Greene et al. found to be correlated with consequentialist judgment in "An fMRI Investigation" and "Neural Bases," so these patients are not clean cases in which only emotional processing is impaired. See Moll and de Oliveira-Souza, "Moral Judgments, Emotions, and the Utilitarian Brain," *Trends in Cognitive Science* 11 (2007): 319–21, and "Response to Greene: Moral Sentiments and Reason: Friends or Foes?" *Trends in Cognitive Science* 11 (2007): 323–24. For Greene's reply, see his "Why Are VMPFC Patients More Utilitarian? A Dual-Process Theory of Moral Judgment Explains," *Trends in Cognitive Science* 11 (2007): 322–23. See also Kahane and Shackel, "Do Abnormal Responses Show Utilitarian Bias?"

More importantly, though, it is dialectically problematic *first* to appeal to patients with damage to emotional brain regions when making an empirical case for the dual-process hypothesis and *then* to go on to argue that the verdicts of these brain regions should be neglected (in effect urging us to be more like these patients), since many patients with this sort of brain damage make moral decisions in their personal lives that count as disastrous when evaluated by just about any plausible normative standard (see Antonio Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain* [New York: Putnam, 1994]). So even if studies of such brain-damaged patients end up supporting the dual-process hypothesis, they threaten to do so only at the cost of making Greene and Singer's normative claims less tenable.

48. See Greene, "From Neural 'Is' to Moral 'Ought' "; Greene, "Secret Joke"; and Singer, "Ethics and Intuitions."

us good reason to conclude that characteristically deontological moral intuitions should not be trusted. Moreover, they think that these findings impugn the epistemic credentials of characteristically deontological moral intuitions *without* impugning the epistemic credentials of characteristically consequentialist moral intuitions.

Both Singer and Greene draw further conclusions from their claim about the comparative epistemic standing of deontological and consequentialist intuitions. Revisiting an old debate with Rawls,[49] Singer argues that the untrustworthiness of deontological moral intuitions also shows that the method of reflective equilibrium is fundamentally misguided.[50] And Greene argues that, in addition, the untrustworthiness of deontological moral intuitions shows that all deontologists—even those who, like Kant,[51] don't explicitly rely on particular-case moral intuitions in their moral theorizing—are rationalizers who construct flimsy post hoc justifications for the very verdicts that they would be led by their emotions to make anyway.[52] I won't be discussing either of these additional arguments here, since they both obviously depend on the antecedent claim that we have good reason to discount deontological but not consequentialist particular-case moral intuitions.[53]

Before turning to Greene's and Singer's central argument against the probative force of deontological intuitions, though, I want to briefly discuss three bad arguments for that conclusion. On a charitable interpretation of Greene and Singer, these are arguments that they don't actually make but which it is extremely tempting to see them as making;

49. See Peter Singer, "Sidgwick and Reflective Equilibrium," *The Monist* 58 (1974): 490–517.

50. Singer, "Ethics and Intuitions," pp. 343–49.

51. On one interpretation of Kant, although he appeals to particular-case intuitions about the conditions under which something is good in the opening paragraphs of *Groundwork* I, he takes himself to have discharged any appeal to particular-case moral intuitions once he has completed his argument for the form and existence of the Categorical Imperative by the end of *Groundwork* III.

52. Greene, "Secret Joke," pp. 66–72.

53. For a reply to Singer's "Ethics and Intuitions" that focuses on the question of whether Greene et al.'s research poses problems for the method of reflective equilibrium, see Folke Tersman, "The Reliability of Moral Intuitions: A Challenge from Neuroscience," *Australasian Journal of Philosophy* 86 (2008): 389–405. However, to my mind Tersman is too quick to concede to Greene and Singer that Greene et al.'s research might demonstrate that deontological moral intuitions are unreliable; his main point is that even if this is so, the method of *wide* reflective equilibrium can take this fact into account.

on an uncharitable interpretation of Greene and Singer, these are bad arguments that they sloppily mix in with their main argument. My guess is that the truth lies somewhere in between: although Greene's and Singer's primary and most promising line of argumentation does not rely on these three arguments, I think they occasionally give their main argument more rhetorical force by invoking versions of these arguments. So it is worth showing just how unconvincing these three arguments are before we consider Singer's and Greene's main reason for thinking that Greene et al.'s neuroscientific research gives us good reason to privilege our characteristically consequentialist intuitions over our characteristically deontological ones.

The crudest possible argument for the conclusion reached by both Greene and Singer would proceed as follows:

The "Emotions Bad, Reasoning Good" Argument:

P. Deontological intuitions are driven by emotions, whereas consequentialist intuitions involve abstract reasoning.
C. So, deontological intuitions, unlike consequentialist intuitions, do not have any genuine normative force.

This is a bad argument. We need a substantive reason for thinking that intuitions based in emotion are less likely to be reliable than those based in "reasoning" for this argument to be at all convincing. After all, there is a venerable tradition that sees emotions as an important way of discerning normative truths.[54] One might disagree with this tradition, but showing that it rests on a mistake requires more than mere name-calling. Furthermore, even if the above argument were anything less than a howler, Greene et al.'s findings with regard to the posterior cingulate would cause additional problems for the argument. If consequentialist intuitions also recruit emotional processes, the pernicious influence of

54. For contemporary expressions of this sentiment, see Robert C. Solomon, *The Passions* (Garden City, N. Y.: Anchor Press/Doubleday, 1976); Ronald de Sousa, *The Rationality of Emotion* (Cambridge, Mass.: MIT Press, 1987); Patricia Greenspan, *Emotions and Reasons: An Inquiry into Emotional Justification* (New York: Routledge, 1988); Michael Stocker, with Elizabeth Hegeman, *Valuing Emotions* (Cambridge: Cambridge University Press, 1996); Bennett Helm, *Emotional Reason: Deliberation, Motivation, and the Nature of Value* (Cambridge: Cambridge University Press, 2001); and Martha Nussbaum, *Upheavals of Thought: The Intelligence of Emotions* (Cambridge: Cambridge University Press, 2001).

the emotions can hardly be used to drive an epistemic wedge between deontological and consequentialist intuitions.[55]

One natural way of improving on the "emotions bad, reasoning good" argument would involve arguing as follows:

The Argument from Heuristics:

P1. Deontological intuitions are driven by emotions, whereas consequentialist intuitions involve abstract reasoning.
P2. In other domains, emotional processes tend to involve fast and frugal (and hence unreliable) heuristics.
C1. So, in the moral domain, the emotional processes that drive deontological intuitions involve fast and frugal (and hence unreliable) heuristics.
C2. So, deontological intuitions, unlike consequentialist intuitions, are unreliable.

This is also not the best argument. Usually when we deem something to be a heuristic, we have a good handle on what the right and wrong answers in the relevant domain are; this is certainly the case in most of Daniel Kahneman and Amos Tversky's well-known examples of heuristics for logical and probabilistic reasoning. However, in the moral case it is very much up for debate what the right and wrong answers are. So it is question begging to assume that the emotional processes underwriting deontological intuitions consist in heuristics. Or more precisely: it is question begging to assume that just because emotional processes in other domains consist in heuristics, therefore emotional processes in the moral domain consist in heuristics. How can we proclaim these emotional processes to be quick but sloppy shortcuts for getting at the moral truth unless we already have a handle on what the moral truth is?[56]

I have just identified the inference from P1 and P2 to C1 as the major problem with the argument from heuristics. However, it is worth briefly mentioning two additional problems with the argument. First, it is a matter of some dispute whether premise P2 is even true. A number of

55. Replacing the "emotions bad, reasoning good" argument with an "alarm-like emotions bad, currency-like emotions plus reasoning good" argument is clearly no dialectical improvement either, unless something substantive is said about *why* currency-like emotions are less problematic than alarm-like ones.

56. Cass Sunstein makes a similar point in "Moral Heuristics," *Behavioral and Brain Sciences* 28 (2005): 531–73, at pp. 533–34.

authors have argued that, in the nonmoral domain, the fast and frugal heuristics underwriting emotional processes are often *more reliable* than their slow but flexible counterparts involving deliberate reasoning.[57] Second, the inference from C1 to C2 is also questionable. After all, it is doubtful that our mental machinery computes all of the actual and expected consequences of an action whenever we make a characteristically consequentialist judgment about it. So even if the neural processes underlying deontological intuitions rely upon heuristics, it is likely that the neural processes underlying consequentialist intuitions *also* make use of heuristics.[58] All told, the argument from heuristics faces serious and, to my mind, fatal problems.[59]

The first two arguments I have just considered involve fixing on the emotional nature of the processes that, according to the dual-process hypothesis, underlie deontological judgments; the third argument I want to consider takes a different tack. Greene and Singer motivate their main argument by telling a "just so" story about the evolution of our faculty for making deontological judgments about personal moral dilemmas. Here is Greene's version of that story:

> The rationale for distinguishing between *personal* and *impersonal* forms of harm is largely evolutionary. "Up close and personal" violence has been around for a very long time, reaching back into our primate lineage. . . . Given that personal violence is evolutionarily ancient, predating our recently evolved human capacities for complex

57. Gerd Gigerenzer has been arguing this point for several decades now. See, among other places, his "Moral Intuition = Fast and Frugal Heuristics?" in *Moral Psychology, Vol. 2: The Cognitive Science of Morality: Intuition and Diversity*, ed. Walter Sinnott-Armstrong (Cambridge, Mass.: MIT Press, 2008), pp. 1–26. John Allman and James Woodward also provide a number of nice examples from the psychology literature in which the utilization of automatic, emotional processes seems to result in better nonmoral decisions than the utilization of more cognitive, deliberative processes. See Woodward and Allman, "Moral Intuition: Its Neural Substrates and Normative Significance," *Journal of Physiology-Paris* 101 (2007): 179–202, at pp. 189–91, 195–96; and Allman and Woodward, "What Are Moral Intuitions and Why Should We Care about Them? A Neurobiological Perspective," *Philosophical Issues* 18 (2008): 164–85, at pp. 170–71, 174.

58. I am grateful to Louis Menand for this point.

59. Another complication: in formulating the argument from heuristics, I have assumed that heuristics are, by definition, unreliable. However, as Frances Kamm has reminded me, this assumption is, strictly speaking, false. For example, the rule "Add up all the digits and see if the result is divisible by three" is a useful heuristic for determining whether a natural number is divisible by three, but it is also perfectly reliable.

> abstract reasoning, it should come as no surprise if we have innate responses to personal violence that are powerful but rather primitive. That is, we might expect humans to have negative emotional responses to certain basic forms of interpersonal violence. . . . In contrast, when a harm is *impersonal*, it should fail to trigger this alarmlike emotional response, allowing people to respond in a more "cognitive" way, perhaps employing a cost-benefit analysis.[60]

Similarly, Singer writes,

> For most of our evolutionary history, human beings have lived in small groups. . . . In these groups, violence could only be inflicted in an up-close and personal way—by hitting, pushing, strangling, or using a stick or stone as a club. To deal with such situations, we have developed immediate, emotionally based responses to questions involving close, personal interactions with others.[61]

In light of their appeal to such an evolutionary story, it is very tempting to read both Greene and Singer as making something like the following argument:

> *The Argument from Evolutionary History*:
>
> P. Our emotion-driven deontological intuitions are evolutionary by-products that were adapted to handle an environment we no longer find ourselves in.
> C. So, deontological intuitions, unlike consequentialist intuitions, do not have any genuine normative force.

However, this is another bad argument. Presumably consequentialist intuitions are just as much a product of evolution—whether directly or indirectly—as deontological intuitions are, so an appeal to evolutionary history gives us no reason to privilege consequentialist intuitions over deontological ones. At one point Singer contends that consequentialist intuitions "[do] not seem to be . . . the outcome of our evolutionary

60. Greene, "Secret Joke," p. 43.

61. Singer, "Ethics and Intuitions," pp. 347–48. (Actually, it is somewhat controversial whether the emotional underpinnings of our moral judgments have been retained in relatively unchanged form since our early ancestors. For some empirical evidence that this might not be so, see Woodward and Allman, "Moral Intuition: Its Neural Substrates and Normative Significance," pp. 183, 187–88.)

past,"[62] but I find this claim rather hard to believe. And at one point Greene declares that "it is unlikely that inclinations that evolved as evolutionary by-products correspond to some independent, rationally discoverable moral truth," without realizing that such a claim poses as much a problem for the epistemic efficacy of consequentialist inclinations as it does for the epistemic efficacy of deontological inclinations.[63] In fact, a crass evolutionary argument of this sort poses problems for more than that. Anyone drawing normative implications from scientific findings is committed to mathematical and scientific judgments having genuine normative force, yet presumably our faculty for making such judgments also has an evolutionary basis. Sensing this sort of worry, Singer calls for us to engage in "the ambitious task of separating those moral judgments that we owe to our evolutionary basis and cultural history, from those that have a rational basis."[64] However, this is clearly a false dichotomy.

Richard Joyce and Sharon Street offer more careful versions of the argument from evolutionary history, yet their conclusion is that all of our moral judgments are unjustified (Joyce), or that all of our value judgments would be unjustified if certain realist conceptions of value were true (Street).[65] The crucial premise in both Joyce's and Street's argument is that moral judgments/intuitions don't need to be truth-tracking in order to conduce toward reproductive fitness, or at least that on a realist construal of what they amount to they don't need to; it is this premise that gives Joyce and Street some hope of undercutting the epistemic status of moral judgments without also undercutting the epistemic status of mathematical and scientific judgments. So maybe one could argue that although deontological intuitions don't need to be truth-tracking in order to conduce toward reproductive fitness, consequentialist intuitions do, and in this way resuscitate the argument from evolutionary history. However: it is far from clear how this additional piece of argumentation would go. Also: now all the work in the argument

62. Singer, "Ethics and Intuitions," p. 350.

63. Greene, "Secret Joke," p. 72.

64. Singer, "Ethics and Intuitions," p. 351.

65. Richard Joyce, *The Myth of Morality* (Cambridge: Cambridge University Press, 2001), chap. 6; Richard Joyce, *The Evolution of Morality* (Cambridge, Mass.: MIT Press, 2006); and Sharon Street, "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies* 127 (2006): 109–66.

is being done by armchair theorizing about the connection between being truth-tracking and being evolutionarily beneficial; the neuroscientific results have completely dropped out of the picture.[66]

V. NORMATIVE IMPLICATIONS: A BETTER ARGUMENT

Now that we have set aside three bad, but tempting, arguments for why Greene et al.'s neuroscientific findings have normative implications, we can consider Greene and Singer's main argument for that conclusion. The crucial move they make is to insist that if Greene et al.'s research is correct, then our deontological intuitions are *responding to factors that are morally irrelevant*, and as such should not be trusted. This suggests the following argument:

The Argument from Morally Irrelevant Factors:

P1. The emotional processing that gives rise to deontological intuitions responds to factors that make a dilemma personal rather than impersonal.
P2. The factors that make a dilemma personal rather than impersonal are morally irrelevant.
C1. So, the emotional processing that gives rise to deontological intuitions responds to factors that are morally irrelevant.
C2. So, deontological intuitions, unlike consequentialist intuitions, do not have any genuine normative force.

When summarizing his central argument, Greene writes, "There are good reasons to think that our distinctively deontological moral intuitions . . . reflect the influence of morally irrelevant factors and are therefore unlikely to track the moral truth."[67] And when responding to a commentator on that article, Greene adds, "I have . . . argued that these [deontological] judgments can be explained in terms of patterns of

66. Moreover, even if Greene and Singer could somehow adapt Joyce's and Street's arguments for their purposes, there is another reason why I don't think this strategy would work: I don't believe that Joyce's and Street's original versions of the evolutionary argument are convincing. I argue for this claim in a companion piece to this article titled "The Metaethical Irrelevance of Evolutionary Theory."

67. Greene, "Secret Joke," pp. 69–70.

emotional response and that these patterns reflect the influence of morally irrelevant factors."[68] Similarly, Singer writes,

> If . . . Greene is right to suggest that our intuitive responses are due to differences in the emotional pull of situations that involve bringing about someone's death in a close-up, personal way, and bringing about the same person's death in a way that is at a distance, and less personal, why should we believe that there is anything that justifies these responses? . . . [W]hat is the moral salience of the fact that I have killed someone in a way that was possible a million years ago, rather than in a way that became possible only two hundred years ago? I would answer: none.[69]

And elsewhere in that same article, Singer insists that "there are no morally relevant differences" between the trolley driver and footbridge dilemmas, which is why we should ignore our emotional responses to the latter sort of dilemma.[70]

What are we to make of the argument from morally irrelevant factors? The first thing to note is that, as Greene and Singer both admit,[71] premise P2 in this argument appeals to a substantive normative intuition, which presumably one must arrive at from the armchair, rather than directly read off from any experimental results; this is why the argument does not derive an "ought" from an "is." I believe that this feature is a virtue of the argument; however, it is also its ultimate undoing. In what follows, I mention four worries that I have about the argument from morally irrelevant factors, in order of increasing significance.

First Worry: Since Greene et al.'s initial characterization of the personal versus impersonal dilemma distinction does not track the gives-rise-to-a-deontological-judgment versus gives-rise-to-a-consequentialist-judgment distinction, it is far from clear that premise P1 is true. Greene thinks that the eventual account of the features to which the deontological moral faculty is responding will have something

68. Greene, "Reply to Mikhail and Timmons," p. 117. See also Greene, "Secret Joke," p. 70, where he refers to "contingent, nonmoral feature[s]"; Greene, "Secret Joke," p. 75, where he again refers to "morally irrelevant factors"; and Greene, "Reply to Mikhail and Timmons," p. 116, where he refers to "arbitrary features" in the course of arguing that deontologists suffer a "garbage in, garbage out" problem.

69. Singer, "Ethics and Intuitions," pp. 347, 348.

70. Ibid., p. 350.

71. See Greene, "Secret Joke," pp. 66–67; and Singer, "Ethics and Intuitions," p. 347.

to do with "personalness," broadly construed,[72] but I have my doubts.[73] Moreover, any attempt to precisely characterize the features that give rise to distinctively deontological judgments reintroduces many of the intricacies of the original trolley problem: formulating a principle that

72. Greene, "Reply to Mikhail and Timmons," p. 112.

73. In a more recent study (Joshua D. Greene, Fiery Cushman, Lisa E. Stewart, Kelly Lowenberg, Leigh E. Nystrom, and Jonathan D. Cohen, "Pushing Moral Buttons: The Interaction between Personal Force and Intention in Moral Judgment," *Cognition* 111 [2009]: 364–71), Greene and his colleagues claim to have discovered that what explains people's moral judgments about footbridge-like cases is whether the agent's action *intentionally harms* someone through the use of what they call *personal force*, which is present when "the force that directly impacts the victim is generated by the agent's muscles" (p. 364). However, there are a number of problems with the study. For instance, many of the contrasting cases have a variety of differences beyond those identified as candidate explanatory factors, and a number of obvious potential counterexamples to their proposal were not tested (for example, do people judge it just as morally unacceptable to force the man off the footbridge by menacing him with a knife, or by threatening to harm his family, or by tricking him into taking a step backwards?). But, ironically, the biggest problem with the study is that Greene et al. seem to have identified, without realizing it, a competing explanation for their respondents' verdicts. Greene et al. gathered evidence about the degree to which their respondents unconsciously filled in more realistic assumptions when imagining the scenarios in question, and they found a high degree of correlation between a tendency to refuse to assume that it was absolutely certain that the five would be saved if the one is killed and a tendency to judge that such a course of action is morally unacceptable (pp. 367–68). So, by their own lights, not all of their subjects were responding to the same scenario, and the variation in responses can be partially explained by the variation in assumptions about the likelihood of the proposed action succeeding. (I suspect that varying assumptions about the degree to which the man on the footbridge might resist, thereby endangering the life of the agent trying to harm him, could also go a long way toward explaining people's differing verdicts.)

More importantly, though, it is simply a mistake to think that by merely surveying people's opinions about the moral permissibility of certain actions, we can empirically study what sorts of factors elicit characteristically deontological judgments. All these studies tell us is that these people make certain moral judgments about certain scenarios, and certain other moral judgments about certain other scenarios; which of these judgments count as characteristically deontological is not something we can just read off from the empirical results. (Sometimes Greene suggests that we can sidestep this worry by postulating that philosophers are confused about the meaning of the term "deontology," and although they think it refers to an abstract moral theory, in fact it refers to a psychological natural kind, namely the verdicts of the emotional subsystem; see his "Secret Joke," pp. 37–38. However, in making this claim Greene is committing himself to an incredibly controversial claim in the philosophy of language. It is one thing to say that although we think "water" refers to one sort of physical substance, in fact it refers to another sort of physical substance. Greene's claim, however, is akin to saying that although we think that "Goldbach's conjecture" refers to an abstract mathematical theory, in fact it refers to the physical process of digestion, or to saying that although we think that the name "Barack Obama" refers to a certain person, in fact it refers to the number seven.)

distinguishes what separates cases-eliciting-a-deontological-judgment from cases-eliciting-a-consequentialist-judgment is likely to be as difficult as the old problem of formulating a principle that distinguishes the permissible options in trolleylike cases from the impermissible ones. After all, Greene et al.'s initial "ME HURT YOU" criteria were inspired by Thomson's proposed solution to the trolley problem in her 1985 article "The Trolley Problem," and it fell victim to Kamm's Lazy Susan Case, which was originally offered as a counterexample to Thomson's very proposal. So settling on a fully adequate account of the sorts of features to which deontological judgments are responding is likely to be an extremely difficult, if not impossible, task, and until that task has been completed, we cannot be sure whether P1 is true.

Second Worry: Even if we *were* able to find a way of characterizing the factors which deontological judgments are responding to that makes P1 true, it is far from clear that P2 would still seem plausible. It is one thing to claim that a faculty which responds to *how "up close and personal" a violation is* is responding to morally irrelevant features, but quite another thing to claim that a faculty which responds to *whatever the sorts of features are that distinguish the footbridge case from the trolley driver case* is responding to morally irrelevant features. Once we fix on what those features are, P2 may well strike us as false.[74]

Third Worry: Even if P2 *does* strike us as true, the argument's conclusion does not follow, for C2 does not follow from C1. Suppose we deem some of the features triggering deontological intuitions to, intuitively, be morally irrelevant, thus granting P2. This is a strike against deontological intuitions. However, we can only conclude that consequentialist intuitions should be privileged over deontological intuitions if a parallel case

74. Indeed, there is a sense in which this objection is already apropos even if we assume Greene et al.'s "ME HURT YOU" proposal to be fully adequate. Greene performs a sort of shell game here: first he proposes that the dilemmas eliciting deontological reactions are dilemmas that, intuitively, involve harm committed in an "up close and personal" manner, and then he glosses dilemmas that, intuitively, involve harm committed in an "up close and personal" manner in terms of the "ME HURT YOU" criteria. However, when it comes time to decide whether deontological judgments are responding to morally relevant factors, Greene switches back to evaluating things in terms of the intuitive up-close-and-personal-harm distinction, rather than in terms of the "ME HURT YOU" criteria. However, it's one thing to say that *whether one has committed a harm in an "up close and personal" manner* is a morally irrelevant factor, and quite another thing to say that *whether one has initiated a new threat that brings about serious bodily harm to another individual* is a morally irrelevant factor.

cannot be made against consequentialist intuitions. Moreover, it is open to the defender of deontology to reply that, intuitively, the faculty eliciting consequentialist reactions is also responding to morally irrelevant factors, or failing to respond to morally relevant ones. For example, a deontologist could contend that the neural processes giving rise to consequentialist judgments are failing to respond to morally relevant factors by ignoring the separateness of persons, or by treating people as vats of well-being, or by assuming that all value is to-be-promoted, or by making morality incompatible with integrity, or . . . [insert your favorite anticonsequentialist intuition here]. So basically we have just recapitulated the same old battle of intuitions over the plausibility of consequentialism versus deontology in our evaluation of which sorts of factors are and are not morally relevant.

This problem leads to my most pressing worry: the neuroscientific results seem to be doing no work in this argument. The epistemic efficacy of consequentialist versus deontological intuitions now appears to be purely a function of *what sorts of features out there in the world they are each responding to*. We have three distinctions on the table:[75]

(1) dilemmas that engage emotion processing versus dilemmas that engage "cognitive" processing;
(2) dilemmas that elicit deontological judgments versus dilemmas that elicit consequentialist judgments;
(3) personal moral dilemmas versus impersonal moral dilemmas.

Greene et al.'s dual-process hypothesis posits that the first of these distinctions matches up with the second. In order to experimentally assess this hypothesis, Greene and his colleagues identified the second distinction with the third one, and then directly tested whether the first distinction matches up with the third. But the argument from morally irrelevant factors *only depends on Greene et al.'s identification of the second distinction with the third one*. Thus the neuroscientific results are beside the point. In particular:

75. More precisely, we have four distinctions on the table, since we need to distinguish between the intuitive way of cashing out the personal versus impersonal moral dilemma distinction, and the more regimented way of cashing out that distinction in terms of the "ME HURT YOU" criteria.

> The "emotion-based" nature of deontological intuitions has no ultimate bearing on the argument's cogency. (Delete "emotional" from P1 and C1, and the argument is just as plausible.)
>
> Issues about the evolutionary history of our dispositions to have deontological and consequentialist intuitions are also irrelevant to the argument's cogency.
>
> Even the claim that these two sets of intuitions stem from separate faculties is irrelevant to the argument's cogency. (The argument would be just as plausible if it turned out that only one faculty was responding to two different sorts of factors.)

So the appeal to neuroscience is a red herring: what's doing all the work in the argument from morally irrelevant factors is (a) Greene's identification, from the armchair, of the distinction between dilemmas-eliciting-deontological-reactions and dilemmas-eliciting-consequentialist-reactions with the distinction between personal and impersonal moral dilemmas, and (b) his invocation, from the armchair, of a substantive intuition about what sorts of factors out there in the world are and are not morally relevant.

The basic problem is that once we rest our normative weight on an evaluation of the moral salience of the factors to which our deontological and consequentialist judgments are responding, we end up factoring out (no pun intended) any contribution that the psychological processes underlying those judgments might make to our evaluation of the judgments in question. So we are left with a dilemma: appeal to a substantive intuition about what sorts of factors are morally relevant, and the neuroscientific results drop out of the picture; or keep those results in, and it looks as though our only recourse is to one of the bad arguments we have already dismissed.

Thus I conclude that the argument from morally irrelevant factors does not advance the dialectic on the relative merits of deontology versus consequentialism. No reasonable philosopher is going to deny that it makes no moral difference whether one harms someone with one's bare hands or from a distance. However, deontologists are most definitely going to deny that so crude a distinction is what really underlies their distinctively deontological moral judgments. And once we have in hand the true account of what sorts of factors underwrite deontological judgments about cases, I claim that evaluation of their moral relevance will

depend on a substantive normative judgment as to whether observing those sorts of moral distinctions is more or less plausible than ignoring the sorts of moral distinctions that consequentialists typically ignore. The appeal to neuroscience provides no new traction on this old debate.[76]

VI. AN INDIRECT ROLE FOR NEUROSCIENCE?

One of the things that seemed so exciting about Greene's research was that it promised a new way of finally resolving the trolley problem. Rather than having to hit upon a compact, exceptionless principle that delivers the intuitive verdict about every trolleylike case, we could use the neuroscientific results and some philosophical theorizing to discount certain of those intuitive verdicts, making it easier to find a principle that fits the leftover data. However, one of the things I have just argued is that in order to use the neuroscientific results and some philosophical theorizing to discount certain intuitive verdicts about trolleylike cases, Greene in effect needs to have already solved the trolley problem. Since the argument for why we should discount a certain set of intuitive verdicts depends on evaluating the features to which those intuitive verdicts are responding, we need an exceptionless (but not necessarily compact) principle delineating the sorts of features that make those sorts of intuitive verdicts kick in. So there is one way in which Greene's task is more difficult than the traditional trolley problem: rather than needing to find a single principle that states *what it takes for an option in a trolleylike dilemma to count as permissible or impermissible*, he needs to find two principles: one principle stating *what it takes for the deontological faculty to count an option in a trolleylike dilemma as permissible or impermissible*, and another principle stating *what it takes for the consequentialist faculty to count an option in a trolleylike dilemma as permissible or impermissible* (plus an account of how conflicts between the two faculties are resolved). But

76. To see how little work the neuroscience is now doing in the argument from morally irrelevant factors, consider this: should we perform additional experiments to see what parts of the brain light up when certain people make a judgment *that such-and-such-factors-picked-out-by-deontological-judgments are not morally relevant*, and what parts of the brain light up when other people make a judgment *that such-and-such-factors-ignored-by-consequentialist-judgments are in fact morally relevant*? What would that possibly tell us? (Shall we then perform yet more experiments to see what parts of the brain light up when people make a judgment about the relevance of the factors to which those second-order judgments are responding, and so on, ad infinitum?)

there is another way in which Greene's task is easier than the traditional trolley problem: he doesn't need his account of the features to which either of these faculties is responding to be rationally defensible. Moreover, it seems that there is a way in which neuroscience *could* play an indirect role in this task that Greene has set for himself—and, by extension, a way in which neuroscience *could* play an indirect role in more traditional attempts at solving the trolley problem.

Suppose we have established that a certain region of the brain is activated when we contemplate a certain class of cases that yield characteristically deontological verdicts about what it is morally permissible to do. Suppose, also, that we have independent knowledge that in nonmoral cases this brain region is recruited to distinguish between (say) intentional and nonintentional action. Then we might try seeing whether what distinguishes this class of moral dilemmas from others has something to do with the intentional versus nonintentional action distinction. Since neuroscience only provides evidence of correlations, it is not certain that when the brain region in question is recruited for moral cases it is responding to the same sorts of features as when it is recruited for nonmoral cases. But the neuroscientific results can give us clues for where to look when trying to characterize what sorts of features out there in the world each moral faculty is responding to. And this is true whether our ultimate aim is to debunk or to vindicate those verdicts. However, note that, even here, the neuroscientific results play no role *after* we have the principles stating what sort of features each faculty is responding to: at that point, the argument for whether we should or should not discount the verdicts of one of these faculties proceeds entirely via armchair theorizing about whether the sorts of features to which that faculty is responding are or are not morally relevant. Still, providing clues for where to look when attempting to characterize the features to which distinctively deontological and distinctively consequentialist judgments respond is no small matter.[77]

Could neuroscience play a more direct role in our theorizing about the evidential status of moral intuitions? It seems to me that the best-case scenario is this:

77. Although Greene et al.'s attempt in "An fMRI Investigation" at characterizing the features that give rise to deontological judgment did not rely on neuroscience in the way I have just sketched, there is a sense in which *evolutionary theory* played just that sort of indirect role, since evolutionary considerations are what partially led them to try out the proposal they put forward.

> *The Best-Case Scenario*: We notice that a portion of the brain which lights up whenever we make a certain sort of obvious, egregious error in mathematical or logical reasoning also lights up whenever we have a certain moral intuition.

In this case, should we discount the moral intuition? That depends on how we fill in the details of the case. If, for all we can see, there is no connection between the content of the moral intuition and the content of the mistaken bit of mathematical/logical reasoning, then I am inclined to think we should continue to trust the intuition and hold out for later neuroscience to make finer distinctions between the portions of the brain activated in the moral and nonmoral cases. (Suppose the same part of the brain that lights up whenever we affirm the consequent also lights up whenever we have an intuition that infanticide is impermissible; would you be willing to start killing babies on those grounds?) If, on the other hand, we come to see that the moral intuition in question rests on the same sort of confusion present in the mistaken bit of mathematical/logical reasoning, then of course we should discount the moral intuition, but in that case the neuroscience isn't playing a direct justificatory role. Again, we might not have thought to link the moral intuition to that sort of mathematical/logical blunder if we hadn't known the neuroscientific results; but again, once we do link them, it seems that we do so from the comfort of an armchair, not from the confines of an experimental laboratory. It is as if, while trying to prove whether or not some mathematical claim is true, your mathematician friend had said to you, "Why don't you try using the Brouwer fixed point theorem?" If you end up proving the claim to be true using that theorem, your justification for the claim in no way depends on your friend's testimony. (After all, she didn't give away whether she thinks the claim is true or false.) Nonetheless, your friend's testimony gave you a hint for where to look when trying to prove or disprove the mathematical claim. So too, I speculate, neuroscience can provide hints for where to look during our normative theorizing, but ultimately it can play no justificatory role in that task. Despite Greene's and Singer's claims to the contrary, learning about the neurophysiological bases of our moral intuitions does not give us good reason to privilege certain of those intuitions over others.