

Dear Readers,

What follows is a set of notes that I prepared for a small meeting of philosophers and scientists at Arizona State University. Attendees circulated papers to one another prior to the meeting. Selim Berker, one of the attendees, circulated his paper “The Normative Insignificance of Neuroscience,” which criticizes my work. I was asked to prepare a reply. My reply, for now, consists of the following set of notes. My plan is to turn these notes into a proper manuscript, suitable for publication.

You are welcome to cite this set of notes. You are welcome to pass this set of notes on to whomever you please. My only conditions on your doing so are (a) that you pass on this document in its entirety *including this message* and (b) that any citation of these notes makes it clear that this is a work in progress. If you cite this document, I ask that you cite it as “Notes on ‘The Normative Insignificance of Neuroscience’ by Selim Berker.”

Feedback is very much welcome.

JDG

12/9/10

Notes on “The Normative Insignificance of Neuroscience” by Selim Berker

Joshua Greene, Nov. 2010

Selim Berker (2009) presents a thoughtful critique of my ideas concerning the relationship between recent scientific research on moral judgment and longstanding issues in normative ethics. Berker has forced me to clarify my views and sharpen my arguments, and for this I am genuinely grateful.

I am, to put it mildly, less enthusiastic about Berker’s critique of my scientific methodology, which appears in Section III of his paper. Here, Berker’s commentary is extremely misleading and in some instances makes statements that are simply false, reflecting misunderstandings of statistical methods and experimental design. In challenging the evidence for my dual-process theory of moral judgment, Berker relegates most of the relevant evidence to footnotes, where that evidence is hastily dismissed, and much the most relevant evidence is never cited at all.

The primary purpose of this set of notes is to sketch out a reply to Berker’s critique of the normative argument presented in Greene, 2007b (“The Secret Joke of Kant’s Soul,” appended). This normative argument, however, is only of interest to the extent that the science on which it is based is reliable. Therefore, before moving on to the defense and further development of my normative argument, I must correct the scientific record. In the following section I will briefly summarize Berker’s errors and explain why his critique is misleading. Further details concerning Berker’s erroneous and misleading presentation follow in the appendix at the end of this document.

Summary of problems with Berker’s methodological critique

I’ll begin with Berker’s false statements.

1. The statistical test used in Greene et al. 2001 is not “statistically invalid.”

Berker claims that the statistical analysis used to analyze the reaction time data in my 2001 fMRI paper is “statistically invalid.” This is false. The test we used (a mixed effects ANOVA, modeling subject as a random effect and judgment type and condition as fixed effects) is a perfectly standard test and by no means “statistically invalid”—not in this instance or in general. (See, for example, Howell, 2001, Chapter 14). On the contrary, good methodological practice *required* us to perform a test of this kind in this instance.

That said, Berker does make a valid point about this analysis. Had we performed an additional statistical test, sometimes known as an “item analysis,” we would have realized that the results from the reaction time data (not the fMRI data, which were the paper’s primary focus) do not add additional support to the conclusions drawn in the paper. Most researchers in our field do not perform item analyses, though they probably

should do so more often, and in this case doing so would have made an important difference. My colleagues and I were not as statistically savvy as we could have been, but it is simply false to say that we conducted a “statistically invalid” test, an accusation implying a lack of basic methodological competence.

I became aware of the problem with the 2001 reaction time data after the paper was published. My awareness of the problem was one of my motivations for performing a subsequent study, which makes the same point that we attempted to make in the 2001 paper, but in a better way. This brings us to Berker’s next set of errors.

2. The test used in Greene et al. 2008 is also not “statistically invalid.” Moreover, Berker dismisses the results of this study because he fails to understand that we also performed the kind of statistical test that he advocates (an item analysis).

The aforementioned more recent paper was published in *Cognition* in 2008. In short, this paper shows that performing a distracting secondary task significantly decreases reaction time for characteristically utilitarian judgments, but not for characteristically deontological judgments. This result supports the dual-process theory by showing that utilitarian moral judgments are preferentially associated with additional controlled cognitive processes, the point we were attempting to make with the reaction time data in the 2001 paper. This paper also provides an explanation for why we did not observe the predicted effect in the 2001 paper.

In footnote 37 of his paper, Berker erroneously dismisses the results of the 2008 paper on the grounds that it uses the same “statistically invalid” analysis. First, as noted above, this analysis is required and is by no means invalid. Second, in the 2008 study we (a) excluded from analysis the items that distorted the results in the 2001 study and (b) performed an item analysis in addition to the standard analysis. Employing both of these safeguards, we obtained the predicted results.

In an unwitting contradiction, Berker concludes footnote 37 as follows:

Greene et al. continue to present their response-time data in the statistically invalid manner... On the other hand, they write in that same study, “This general pattern also held when item, rather than participant [subject], was modeled as a random effect, though the results in this analysis were not as strong” (ibid., p. 1149).

The second sentence in the above quotation includes a direct quotation from the 2008 paper. Berker quotes this material directly because he apparently does not understand what it means. What it means is that we performed the analysis the way he recommends and still got the predicted results.

In sum, the 2008 study was designed in part to address the problem that Berker highlights in his objection to the 2001 reaction time data. Unfortunately, Berker mentions this study only in footnotes where it is erroneously dismissed.

3. Greene et al (2009) did not identify “without realizing it” a competing explanation for their results. Berker fails to understand that the analyses in this study *control for the competing explanation in question.*

Of all of the empirical papers I’ve published, the one with the most direct relevance to normative issues was recently published in *Cognition* (2009). (See appended) This paper achieves two relevant goals. First, it revises the personal/impersonal distinction presented in the 2001 paper, the distinction we tentatively used to distinguish the *switch* and *footbridge* dilemmas (among others) from each another. (What I call this *switch* case is often called the “bystander” case. I prefer *switch* because it highlights the physical mechanism, which is critically important for my purposes.) Second, this paper identifies “personal force”—roughly, the distinction between directly pushing someone vs. harming someone more indirectly, e.g. by hitting a switch—as a critical psychological factor in the Trolley Problem. (This is important for our normative discussion because “personal force” does not appear to be a difference that ought to make a difference. More on this later.)

A common complaint about trolley dilemmas is that one can account for people’s judgments simply by assuming that people are importing real-world assumptions that conflict with the premises of the dilemma. For example, people might disapprove of pushing the man off the footbridge because they think that pushing him won’t actually save the lives of the five on the tracks. In our 2009 paper, we employed a novel approach to this problem. We collected information about people’s “real-world” expectations in these dilemmas and then included that information in our statistical analyses. The idea is that *if* these judgments are driven by real-world expectations, the judgments for the different dilemmas should all look roughly the same after one has *controlled for* people’s real-world expectations.

Using this method, we found that people’s real-world expectations have a small but statistically significant effect on their judgments and that this alternative hypothesis *cannot* explain our main results. For example, the difference between the ratings for the standard switch and footbridge cases is only reduced by about 10% when one controls for people’s real-world expectations. All of the results reported in the 2009 paper were obtained *after having controlled for the effects of people’s real world expectations.*

In footnote 73, Berker writes:

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 ... 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.

Berker misunderstands these results. We did not, as Berker claims, identify a competing explanation “without realizing it.” We explicitly identified the relevant explanatory factor and explicitly *controlled for it*. The fact that this additional factor was, so to speak,

“worth controlling for” is not in any way a shortcoming of our study, let alone “the biggest problem with the study.”

I began with the three errors described above because they are the clearest instances of objectively false statements. There are other mistakes as well, albeit ones that require a bit more theoretical background to understand. For example, Berker describes a null result presented in our 2008 paper as a “troubling piece of counterevidence,” when it is not, in fact, counterevidence. Berker states that there are “obvious potential counterexamples” to the claims made in our 2009 paper, when in fact the examples alluded to are not counterexamples to the claims made in that paper. Berker misleadingly highlights ambiguous results in my 2004 fMRI paper as if they are dirty secrets that, once known, seriously compromise the paper’s main conclusions. And so on. (See Appendix.) However, the most serious problem with Berker’s critique is its general tendency to minimize, dismiss, or ignore entirely the available evidence for the dual-process theory.

4. Berker’s critique is extremely misleading. Most of the available evidence for the dual-process theory is mentioned only in footnotes, where it is hastily dismissed, and some of the most critical pieces of evidence are not cited at all.

The purpose of Berker’s methodological critique is to suggest that the evidence for the dual-process theory of moral judgment is surprisingly weak. Given this purpose, Berker should consider all of the available evidence, and his focus should be on the strongest evidence. (To Berker’s credit, and following a parallel “best practice” in philosophy, Berker attempts to formulate the strongest version of my normative argument before critiquing it.) And yet, the main text of Berker’s methodological critique is written as if the only evidence for the dual-process theory of moral judgment comes from my first two fMRI papers. This is a deliberate choice, as Berker explains in footnote 26:

“In the body of this article I have focused on the [2001 and 2004] 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....”

This is poor scientific scholarship. If one’s task is to critique a scientific theory for lack of evidential support, it is not acceptable to focus on a subset of the evidence simply because it is “particularly vivid” and because journalists have written about it. If my early work “captured the public’s imagination” it is because it raised novel questions and suggested some intriguing answers, not because it conclusively answered those questions. To focus on only the earliest pieces of evidence is to seriously distort the scientific picture. For example, Berker’s readers would never know that researchers from three different labs (at UCLA, the University of Iowa, and the University of Bologna) read my fMRI papers and independently recognized that the dual-process theory makes specific predictions concerning the moral judgments of neurological patients. All three labs tested those predictions and confirmed them. For Berker to deny his readers this information while casting doubt on the evidence for my theory is irresponsible.

Berker begins his article by dismissing my research as part of a fad—an inevitable collision between newfangled brain imaging technology and ethics:

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.”... by one estimate, an average of eight peer-reviewed articles employing fMRI were published per day in 2007. So perhaps it was inevitable that empirically minded philosophers would take some of these fMRI studies to have profound implications for philosophy. (pg. 293)

Had Berker paid more attention to the evidence beyond my first two fMRI papers, he would have realized that my fMRI research cannot be reduced to “ethics meets fMRI.” My fMRI research was designed to test a psychological theory, and that theory is in no way bound to the technology that was first used to test it. Indeed, several independent researchers correctly recognized that my original fMRI data support a theory that makes a wide range of testable predictions using a wide range of methods. These methods include the testing of neurological patients (Mendez et al, 2005; Koenigs et al., 2007, Ciaramelli et al, 2007) as well as behavioral methods such as emotion induction (Valdesolo & DeSteno, 2006), interference by cognitive load (Greene et al, 2008), and the examination of individual differences in cognitive style (Bartels, 2008; Moore et al., 2008). And since Berker’s manuscript was written, researchers have tested the dual-process theory using other methods such as psychophysiology (Moretto et al., 2010) and pharmacological interventions (Crockett et al., 2010). This is not fad science. On the contrary, this kind of international, multi-methodological production of convergent evidence is a model of scientific progress. It is especially unfortunate that Berker chooses to dismiss this research as a fad because, in doing so, he denies the field of philosophy credit for having had a positive influence on science.

There is one passage in Berker’s methodological critique with which I whole-heartedly agree:

In general it is dangerous... 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. (pg. 305)

I’ll only add that, if the scholarship is particularly poor, the points made in a philosopher’s critique may also be overcome by past research. It is regrettable that Berker’s methodological commentary was not itself subjected to scientific peer review.

In contrast to Berker’s critique of my scientific research, his critique of my normative agenda is admirably clear and entirely reasonable. That said, I do not agree with it, and I will explain why in what follows.

Why cognitive (neuro)science matters for ethics

Berker summarizes his argument in the second paragraph of his paper:

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. [emphasis mine] (pg. 294)

My fundamental disagreement with Berker resides in the contrast between the two underlined portions of the above passage. Berker believes that the following two statements are equivalent:

Greene's argument relies on substantive normative judgments that are independent of the scientific results.

Scientific results do no work in Greene's argument

These statements are not equivalent. The first statement is true. The second is false.

Like many philosophers, I believe that one cannot derive a substantive moral "ought" from a scientific "is" (Greene, 2003). More specifically, I agree with Berker that substantive moral conclusions cannot be deduced from scientific findings, neuroscientific or otherwise. Thus, as Berker argues, any valid normative conclusions reached on the basis of scientific research must also invoke one or more non-scientific normative premises. However, it does not follow from this conclusion that scientific results inevitably do "no work" in such normative arguments. In what follows I'll explain how scientific results can have normative implications—that is, how they can do important work in normative arguments—without illicitly hopping the is/ought gap.

It's not about neuroscience

Before we move on, the central issue must be reframed. Berker asks whether *neuroscience* has any normative implications. In his introduction, he frames the issue even more narrowly, questioning the moral implications of *brain imaging* data. This narrow framing threatens to caricature my arguments. One imagines a benighted neuroscientist, clad in white coat, triumphantly waiving a pixilated picture of a human brain: "*Eureka! Kant was wrong! It's all here in this brain scan!*" Such a character is easy to dismiss.

I have never attempted to draw normative conclusions from brain scans alone, or even from neuroscientific data alone. My ideas and my research program—at least as they relate to moral philosophy—are not essentially neuroscientific. My dual-process theory of moral judgment is a psychological theory, and neuroscientific data are just one among several types of supporting evidence (Greene 2007b, 2009b). The dual-process theory *is* supported by neuroscientific evidence, but it could also be well-supported *without* neuroscientific evidence. In light of this, one might wonder why one would go through the trouble and expense of collecting neuroscientific data if cheaper, lower-tech data will do.

First, studying the neuroscience of moral judgment adds value to neuroscience. Second, it adds value to both neuroscience and psychology by furthering their integration. Third, neuroscientific tests of psychological theories can be especially powerful because neuroscientific results, unlike behavioral results, cannot be predicted by common sense. (Neuroscientific results are rarely met with “duh.”) Fourth, understanding the neuroscience of moral judgment may also play a valuable role in dispelling mind-body dualism (Greene, in press)—a worthwhile philosophical endeavor, but not one that is likely to challenge the views of contemporary ethicists. (Insofar as my early fMRI work “captured the public’s imagination” it is probably because of the challenge it poses to mind-body dualism. See Bloom, 2006) In short, studying the neuroscience of moral judgment is valuable for several reasons that have little to do with normative ethics. Thus, if my fMRI studies of moral judgment, taken in isolation, fail to have definitive moral implications, this may be because having moral implications is not their most immediate purpose.

I do believe that neuroscientific data have implications for normative ethics, but I believe that the relevance of neuroscience is rather indirect and that it depends on further non-neuroscientific ideas and assumptions. In my writings that have explored the possible normative implications of neuroscientific research, I have discussed the neuroscientific evidence in the context of a much broader set of psychological and biological ideas (Greene, 2003; Greene, 2007b). It is this broader set of ideas from which the normative implications follow, if any do. Thus, to frame the present debate as one concerning the normative significance of neuroscience is misleading.

With this in mind, I will respond to Berker’s argument in three stages. First, I will give two examples that illustrate how psychological research can “do work” in a normative argument. Then, using these examples as models, I will explain how behavioral research in moral psychology “does work” in my normative argument. Third, I will explain how neuroscience, by supporting the dual-process theory of moral judgment, “does work” in my normative argument, albeit more indirectly.

The role of scientific premises in normative arguments: Two examples

I'll begin with a simple example that illustrates how scientific evidence can do important work in a normative argument.

Let's begin with this question: Do juries in capital cases make fair decisions? This is clearly a normative question, as well as a controversial one. There is a lively debate concerning whether and to what extent capital juries make unfair decisions. Can science help? Scientific research supports the following descriptive claim.

D1. Juries' decisions are affected by the race of the defendant (Baldus et al, 1998; Eberhardt et al., 2005).

May we now reach the following normative conclusion?

NC1. Juries in capital cases make (at least some) decisions that are unfair.

Not yet. To do that, we need an additional normative premise.

NP1. Capital juries ought to regard a defendant's race as irrelevant.

It is only in conjunction with NP1 that D1 can play a role in getting us to NC1. Moreover, NP1 is in no way supported by science, at least not any we've discussed. The truth of NP1 depends entirely on the "intuitions" of the people who accept it. And yet, it would be crazy to say that descriptive premise D1 does "no work" in this argument. Lesson: It's a mistake to assume, as Berker does, that science does "no work" in a normative argument simply because the argument depends critically on "substantive normative intuitions" that go beyond the science.

At this point, one might agree that the science "does work" in the above normative argument, but deny that the above argument is usefully analogous to the kind of argument that I have made. More specifically, one might argue that the above argument does nothing to challenge anyone's *values*. Insofar as this argument is compelling, it's because present company is unanimously opposed to racism. In this case, the science merely draws our attention to the presence of racism, thus highlighting a specific application of our values. But if one were to defend racism (by denying NP1), this argument would get us nowhere. I am arguing that psychology can "do work" in determining which values we should have, and therefore one might object that an argument in which no one's values are challenged is not analogous.¹

I believe that the analogy is closer than this argument suggests. One might, for example, have the firm conviction that the juries in question, consisting as they do of upstanding citizens of the Great State of X, could never make such grossly unfair decisions. Almighty God, who speaks through their solemn verdicts, would not allow such injustice to enter their hearts. Now the above argument *does* challenge somebody's values, and the science continues to do important work in the argument. But let's consider a different

¹ Thanks to T.M. Scanlon for this point.

example that more clearly poses a scientific challenge to people's values. The following passage is from the current draft of my forthcoming book. (Pardon the missing references. Note also that I've dropped footnotes to save space.):

Take, for example, the taboo against incest. Science tells us that the incest taboo derives from a biological adaptation for avoiding genetic diseases[refs]: Offspring of close relatives are relatively likely to produce children with genetic diseases (due to a higher frequency of shared deleterious genes), and therefore an emotional aversion to mating with close relatives is adaptive. Now, consider the case of Danielle Heaney and Nick Cameron, a half-brother-sister pair living in the U.K. who met as adults and instantly fell in love.[dailymail, 2/7/08]. Their mother reported them to the police after walking in on them *in flagrante delicto*. "What you are doing is morally wrong!" she shouted. Others agree, including a judge who has threatened them with imprisonment if they continue their sexual relationship.

Let's suppose that Nick and Danielle intend to continue their relationship but have no intention of having biological children. Let's suppose further that Nick has had a vasectomy and that Danielle has had a tubal ligation, rendering their odds of having a child with a genetic disease far lower than that of an ordinary couple. Should the law keep them apart? And ought our scientific knowledge affect our answer? Some of us are repulsed by their behavior (Haidt, 2001), but if our repulsion is grounded in an adaptation for avoiding genetic diseases, and avoiding genetic disease is not a serious concern in their case, then why should we give that feeling of repulsion any moral weight?

Incest between consenting adults is more of a moral curiosity than a major social issue, but it nicely illustrates the relationship between the "is" of science and the "ought" of morality. In a limited way, the "is"/"ought" separatists are correct: The scientific "is" concerning the psychological origins of the incest taboo does not, by itself, tell us what we ought to do about Nick and Danielle. One can understand the relevant science and still (legitimately) ask, "But is what they're doing wrong?" And yet, the science is clearly relevant, and in a way that challenges our values rather than merely complementing them with useful information.

The Nick-and-Danielle problem involves a conflict between competing values. On the one hand, we in free societies have a presumption in favor of individual freedom. On the other hand, incest, even between consenting adults, strikes many people as deeply morally wrong. Our scientific knowledge casts this moral aversion in a new light, threatening its moral authority, perhaps enough to change our minds, if not our feelings, about Nick and Danielle. As this case illustrates, scientific knowledge can cause reasonable people to rethink their moral values, even if there is no strict, logical relationship between the "is" of science and the "ought" of morality.

Let's break this argument down as above. Science gives us a descriptive premise, something like this:

D2. People are opposed to incest between consenting adults because such behavior in the environment of our ancestors increased the probability of having children with genetic diseases.

For present purposes, it's not important whether science has definitively shown the above claim to be true. All that matters is that D2 is the kind of descriptive claim that can be supported by scientific research. Do we now have an argument in favor of letting Nick and Danielle do as they please? Again, not yet. We need an additional normative premise:

NP2. Whether or not a behavior increased the probability of deleterious consequences in the environment of our ancestors is irrelevant to its present moral acceptability, so long as this behavior does not also causes similar harm in our present environment.

We're not quite there yet. We will likely need the help of additional descriptive premises, for example:

D2.1. Incest among consenting adults who use reliable forms of birth control (etc) is not on the whole harmful.

But we're still not where we'd like to be. Here, in contrast to the previous example, the argument is a *debunking* argument. That is, it's an argument that undermines a set of values by explaining their adoption in a way that makes it unnecessary or unlikely that those values are true or otherwise defensible. Because the values that we are attempting to debunk have taken on a life of their own, independent of their origins, they won't die so easily. D2 explains why people do, as a matter of fact, oppose consensual adult incest. And with the help of NP2, D2 tells us that people's judgments are, in this instance, determined by their sensitivity to a morally irrelevant factor. However, this argument doesn't prove that there couldn't be some *other* justification for condemning incest among consenting adults. D2.1 rules out standard consequentialist justifications, but there could be others. Thus, we cannot definitively conclude that that consensual adult incest a la Nick and Danielle is morally acceptable, although something like the above argument is enough to convince many people (Paxton, Ungar, & Greene, submitted). We'll have to settle for something like this:

NC2. Insofar as consensual adult incest is not on the whole harmful, and insofar as we lack a non-intuition-based justification for condemning consensual adult incest, we have no reason to believe that it is wrong.

Still, that's pretty good. Once again, we have shown that scientific information can do critically important work in a normative argument, even if the argument requires one or more additional normative premises. Moreover, this argument shows that scientific

information can do work, not merely by drawing our attention to instances in which our shared values apply, but by challenging values that not all of us share.

The “Argument from Morally Irrelevant Factors”

Before critiquing what he regards as my strongest argument, Berker considers a number of others, which he calls the “Emotions Bad, Reasoning Good” argument, the “Argument from Heuristics,” and the “Argument from Evolutionary History.” Berker’s objections to each are in the spirit of G.E. Moore: Yes, these judgments are driven by emotion, but are those emotions leading us astray? Yes, these judgments depend on heuristics, but are those heuristics leading us astray? Yes, these judgments depend on features of our psychology that evolved for other purposes, but are those features of our psychology leading us astray?

Finally, Berker settles on a version of my argument that he calls the “Argument from Morally Irrelevant Factors.” In a nutshell, the claim is that deontological judgments made in response to trolley dilemmas reflect our sensitivity to morally irrelevant factors, and are therefore morally suspect.

I agree that this is the clearest way of putting the argument. The arguments rejected above, presented as deductive arguments, are clearly false. As I will explain below, however, whether a judgment is produced by a process that is emotional, heuristic, or a by-product of our evolutionary history is not unrelated to whether that judgment reflects a sensitivity to factors that are morally irrelevant. (See the camera analogy, below) It’s just that the relationship is a contingent, probabilistic, and empirical one, not a logical one. More on this later.

Berker sketches the Argument from Morally Irrelevant Factors as follows:

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.

I might quibble with the wording in various places. In P1 I would make it clear that the technical meaning of “personal” has been updated since 2001 and may undergo further revision. In P2, I would simply say that “personalness” in the above sense is morally irrelevant. With respect to C2, I think that “consequentialist intuitions” is misleading

(for reasons explained below), and I would word this conclusion less strongly. I would rather say that the results cast doubt on the reliability of deontological judgments without doing the same for consequentialist judgments. But, overall, I think this is a fair formulation of the argument.

Berker lays out four “worries” about the argument above.

Worry 1: Berker’s first worry is that P1 might be false. Unfortunately, Berker largely ignores and erroneously dismisses the evidence that provides the strongest support for P1 (Greene et al., 2009). I will describe this evidence below.

Worry 2: I confess that Worry 2 is not entirely clear to me. It reads:

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 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. (pg 324)

Here is my understanding of the above. At this point, Berker is not denying that personalness is morally irrelevant. But the factors that make an action personal might also endow an action with features that are morally relevant. To put this in terms native to experimental psychology, Berker is saying that personalness might be *confounded* with other features that are morally relevant. If that’s the argument, then I believe the argument is confused. P1 can only be supported if personalness is shown to have an effect that it *independent* of other factors. Thus, this is really a worry about P1, not P2. It’s the worry that something confounded with personalness, rather than personalness itself, is affecting people’s judgments. As I will explain below, the experimental evidence does a good job of showing that personalness itself has an effect (although this is not the only factor that has an effect). If that’s correct, then P1 is well-supported.

The move from P1 to P2 is not about confounding factors or otherwise about what’s “really going on” in the Trolley Problem. The move from P1 to P2 is justified by a simple, non-scientific normative judgment that, so far as I can tell, Berker does not dispute: Personalness *per se* is not morally relevant.

Worry 3: Here Berker acknowledges that it would be a “strike against” deontology if P2 were true, i.e. if deontological judgments are driven by a sensitivity to morally irrelevant factors. This is an important concession because it effectively acknowledges that science can “do work” in a normative argument. Here, Berker’s worry is that consequentialist judgments are similarly suspect. As I will explain below, the dual-process theory (aided by the camera analogy) counters this argument. Berker writes:

The “emotion-based” nature of deontological intuitions has no ultimate bearing on the argument’s cogency.

I disagree. The psychology behind deontological judgments is not only qualitatively different from the psychology behind consequentialist judgments. That psychology is different in a way that makes it more likely to go wrong, at least in certain contexts. In other words, the dual-process theory gives us less reason to accept a version of P1 in a parallel argument that would cast doubt on consequentialist judgments. More on this below.

Worry 4: Berker’s “most pressing” worry is that neuroscience plays no role in the argument. This shouldn’t be Berker’s most pressing worry. The neuroscientific elements of my research may have been essential for getting my name in the newspaper, but they are not essential to my normative argument. As noted above, and explained in more detail below, the dual-process theory of moral judgment *is* essential to my argument. The neuroscientific data are “significant” because they support the dual-process theory, but they are not essential because the dual-process theory can be supported, and is supported, in other ways. Berker summarizes Worry 4:

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.

Here, my reply is threefold.

First, the claim made in (a) is not “from the armchair.” It is in part definitional and in part empirical. It is effectively a matter of definition that the judgment typically elicited by the *switch* case (better to hit the switch in order to save more lives) is a characteristically consequentialist judgment, by which I mean the kind of judgment that is easily justified in consequentialist terms. (For more on “characteristically” see Greene, 2007b, appended.) Likewise, it is effectively a matter of definition that the judgment typically elicited by the *footbridge* case is characteristically deontological. Next, there is an empirical psychological question concerning which factors cause most people to approve of the action in the *switch* case but not in the *footbridge* case. I have argued based on *empirical evidence*—evidence that Berker erroneously dismisses—that personalness is one of these factors.

Second, it is true that I have invoked “from the armchair,” a substantive intuition about what sorts of factors out there in the world are and are not morally relevant. And it is true that my doing this plays a key role in my argument. However, this “substantive intuition” is one that nearly all of us share, whether or not we have deontological or consequentialist proclivities. And, thus, my making this assumption is not question-

begging. As I will explain below, this assumption is analogous NP1 and NP2 in the above arguments.

Third, as noted above, it is incorrect to say that (a) and (b) do *all* the work. The dual-process theory explains why a parallel argument casting doubt on consequentialism is unlikely to go through. What follows would be the first premise in such an argument:

Anti-Consequentialist P1. The emotional processing that gives rise to consequentialist intuitions responds to factors that make a dilemma personal rather than impersonal.

This premise is not so plausible. According to the dual-process theory, consequentialist judgments are not emotionally driven (though affect may play a more indirect role in their genesis, Greene, 2007b) and are not in the relevant sense “intuitive.” With this in mind, one might revise the above premise as follows, dropping the words “emotional” and “intuitive,” on the grounds that they are unnecessary:

Anti-Consequentialist P1.1. The processing that gives rise to consequentialist judgments responds to factors that make a dilemma personal rather than impersonal.

But this is also not true. Consequentialists consciously make their judgments based on the body count (in trolley cases) and say to hell with everything else. If their intuitions tell them otherwise, they “bite the bullet” and say, “So much the worse for our intuitions.” (Of course not all consequentialists do this, but I have no interest in defending the ones who don’t. I have no problem, however, with consequentialists such as Hare (1981) who argue that simple forms of consequentialism can justify actions and policies that would appear to be non-consequentialist.)

The effect of personal force in the Trolley Problem

I will now flesh out the “Argument from Morally Irrelevant Factors” using the data that Berker dismisses. I will model my argument on those above concerning racial bias and incest.

In Greene et al. (2009, appended) my co-authors and I compared people’s responses to a number of trolley variations. More specifically, we compared four versions of the *footbridge* dilemma. In the *standard footbridge* case the agent pushes the victim off of the footbridge with his hands. In the *footbridge pole* version, the agent pushes the victim with a pole. In the *footbridge switch* version, the agent drops the victim onto the tracks through a switch-operated trap door, where the switch is next to the victim on the footbridge. In the *remote footbridge* version, the switch is located elsewhere, away from the footbridge. We asked separate groups of subjects to judge whether the action proposed is morally acceptable. The results are as follows:

standard footbridge: 31% Yes

<i>footbridge pole:</i>	33% Yes
<i>footbridge switch:</i>	59% Yes
<i>remote footbridge:</i>	63% Yes

The results for the first two cases do not differ significantly. Nor do the results of the last two cases. However, the difference between the first two cases and the second two is highly significant. As you can see, people are about twice as likely to approve of the action if it is performed by hitting a switch rather than by pushing someone. I call the factor that differentiates the first two cases from the second two cases “personal force,” which is distinct from both body contact and spatial distance. The concept of “personal force” may undergo further refinement (For example, does throwing projectiles count?), but for present purposes, we can think of personal force as being about pushing.

Now we’re ready to make an argument, modeled on those above. We’ll start with a scientifically supported descriptive premise.

D3. People’s judgments in response to trolley problems are strongly influenced by the presence/absence of personal force.

By itself, this descriptive premise gets us nowhere. To gain normative traction, we need an additional normative premise:

NP3. The presence/absence of personal force is morally irrelevant to the moral acceptability of actions such as these.

And thus we reach the following conclusion:

NC3. People’s judgments in response to trolley problems are strongly influenced by at least one morally irrelevant factor, personal force, and are therefore at least somewhat unreliable.

Let’s take stock. We have now dealt with Berker’s first two worries. The data show that personal force (or something very much like it) has an effect on people’s judgments, and therefore P1 is well-supported. (See above and below concerning whether it’s the *characteristically deontological* judgments that are sensitive to personal force.) We have also dealt with Berker’s second worry. The above experiment is very well controlled. More specifically, the *footbridge switch* and *footbridge pole* cases are very close. Perhaps something other than personal force differentiates them, but whatever that “something else” may be, it’s not going to be morally relevant. (One might suppose that the action is more likely to work in the *footbridge switch* case, but, as explained in my response to Berker’s methodological critique, this experiment *controlled for* such real-world expectations.)

Finally, this argument deals with part of Berker’s fourth worry. Once again, this is the worry about what is “doing work” in the argument. Berker levies two distinct charges. The first is about what *is* doing the work. Here, Berker says that it’s nothing but question-begging armchair assumptions. The second charge is about what is *not* doing any work. Here, Berker says that it’s not neuroscience. So far, neuroscience has

played not role, but that will change shortly. With respect to the first charge, Berker is incorrect. To make this argument work, we needed a non-scientific normative premise, NP3. But, as noted above, this premise is not *question-begging*. Rather, it is a premise that nearly all of us share. Whether your normative proclivities are consequentialist, deontological, or otherwise, it's hard for you to argue that personal force is morally relevant.

NC3 may not say everything an empirically-minded Trolleyologist might want to say, but it is a substantive normative conclusion. Moreover, the scientific evidence presented above played a critical role in reaching that conclusion and thus “does work,” despite the fact that the argument also requires NP3, a substantive normative premise. The fact that the above argument requires an additional normative premise does not mean that the argument is question-begging or that the science does “no work.”

We now have one and a half worries left. Worry 3 acknowledges that the argument, as it currently stands, is a “strike against” deontology, but it asks whether a similar argument might not make an equally powerful “strike against” consequentialism. The remaining half worry concerns the specific role of neuroscience. We've already shown that science matters. But what about *neuroscience*? The response to both of these worries brings us back to the dual-process theory. The dual-process theory explains why there is an asymmetry between deontology and consequentialism, and it makes neuroscience relevant because the dual-process theory is supported (though not essentially supported) by neuroscience.

Before moving on, let me head off a tempting counter-argument. Further experiments in Greene et al. (2009) show that people's judgments are also sensitive to the means/side-effect distinction. This may sound like good news to traditional trolleyologists and, in the absence of further information, it is to some extent. This means that people, in addition to being sensitive to a morally irrelevant factor, are also sensitive to a factor that is considered highly morally relevant.

For now, I will simply emphasize that this good news is not enough to *undo* the bad news above. The effect exerted by the mean/side-effect distinction is not an *alternative explanation* for the effects described above. That is, the means/side-effect distinction cannot explain why people think the actions in the *footbridge* and *footbridge pole* cases are worse than those in the *remote footbridge* and *footbridge switch* cases. Something is still going wrong with people's intuitive judgments.

Moreover, I would advise against taking refuge in the means/side-effect distinction. In my forthcoming book I argue that our tendency to draw the means/side-effect distinction is a by-product of the action-representation format used by the system that triggers our automatic emotional responses to harmful actions. That's a mouthful, and I won't unpack it here.

What's gone wrong?

The above argument tells us that something stinks in Trolleytown—and that’s important news—but it doesn’t identify the source. Why not suppose, as Berker does in his third worry, that consequentialist “intuitions” are as much to blame as deontological ones? The answer is that there is a deep cognitive asymmetry between consequentialist and deontological thinking, as posited by the dual-process theory. The fact that characteristically deontological judgments are driven by automatic emotional responses is not irrelevant to figuring out what’s gone wrong with our judgments in the Trolley Problem.

To definitively identify the source of the problem will require a detailed cognitive and evolutionary theory that explains how and why we make the judgments we make. Such a project goes well beyond the scope of this set of notes. I address this question further in my forthcoming book. For now I would like to offer a more general argument for the claim that the stink in Trolleytown is more of a problem for deontologists than consequentialists.

I’ll begin by drawing a distinction between what I’ll call *consequentialist thinking* and *consequentialism*. Consequentialist thinking is simply thinking aimed at producing the best possible overall consequences. This kind of thinking is perfectly ordinary and not particularly controversial. Everyone agrees that, *if all else is equal*, it’s better to produce better overall consequences. *Consequentialism* is the highly controversial idea that aggregate consequences are the *only* things that ultimately matter. In other words, everyone thinks that consequentialist thinking is reasonable and that it is often right to act on it. The debate is about whether there are *other* kinds of thinking that capture important features of morality that consequentialism misses.

So what’s wrong with *consequentialism*? Why do we need anything else? Philosophers have various theoretical arguments against consequentialism, but what makes those arguments compelling are the myriad intuitive counterexamples that philosophers have generated over the years, from Nozick’s (1974) utility monster, to Rawls’ (1971) utilitarian slave society, to the *footbridge* case (Thomson, 1985). In short, consequentialism’s problem is not that it doesn’t make sense in the abstract, but that it offends our moral sensibilities when it comes to certain specific cases. Because consequentialism’s biggest problem is its counter-intuitive implications, the news that our moral intuitions are at least somewhat unreliable is good news for consequentialists.

At this point one might ask whether this news is not equally good for deontologists. I don’t think it is. This is because counter-intuitive implications are not a nuisance for deontologists in the way they are for consequentialists. This is nicely illustrated by philosophical work on the Trolley Problem. Deontological trolleyologists are in search of a theory that offends our intuitions as little as possible. For example, the doctrine of double effect allows us to say “yes” to the *switch* case and “no” to the *footbridge* case, which is what most people want to say. But then there’s that pesky *loop* case (Thomson, 1985), in which the action appears to be morally acceptable, despite the fact that the victim is harmed as a means. And so, the deontologist presses on, looking

for ever more sophisticated principles that are consistent with ever more intuitive judgments in response to an ever growing library of cases Kamm (1993, 1996, 2006). The deontological trolleyologist is not slavishly bound to his intuitions. For example, such a philosopher might *feel* the difference between the *footbridge* and *footbridge switch* cases, but nevertheless reject that feeling as invalid, especially when the source of the feeling is revealed. However—and this is the critical point—a deontological trolleyologist cannot dismiss his intuitions as *generally* invalid because they are his guiding lights.

At this point, one might object that consequentialists need their intuitions just as much as deontologists. What about the consequentialist intuition that it's morally acceptable to hit the switch in the *switch* case? Isn't that intuition just as vulnerable as the intuition that it's wrong to push the man off the footbridge? I don't think so. This is where the dual-process theory comes in.

When philosophers talk about “intuitions,” they mean something like “natural, untutored judgments.” But when psychologists talk about “intuitions” they mean something even more specific. For psychologists, an intuition is, among other things, the output of a process to which the individual has no conscious access. For example, since childhood you've been able to reliably identify faces as male or female. But you have no idea—or at least had no idea when you were eight—what features of faces you used to identify faces as male or female (e.g. the height of the cheekbones, the distance between the nose and the lips). This is because you identify faces *intuitively* rather than through the conscious application of a male/female face-sorting rule.

Consequentialist judgments are based on “intuitions” in the philosophers sense, but they are not based on intuitions in the psychologist's sense. For example, the perfectly natural, untutored judgment that it's acceptable to turn the trolley in the *switch* case is *not* an intuitive judgment. It is, I claim, a very simple *reasoned* judgment, the result of explicitly applying a utilitarian decision rule. In contrast, the judgment that it's wrong to push the man off the footbridge is an intuitive judgment. It's based on an automatic emotional response. This is not an arbitrary conjecture. There is evidence for it.

1. Conscious access: When people make characteristically consequentialist judgments, they always know why they are judging as they do. Indeed, even when people make characteristically deontological judgments, they are explicitly aware of the consequentialist rationale for doing otherwise. In contrast, people can and do exhibit classic deontological patterns of judgment (e.g. conforming to the doctrine of double effect or the doctrine of doing and allowing) and yet have no idea why they are judging as they do (Cushman et al., 2006).
2. Lesion patients: If characteristically consequentialist judgments were just as intuitive as characteristically deontological judgments, then there would be no reason for patients with emotion-related brain damage to be more consequentialist (Mendez et al., 2005; Ciarmelli et al., 2007; Koenigs et al., 2007). Patients who can reason quite well, but who

lack emotional intuitions, become more consequentialist because consequentialist thinking is not intuitive in the psychological sense.

3. fMRI data: Perhaps the most direct evidence on this point comes from the results of a new fMRI study (as yet unpublished) in which we compared the neural activity associated with characteristically deontological disapproval (e.g. saying “no” to the *footbridge* case) and characteristically consequentialist disapproval (e.g. saying no to a *reversed* switch case in which one can turn the trolley away from one and onto five). To a philosopher, these two judgments are equally “intuitive,” but the brain imaging data reveal that, psychologically speaking, they are qualitatively different. Consequentialist disapproval was, as predicted, associated with increased activity in the dorsolateral prefrontal cortex (associated with reasoning), while deontological disapproval was associated with increased activity in the medial prefrontal cortex (associated with emotion and social cognition).

4. Priming counter-intuitive behavior: In a recent study Joe Paxton and I primed people to distrust their intuitions by giving them a test with tricky math problems (the Cognitive Reflection Test; Frederick, 2005). As predicted, taking the test before responding to moral dilemmas made people’s judgments more consequentialist. This shows that consequentialist judgments are not simply based on different intuitions, but are genuinely counter-intuitive (when the utilitarian action is emotionally aversive).

In short, characteristically consequentialist judgments are not intuitive in the psychological sense, but characteristically deontological judgments are. (This is not to say that emotion plays no role in consequentialist thinking. See Greene, 2007b.) More generally, our mysteriously variable moral intuitions are a nuisance for consequentialists, but they are—as I have argued elsewhere based on a wide range of evidence (Greene, 2007b)—the lifeblood of deontological theorizing. For these reasons, evidence that our intuitions are unreliable is a point in favor of consequentialism and a point against deontology. Consequentialism can do just fine without intuitions, but deontology is, in practice if not in theory, all about finding theories that cohere with our mysteriously variable intuitions. Or so I claim, as an empirical matter (Greene, 2007b).

Does neuroscience do any work in the above argument? Yes. Neuroscientific data support the dual-process theory, and the dual-process theory explains the asymmetry between consequentialist and deontological judgment. As noted above, the dual-process theory does not *require* evidence from neuroscience, and so it’s true that there is nothing *essentially* neuroscientific about the argument. Once again, it’s not about neuroscience.

The 2009 data show that our trolley judgments are at least somewhat unreliable, but they do not tell us what’s gone wrong. The dual-process theory tells us that the characteristically deontological judgments are the more intuitive ones, in the psychologist’s sense of “intuition.” But why should we think that intuitions, in the psychologist’s sense, are responsible for what’s gone wrong with our trolley judgments?

One answer, noted above, is that it cannot be our utilitarian reasoning that's giving rise to the personal force effect. Consequentialists know exactly what they are doing. In trolley cases, consequentialists care about the body count, and only the body count. Thus, it cannot be consequentialist thinking that is differentiating between cases that vary in the presence/absence of personal force, but that keep the body count constant. A deontologist will, of course, disavow any allegiance to the "principle of personal force." The point, however, is that deontologists often don't know what "principles" are underlying their judgments. One possibility—one that I favor—is that once all of the inner workings of our judgments are revealed by science, there will be nothing left for deontologists. All of the factors that push us away from consequentialism will, once brought into the light, turn out to be things that we will all regard as morally irrelevant. That's the grand ambition. The argument made here is just a first step. In my forthcoming book I make a few more.

The argument in the special case of the Trolley Problem has gone like this: Something is going wrong in our trolley judgments, and consequentialist reasoning does not appear to be the culprit. Therefore, it must be the intuitive deontological thinking. This argument works well enough, but it leaves open the possibility that deontological thinking could easily be cleaned up. Yes, deontological thinking is intuitive. And yes, intuitions can go wrong, as illustrated by the data presented above. But this doesn't mean that consequentialist thinking is right and the deontological thinking is wrong. It could be that consequentialist thinking (reasoned though it may be, in the psychological sense) is completely wrong, and that deontological thinking is generally right. It's just imperfectly implemented in our intuitions. With this in mind, I want to offer a more general argument for why we should distrust our deontological intuitions, at least in some important circumstances.

The Camera Analogy: Getting Beyond Point-and-Shoot Morality

The passage below is from the introduction to my forthcoming book. (Again, pardon the missing references.)

The moral brain is like a dual-mode camera with both automatic settings and a manual mode. To solve our biggest problems we need to better understand how our moral brains work. We need to learn when to rely on automatic settings and when to put our moral brains in manual mode.

...A dual-mode camera has automatic settings that are optimized for typical photographic situations ("portrait," "action," "landscape"). The user hits a single button, and the camera automatically configures the ISO, aperture, exposure, etc.—point and shoot. A dual-mode camera also has a manual mode that allows the user to adjust all of the camera's settings by hand. A camera with both automatic settings and a manual mode exemplifies an elegant solution to a ubiquitous design problem, namely the trade-off between *efficiency* and *flexibility*. The automatic settings are highly efficient, but not very flexible, and the reverse

is true of the manual mode. Put them together, however, and you get the best of both worlds, provided that you know when to manually adjust your settings and when to point and shoot.

The moral brain's automatic settings are dispositions to have intuitive emotional responses, gut reactions to actions, people, and other objects of moral evaluation. Suppose, for example, that you're walking along when you see a group of children pour gasoline on a stray dog and set it on fire. You have a strong and immediate sense that their behavior is wrong, a reaction that requires no conscious inference (Zajonc, 1980). That's an extreme case. Far more often, our automatic settings simply emit little emotional blips that tell us "that's good" or "that's bad." For example, you probably got a little blip of badness when you read about the cruel children above.

The moral brain's moral manual mode, in contrast, is our capacity for conscious, deliberate, moral reasoning. This includes our ability to apply explicit moral rules, to evaluate moral rules and judgments for consistency, and to override gut reactions that are at odds with our considered judgments. Suppose, for example, that your vegetarian friend takes your encounter with the dog-burning children as an opportunity to challenge your moral views. She says that the suffering experienced by factory-farmed pigs over the course of their lives is comparable to that experienced by a burning dog. The enjoyment some children derive from dog-burning is comparable to the enjoyment that you and your carnivorous friends derive from pulled pork sandwiches, she says. Thus, your meat-eating behavior is morally comparable to that of the cruel children you've so emphatically condemned. Vigorous debate ensues. You argue that your behavior is different from theirs in many significant ways. Your friend argues that those differences are either illusory or morally irrelevant. And so it goes, back and forth. This is morality in manual mode, though probably not exclusively in manual mode. Unlike a camera's automatic settings, ours are never turned off. We can, however, override our automatic settings. For example, you might choose to stop eating meat, despite its enduring appeal, simply because you've been convinced by your friend's argument.

The camera metaphor connects the two principal questions addressed by this book: *How does the moral brain work?* and *How can it work better?*

[At this point I introduce the Nick and Danielle incest case as above.]

This case nicely illustrates this book's central message, which is that *we rely too much on our automatic settings*, on our emotional moral intuitions. Our emotional aversion to incest is the product of an automatic setting, one that is shaped by both genes and experience[refs]. Like all automatic settings, this one is efficient, but inflexible. The aversion to incest is an efficient mechanism for avoiding birth defects, but this aversion persists even when the rationale for having it no longer applies.

Generalizing from this example, it would be foolish to condemn all of our automatic settings as stupid and irrational. But it would be equally foolish to assume that our automatic settings are always correct. Instead, we should try to

understand our automatic settings—where they come from and how they work—and then apply that knowledge in deciding when to trust our intuitions and when to override them.

Here, too, the camera analogy offers some guidance. A camera's automatic settings are designed to handle familiar photographic problems, ones that the camera's manufacturer could anticipate. The designs of these automatic settings reflect the collective experience of past photographers, who have learned through trial and error which configurations are best for taking portraits, action photos, landscapes, etc. If you're facing a familiar photographic problem, the automatic settings are likely to serve you well. In contrast, if you're trying to do something fundamentally new, pushing the bounds of photography, it's unlikely—not impossible, but unlikely—that you'll accomplish your goals using point-and-shoot automatic settings. Instead, you'll need to put the camera in manual mode. In the same way, we can expect our automatic settings to serve us well when our moral problems are familiar and to lead us astray when our moral problems are fundamentally new. This idea, however, comes with two important caveats concerning the meanings of “familiar” and “well.” [We won't get to the second one.]

First, familiarity: All automatic settings, whether in a camera or a human brain, ultimately rely on knowledge gained from trial-and-error experience. What makes a problem “familiar,” then, is that our automatic reactions to it have been shaped by trial-and-error experience. One's intuitions may be shaped by one's own personal experiences, the experiences of people from whom we've learned, or the experiences of our biological ancestors reflected in our genes. The problem of global warming is “familiar” in the sense that we've all heard plenty about it, but it is not familiar in the relevant sense. No one has had trial-and-error experience with solving the global warming problem because we are still in the midst of our first trial. At best, this problem is familiar in virtue of its similarity to other problems that people have faced. But if the global warming problem has critical features that are fundamentally new, then, from a cognitive perspective, it's not familiar, and we are unlikely to solve it by relying on our automatic settings. More generally, it would be a kind of *cognitive miracle* if our automatic settings could adequately deal with complex new problems. [Note: And this is true *regardless of the standard we use for determining what counts as dealing with our problems “well.”* This point is critical because it explains how psychology can offer us guidance without making any assumptions about our ultimate normative standards.] This would liken our brains to a miraculous point-and-shoot camera that can, with the touch of a single button, do anything an advanced photographer might want to do. This point is important because it connects the “is” of science to the “ought” of morality, now in a more general way: If a moral problem *is* fundamentally new, then we *ought* not rely on our automatic settings in trying to solve it.

If the dual-process theory is correct, then deontology is fundamentally an intuitive philosophy. This may sound strange given how much time deontological philosophers spend engaged in intense, conscious moral reasoning. But, if I'm right, for deontologists,

the intuitions are running the show, keeping the reasoning on a very short leash. (Not a maximally short leash, but a short one.) The goal—sometimes explicit, sometimes implicit—of deontological philosophizing is to articulate a theory that conforms as closely as possible to our intuitive judgments. I believe that this is true even among philosophers, such as Kant and Rawls, who are attempting to build systematic deontological theories from the ground up based on first principles.

And what's wrong with doing things intuitively? Sometimes nothing at all. Automatic settings are great for handling "familiar" problems. But I believe that many of our biggest problems are "unfamiliar" and that deontological thinking gets in the way of solving them. If I'm right, deontological philosophy is essentially an elaborate set of rationalizations for "point-and-shoot" morality. Nietzsche understood this:

Kant's Joke—Kant wanted to prove, in a way that would dumbfound the common man, that the common man was right: that was the secret joke of this soul. He wrote against the scholars in support of popular prejudice, but for scholars and not for the people.

Consequentialism, in contrast, is *geek morality*. It's what you get when you turn the problem of moral thinking over to the brain's manual mode. Manual mode can be dangerous. It's good that we don't ordinarily have to think about whether to keep our promises, take care of our children, or push people off of footbridges (Hare, 1981). Our automatic settings give us the right answers most of the time. But sometimes, I contend, we need to put the automatic setting aside and use manual mode, even when this feels like the wrong thing to do.

To take my favorite example (Peter Singer's, too), I think that deontological thinking is a major obstacle to ending poverty (Greene, 2003). Nature didn't design us to behave morally in a world in which one can save the life of a distant stranger at very little cost. If Singer (2005) and I are correct, we humans rely too much on our emotional intuitions to tell us when we have a moral obligation to help and when we don't. *We feel* an obligation to help when a victim of misfortune is right in front of us (drowning baby), but not when the victim is a distant, "statistical" one (Small & Loewenstein, 2003; Slovic, 2007) on the other side of the world (giving to Oxfam). (My undergraduate thesis student, Jay Musen, has recently shown that people's judgments concerning obligations to help people in need are indeed affected by mere spatial distance.) If everyone in the world were to think about poverty as consequentialists do, instead of relying on their emotions to tell them when to help or not help, poverty might be quickly eliminated.

I am, of course, making many assumptions here and pushing many important questions off to the side. (Doesn't it beg the question to defend consequentialism by appealing to the fact that it endorses actions that produce good consequences?) My goal here, however, is not to make this larger argument. That's what I do in my book, and even then only incompletely. In this closing section, my aim is simply to sketch the picture of human morality that is in the background. If I'm right, deontological thinking is essential for ordinary life and an obstacle to solving many of our biggest problems, which are uniquely modern and "unfamiliar." To solve these problems we need to get beyond point-and-shoot morality and the philosophical temptation to defend it with reason.

Appendix: Further details concerning errors in and misleading features of Berker's methodological critique

1. 2004 fMRI results

In the main text of his critique, Berker raises three specific methodological objections, the first of which concerns the data presented in my 2004 fMRI paper.

The main finding of this paper is that utilitarian moral judgment is associated with increased activity in the dorsolateral prefrontal cortex, the brain region most closely associated with cognitive control. This result was specifically predicted by the dual-process theory, connecting a large literature on the neuroscience of cognitive control (Miller & Cohen, 2001) with research on moral judgment. This key result is never questioned by Berker. Instead he focuses on a different result that is more difficult to interpret.

A brain region in the posterior cingulate cortex, a region associated with emotion, but also implicated in a wide range of cognitive tasks, exhibits the same effect as the dorsolateral prefrontal cortex in this study. Berker writes that this finding “muddies the waters a bit,” a vague statement with which one could hardly disagree. But Berker presents this result to his readers as if it's a kind of dirty secret.

In our paper we reported 43 distinct neuroimaging findings, many of which are difficult to interpret. This is standard for neuroimaging papers, which are expected to present *all* statistically significant results while focusing on a select few that are interpretable based on *a priori* (here meaning prior to the experiment) hypotheses. Nearly all neuroimaging papers present results that are difficult to interpret and that could, with varying degrees of plausibility, be interpreted as running counter to the main conclusions drawn.

In keeping with this standard practice, we presented the full set of data in our paper and argued that, overall, the results support our conclusions. We argued that the most readily interpretable findings, findings that were predicted in advance based on our prior research, support the claim that utilitarian judgments are preferentially supported by controlled cognitive processes. The editors at *Neuron* (a top neuroscience journal) and three expert reviewers agreed that our conclusions were well supported, while fully aware of the result that Berker highlights. It is, of course, possible that the experts who evaluated this paper failed to appreciate the significance of this result, but Berker offers no reason to believe that this is the case. Berker brings to the table no new information, no additional expertise, and no new hypotheses or interpretations. For Berker to raise doubts about the paper's overall conclusions based on one ambiguous finding, without any discussion of the broader context, is extremely misleading, especially for readers who are not familiar with the standards and practices in neuroimaging research.

Berker also fails to mention that our 2008 *Cognition* paper provides additional support for the conclusions drawn in our 2004 fMRI paper. As explained below, Berker mentions the 2008 paper only in footnotes, where he falsely claims that the methods used in this paper are “statistically invalid.” (See above.)

2. Further details concerning statistical methods employed in Greene et al. 2001 and 2008.

In his second main objection, Berker claims that the methods used to analyze the reaction time data in our 2001 paper are “statistically invalid.” This is false. Far from being “invalid,” good statistical practice *required* us to perform a test of the kind that we performed.

We analyzed these data using a statistical test known as a mixed-effects or repeated measures ANOVA. More specifically, we modeled subject (the variable corresponding to the individual subject being tested) as a random variable and condition (personal vs. impersonal vs. non-moral dilemmas) as well as response type (“appropriate” vs. “inappropriate”) as fixed variables. This is a standard analysis for data of this type and is by no means “statistically invalid” (Howell, 2001, Chapter 14). By treating subject as a random variable, one minimizes the influence of irrelevant variability between subjects. Had we not performed this analysis, our results would be open to the objection that they simply reflect irrelevant variability between subjects and that they do not generalize beyond our sample.

The benefits of treating subject as a random effect come at a cost, allowing for the influence of irrelevant variability between items (in this case dilemmas). An alternative method, which minimizes the influence of irrelevant variability between items, but at the cost of allowing irrelevant variability between subjects to have an influence, is known as an “item analysis.” (NB: This term has other meanings in other contexts.) This involves treating the item variable, rather than the subject variable, as a random effect. Neither analysis is “invalid.” They have different strengths and weaknesses. Unfortunately there is no single test that can treat both item and subject as random effects (provided that subjects only respond to each item once). It is a general requirement that one perform an analysis in which subject is treated as a random effect. The ideal approach is to perform such an analysis and then also perform an item analysis whenever possible. Had we been more savvy statisticians we would have done this, but the method that we did use is by no means “statistically invalid.” Once again, it was required.

As noted above, our 2001 reaction time data tell a different story when analyzed using an item analysis, and for the reasons Berker suggests. I became aware of this problem soon after the 2001 paper was published. (See also my recent exchange with McGuire et al., 2009; Greene, 2009a). My awareness of this problem was one of the motivations for conducting the aforementioned cognitive load study, which was published in *Cognition* in 2008. In the 2008 paper we were more savvy statisticians. We excluded the items that gave rise to the problem in the 2001 paper and performed an item analysis in addition to our ANOVA in which subject is modeled as a random effect. As noted above, Berker fails to understand that the 2008 paper effectively deals with the problem faced by the 2001 paper. He mentions the 2008 paper only in the footnotes, where he mistakenly claims that the analysis is “statistically invalid” and fails to understand that we performed the kind of analysis that he favors.

Thus, buried in footnote 37 is a perfectly adequate, empirical response to Berker’s second objection. But Berker’s readers would have no way of knowing this. Berker closes this section of his paper as follows: “it is important to keep in mind that at this point in time the response-time prediction has not been borne out, which in fact is an

empirical strike against the dual-process hypothesis.” This statement, if it is not simply false, is highly misleading. The reaction time data from the 2008 paper support the dual-process theory.

3. Criticism of the criteria we used to sort our dilemmas into the categories of “personal” and “impersonal.”

In our 2001 paper, we divided moral dilemmas into two categories, “personal” (resembling the *footbridge* case) and “impersonal” (resembling the *switch* case). We made our division based on a set of criteria that were explicitly presented as tentative. We did this because our aim in this paper—contrary to the assumptions of many philosophers who read it—was not to solve the Trolley Problem, not even in descriptive terms. Rather, our aim was to test a psychological and neuroscientific theory—the dual-process theory—that was *inspired by* the Trolley Problem. According to this theory, both emotional processes and controlled cognitive processes play important roles in moral judgment—an interesting and controversial claim. Our aim was to provide evidence for this claim about emotion and cognition in moral judgment, and we did this by dividing our dilemmas into the two aforementioned categories and showing that one set elicited a more emotional pattern, while the other elicited a more classically “cognitive” pattern. This was enough to make our psychological and neuroscientific points, and it was not essential for our purposes that our sorting criteria solve the Trolley Problem, or even come close. Thus, while the personal/impersonal distinction drawn in the 2001 paper was in need of improvement for the purposes of solving the Trolley Problem, it was adequate for the purposes to which it was put in the 2001 paper.

I would like to emphasize that the personal/impersonal distinction tentatively posited in 2001 and the dual-process theory of moral judgment are completely orthogonal ideas (Greene, 2009a). The dual-process theory could be completely correct, even if the personal/impersonal distinction is completely wrong, and vice versa. The dual-process theory claims that dilemmas like the *footbridge* case elicit a prepotent negative emotional response to the action that supports the characteristically deontological response (It’s wrong, even though it will save more lives) as well as a competing controlled cognitive response that supports the characteristically utilitarian conclusion (It’s morally acceptable because it will save more lives). The personal/impersonal distinction constitutes a specific theory about what “dilemmas like the *footbridge* case” should mean. That is, it’s a theory that aims to identify the features of the *footbridge* case that make it more emotionally salient than other cases like the standard *switch* case. The personal/impersonal distinction is a modular piece that fits into dual-process framework. And if that module is bad, it can be replaced by another one. Indeed, my colleagues and I, in a more recent paper (*Cognition*, 2009), have attempted to replace that temporary 2001 module with something more permanent.

In his third main methodological point, Berker criticizes the criteria we used to draw the 2001 personal/impersonal distinction. I am happy to accept much of what Berker says about the shortcomings of the original distinction. My main concern with this part of Berker’s critique is the manner in which he dismisses the most relevant recent research, which was conducted with the explicit intention of improving upon the research that Berker here critiques (Greene et al, 2009, appended). This painstaking research—

which used low-tech, paper-and-pencil methods rather than brain imaging—took many years, involved dozens of dilemmas, employed over a thousand subjects, and produced very clear results. Instead of engaging with the new and maximally relevant research, Berker relegates his discussion of it to the footnotes and hastily dismisses it.

As noted above, Berker dismisses the entire 2009 paper based on a misunderstanding of the methods. Earlier in the same footnote (73), Berker writes:

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?).

The above criticisms sounds rather damning (“obvious potential counterexamples”), but the reader is not given anywhere near enough information to properly evaluate them. Allow me to provide some background.

In one of our critical comparisons, we compare two closely matched versions of the footbridge dilemma. In one case, the agent pushes the victim off the footbridge with a pole. In the contrasting case, the agent drops the victim through a trap door by hitting a switch, while standing near the victim on the footbridge. We find that people, on average, judge the action in the first case to be significantly less acceptable than the action in the second case. This is the critical result for our identification of “personal force” (roughly, pushing vs. hitting a switch) as a critical factor. The influence of this factor was then observed, not just in this pair of cases, but in several other pairs of closely matched dilemmas.

Berker’s comments above do not challenge these results in any way. Our claim is that the presence of “personal force” exerts a psychological effect in the moral dilemmas that we tested, which include versions of, and variations on, many classic trolley dilemmas. We did not claim that personal force (and intention) are the only features of harmful actions that could ever lead people to disapprove of them. And, thus, it is no “obvious counterexample” to say that people would also disapprove of an agent if he were to cause harm using threats of physical violence rather than physical violence. The important point, however, is that Berker does not give readers the opportunity to properly evaluate the evidence. He simply declares that the experiments are poorly done, giving enough information to make the charges sound credible, but nowhere near enough information for readers to judge for themselves.

In footnote 46, Berker cites “further evidence for the inadequacy of Greene et al.’s... criteria” by referencing an objection raised by Kahane and Shackel (2008). Kahane and Shackel published a letter in response to a study by Koenigs et al., which was published in *Nature* in 2007. This paper showed that patients with emotion-related brain damage are far more likely than others to make utilitarian judgments in response to dilemmas like the *footbridge* case. This study provides what is, in my estimation, the single strongest set of evidence supporting the dual-process theory. This study (on which I was not an author) used modified versions of the dilemmas that I introduced in my 2001

and 2004 brain imaging papers. Kahane and Shackle question whether certain dilemmas have a truly “utilitarian” option. Kahane and Shackle presented these dilemmas to philosophers and found that only five of them were deemed by a majority to involve a choice between a deontological and consequentialist/utilitarian option. (Kahane and Shackle evaluated 41 dilemmas in this way, but only a subset of them were ever intended to involve a utilitarian option.) In my opinion, Kahane and Shackle’s criteria are unnecessarily stringent. In short, they assume that a judgment favoring the maximization of utility isn’t utilitarian unless there is no other justification for that action. But, for present purposes, I’m happy to put this disagreement aside.

The key point here is that Berker mentions Kahane and Shackle’s objection, but then says nothing about the fact that Koenigs et al. responded to that objection with a further analysis of the data. Koenigs et al. took the five dilemmas that were blessed by Kahane and Shackle and separately reanalyzed the data from those dilemmas. They found, as predicted, that the patients with emotion-related damage gave significantly more utilitarian judgments in response to those five dilemmas. The Koenigs et al. reply was published alongside Kahane and Shackle’s letter, so there was no way to miss it. Nevertheless, Berker leaves his readers with the impression that the conclusions of the Koenigs et al study are now in doubt, when the opposite is true. Kahane and Shackle proposed a more stringent test for the dual-process theory, and the dual-process theory passed with flying colors.

What’s more, Berker fails to cite a study by Mendez et al. (2005) which makes the same point as the Koenigs study while completely avoiding the Kahane and Shackle’s objection. In the Mendez et al. study, the patients with emotion related deficits were only presented with versions of the original *switch* and *footbridge* dilemmas, and therefore there is no concern with the selection of dilemmas. In the Mendez study that Berker fails to cite, the patients with emotion-related brain damage were three times as likely as others to say that it’s morally acceptable to push the man off the footbridge—another confirmed prediction of the dual-process theory.

4. Miscellaneous criticisms

At the end of his methodological critique, Berker adds a footnote raising three additional objections to studies that support the dual-process theory.

First, in our 2004 fMRI paper we find that difficult moral dilemmas are associated with increased activity in the anterior cingulate cortex, a brain region associated with response conflict. We interpret this result as evidence for response conflict, an interpretation consistent with the dual-process theory. We support this interpretation with multiple references to the literature on anterior cingulate function (Botvinick et al., 1999, 2001; MacDonald et al., 2000). In our discussion we note that other interpretations of anterior cingulate activity are possible. As before, Berker seizes upon this interpretative ambiguity, which we openly acknowledged, and presents it as yet another dirty secret. Once again, and as above, Berker provides no new information, no additional expertise, and no new hypotheses. Based on his objection, there is no reason to think that Berker knows anything about the function of the anterior cingulate cortex beyond what he’s read in our 2004 paper. His objection is entirely gratuitous.

Second, Berker questions, in a different way, the results of the aforementioned 2008 cognitive load study. Our main finding, once again, is that the cognitive load selectively increased reaction time for utilitarian judgments. Our study did not find that the cognitive load reduced the frequency of utilitarian judgment, a result that would have been even stronger than the result we reported. Berker describes this non-finding as a “troubling piece of counterevidence.” It is not “counterevidence.” The reaction time results that we reported are highly significant and theoretically meaningful. (That is why these data were published in a top cognitive psychology journal.) The fact that our experiment did not produce an even stronger result does not change the fact that it produced the very interesting result that it produced. This result requires an explanation, and right now the dual-process theory is the theory that best explains it.

Third, Berker questions the results of the Koenigs et al. study, following Moll and Oliveira-Souza (2007), who observe that (some of) the patients tested have damage to the anterior dorsolateral prefrontal cortex, and therefore may have “cognitive” deficits as well as emotional deficits. (See my reply in Greene, 2007a.) Here, too, Berker fails to consider the available evidence. The patients in the Koenigs et al. study were independently evaluated for their cognitive and emotional capacities. These patients were given five different tests of cognitive ability and four different tests of social-emotional competence. The patients did very well on tests of cognitive ability. For example, these individuals, despite their brain severe damage, had a mean IQ of 106--above average. In contrast, all six patients exhibited impaired physiological responses to emotional stimuli. Likewise, all six patients exhibited moderate to severe impairments in empathy. In sum, the claim that these patients have emotional deficits, but not “cognitive” deficits, is backed up by a lot of evidence, evidence that Berker completely ignores.

Berker also fails to consider the two other published studies that tested patients with emotion-related brain damage. First, Ciaramelli et al. (2007) performed essentially the same experiment as Koenigs et al., but with a different set of patients, and got the same results. None of the patients tested by Ciaramelli et al. have damage to the dorsolateral prefrontal cortex. Second, as noted above, Mendez and colleagues (2005) showed that patients with frontotemporal dementia (a disease that compromises emotional function) exhibit the same effect, making more utilitarian judgments than others in response to the footbridge case. These patients, too, have no observable damage to the dorsolateral prefrontal cortex.

Berker fails to mention these two studies in the footnote in which he dismisses the Koenigs et al. study. More important, however, is the fact that Berker fails to mention these papers anywhere in his paper. This oversight is especially egregious because these two studies together counter the objections (again, not well-supported) that Berker makes to the Koenigs et al. study. As noted above, the Mendez et al. (2005) paper uses only versions of the classic switch and footbridge cases, and is therefore immune to the objection that the relevant affirmative judgments are not utilitarian. The Ciaramelli (2007) paper uses patients who have no damage to the dorsolateral prefrontal cortex, and is therefore immune to the objection that their judgments might be principally shaped by “cognitive” deficits, resulting from damage to the dorsolateral prefrontal cortex. This convergent evidence is a model of scientific progress—different studies with different strengths and weaknesses conducted by different labs coming together to support a single

theory. All three papers were published long before Berker's paper was published and are referenced in papers that Berker references. For Berker to ignore two of these three studies (while relegating the third to the footnotes) is shoddy scholarship.

References

- Baldus D, Woodworth G, Zuckerman D, Weiner N, Broffitt B (1998) Racial discrimination and the death penalty in the post-Furman era: An empirical and legal overview, with recent findings from Philadelphia. *Cornell Law Review* 83:1638-1770.
- Bartels D (2008) Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition* 108:381-417.
- Berker S (2009) The normative insignificance of neuroscience. *Philosophy and Public Affairs* 37:293-329.
- Bloom P (2006) Seduced by the flickering lights of the brain. In: Seed.
- Botvinick M, Nystrom LE, Fissell K, Carter CS, Cohen JD (1999) Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature* 402:179-181.
- Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD (2001) Conflict monitoring and cognitive control. *Psychol Rev* 108:624-652.
- Ciaramelli E, Muccioli M, Ladavas E, di Pellegrino G (2007) Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience* 2:84-92.
- Crockett MJ, Clark L, Hauser MD, Robbins TW (2010) Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proc Natl Acad Sci U S A* 107:17433-17438.
- Cushman F, Young L, Hauser M (2006) The role of conscious reasoning and intuition in moral judgment: testing three principles of harm. *Psychol Sci* 17:1082-1089.
- Eberhardt JL, Davies PG, Purdie-Vaughns VJ, Johnson SL (2006) Looking deathworthy: perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychol Sci* 17:383-386.
- Frederick S (2005) Cognitive Reflection and Decision Making. *Journal of Economic Perspectives* 19:25-42.
- Greene J (2003) From neural 'is' to moral 'ought': what are the moral implications of neuroscientific moral psychology? *Nat Rev Neurosci* 4:846-849.
- Greene J (in press) Social neuroscience and the soul's last stand. In: *Social Neuroscience: Toward Understanding the Underpinnings of the Social Mind* (Todorov A, Fiske S, Prentice D, eds). New York: Oxford University Press.
- Greene JD (2007a) Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends Cogn Sci* 11:322-323; author reply 323-324.
- Greene JD (2007b) The Secret Joke of Kant's Soul. In: *Moral Psychology, Vol. 3: The Neuroscience of Morality: Emotion, Disease, and Development* (Sinnott-Armstrong W, ed). Cambridge, MA: MIT Press.
- Greene JD (2009a) Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie. *Journal of Experimental Social Psychology* 45:581-584.
- Greene JD (2009b) The cognitive neuroscience of moral judgment. In: *The Cognitive Neurosciences IV* (Gazzaniga MS, ed). Cambridge, MA: MIT.

- Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD (2001) An fMRI investigation of emotional engagement in moral judgment. *Science* 293:2105-2108.
- Greene JD, Nystrom LE, Engell AD, Darley JM, Cohen JD (2004) The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44:389-400.
- Greene JD, Morelli SA, Lowenberg K, Nystrom LE, Cohen JD (2008) Cognitive load selectively interferes with utilitarian moral judgment. *Cognition* 107:1144-1154.
- Greene JD, Cushman FA, Stewart LE, Lowenberg K, Nystrom LE, Cohen JD (2009) Pushing moral buttons: the interaction between personal force and intention in moral judgment. *Cognition* 111:364-371.
- Haidt J (2001) The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 108:814-834.
- Hare RM (1981) *Moral thinking: Its levels, method, and point*. Oxford, England: Oxford University Press.
- Howell D (2001) *Statistical methods for psychology*: Wadsworth.
- Kahane G, Shackel N (2008) Do abnormal responses show utilitarian bias? *Nature* 452:E5; author reply E5-6.
- Kamm F (2006) *Intricate Ethics*. New York: Oxford University Press.
- Kamm FM (1993) *Morality, Mortality Vol. 1: Death and Whom to Save from It*. New York: Oxford University Press.
- Kamm FM (1996) *Morality, Mortality Vol. II: Rights, Duties, and Status*. New York: Oxford University Press.
- Koenigs M, Young L, Adolphs R, Tranel D, Cushman F, Hauser M, Damasio A (2007) Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature* 446:908-911.
- MacDonald AW, 3rd, Cohen JD, Stenger VA, Carter CS (2000) Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science* 288:1835-1838.
- Mcguire J, Langdon R, Coltheart M, Mackenzie C (2009) A reanalysis of the personal/impersonal distinction in moral psychology research *Journal of Experimental Social Psychology* 45:577-580.
- Mendez MF, Anderson E, Shapira JS (2005) An investigation of moral judgement in frontotemporal dementia. *Cogn Behav Neurol* 18:193-197.
- Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167-202.
- Moll J, Oliveira-Souza (2007) Moral judgments, emotions, and the utilitarian brain. *Trends Cogn Sci*.
- Moore A, Clark B, Kane M (2008) Who shalt not kill?: Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science* 19:549-557.
- Moore GE (1903/1966) *Principia Ethica*. Cambridge: Cambridge University Press.
- Moretto G, Ladavas E, Mattioli F, di Pellegrino G A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *J Cogn Neurosci* 22:1888-1899.
- Nozick R (1974) *Anarchy, State, and Utopia*. New York: Basic Books.

- Paxton J, Ungar L, Greene JD (submitted) Reflection and Reasoning in Moral Judgment. In.
- Rawls J (1971) A theory of justice. Cambridge, MA: Harvard University Press.
- Singer, P (2005) Ethics and intuitions. *The Journal of Ethics* 9, 331–52
- Slovic P (2007) If I look at the mass I will never act: Psychic numbing and genocide. *Judgment and Decision Making* 2:79-95.
- Small DA, Loewenstein G (2003) Helping a Victim or Helping the Victim. *Journal of Risk and Uncertainty* 26:5-16.
- Thomson J (1985) The Trolley Problem. *Yale Law Journal* 94:1395-1415.
- Valdesolo P, DeSteno D (2006) Manipulations of emotional context shape moral judgment. *Psychol Sci* 17:476-477.
- Zajonc RB (1980) Feeling and thinking: Preferences need no inferences. *American Psychologist* 35:151-175.