On the normative significance of experimental moral psychology

Victor Kumar & Richmond Campbell

Available online: 01 Mar 2012

PLEASE SCROLL DOWN FOR ARTICLE

Routledge
Taylor & Francis Group

On the normative significance of experimental moral psychology[†]

Victor Kumar and Richmond Campbell

Experimental research in moral psychology can be used to generate debunking arguments in ethics. Specifically, research can indicate that we draw a moral distinction on the basis of a morally irrelevant difference. We develop this naturalistic approach by examining a recent debate between Joshua Greene and Selim Berker. We argue that Greene's research, if accurate, undermines attempts to reconcile opposing judgments about trolley cases, but that his attempt to debunk deontology fails. We then draw some general lessons about the possibility of empirical debunking arguments in ethics.

Keywords: Consistency Reasoning; Debunking; Joshua Greene; Moral Psychology

1. Introduction

Our purpose in this essay is to illuminate the normative significance of experimental moral psychology. To do so, we examine Joshua Greene's (2003, 2008, 2010) recent experimentally based arguments that there is no rational basis for our deontological moral commitments. In general, empirical debunking arguments in ethics are successful insofar as they can be leveraged upon normative claims that are more plausible than the normative claims they target. Though we share with Greene a naturalistic approach to ethics (Campbell & Kumar, forthcoming), we argue that his attempt to empirically debunk deontology fails because it turns on implausible and undefended normative assumptions, or else confuses them with normative assumptions that are plausible but render his argument invalid. Our discussion, however, reveals how empirically based debunking arguments in ethics can be sound provided they incorporate plausible normative assumptions, and therefore can be a rational basis for moral change.

In sections 2 and 3 we examine Greene's main debunking argument and a leading critique by Selim Berker (2009). Berker argues that Greene's research is—and could

Victor Kumar is a PhD candidate at the University of Arizona.
Richmond Campbell is George Munro Professor of Philosophy Emeritus at Dalhousie University.
Correspondence to: Victor Kumar, Philosophy Department, 213 Social Sciences Bldg, P.O. Box 210027, Tucson, AZ 85721-0027, USA. Email: vckumar@email.arizona.edu

not be other than—irrelevant to the assessment of normative ethical theories.[1] In sections 4 and 5 we show that Greene's experimental findings, once combined with defensible normative premises, are normatively significant but do not threaten deontology. By casting a critical eye on Greene's arguments, we uncover a more defensible form of debunking argument supported by research in experimental moral psychology. In section 6 we outline a schema for empirical debunking arguments and identify possible applications to debates in ethics.

2. Greene's Main Argument

Greene's central empirical finding relates to two well-known variations on the trolley case (Greene, Nystrom, Engell, Darley, & Cohen, 2004). In the *Footbridge* case, one is asked whether it is permissible to push a large person from a footbridge and into the path of an oncoming trolley in order to save five people who would otherwise be run over. Most of us judge that pushing the large person is morally impermissible. Because the choice does not maximize the good, this is for Greene a characteristically deontological judgment. In the *Bystander* case, one is asked whether it is permissible to flip a switch that will divert a trolley from one track to another, running over one person instead of five. Most judge that flipping the switch is morally permissible. One person is sacrificed for the sake of many, and so this is for Greene a characteristically consequentialist judgment.[2]

The philosophical literature on trolley cases is devoted largely to defending a principle that justifies these seemingly conflicting judgments (Foot, 1967; Kamm, 1996, 2007; Thomson, 1976, 1985). For some philosophers, trolley cases and other cases like them show that it makes a difference how harm is brought about. According to the doctrine of double effect, pushing the large person in *Footbridge* is impermissible because the harm is intended; flipping the switch in *Bystander* is permissible because the harm is foreseen but unintended.

Greene's primary aim is to develop a psychological explanation of judgments about trolley cases by probing subjects' responses to *Footbridge*, *Bystander* and a number of other cases that tease apart potentially confounding variables. Surprisingly, Greene finds that variation among subjects' responses is explained by the presence or absence of "personal" harm. Harm is defined as "personal," roughly, if it involves an agent exerting muscular force against an identifiable other (Greene et al., 2009). So, for example, in Greene's cases pushing someone with your hands or with a pole from a footbridge constitutes personal harm. But remotely flipping a switch that opens a trapdoor beneath his feet constitutes impersonal harm. Greene finds that subjects tend to make a deontological judgment about *Footbridge* because the harm is personal; they make a consequentialist judgment about *Bystander* because the harm is impersonal.[3]

Greene thinks this psychological explanation of responses to trolley cases debunks the deontological judgment about *Footbridge* and therefore has normative implications. Without directly attacking its truth, Greene claims that the judgment is based

on poor grounds.[4] His main argument's clearest formulations are in Berker's (2009) critical piece and Greene's (2010) reply, both of which make explicit a normative assumption in the argument. Berker's label is the "argument from morally irrelevant factors":

> (1) The deontological judgment about *Footbridge* is a response to personal harm.
> (2) Personal harm is a morally irrelevant factor.

Therefore,

> (3) The deontological judgment about *Footbridge* is unwarranted.[5]

(1) is an empirical claim, an upshot of Greene's research. According to Greene, (2) is an uncontroversial normative premise: whether harm is personal rather than impersonal cannot alter the moral permissibility of an action. Greene's premises together yield an apparently valid debunking argument against the judgment that one should not push the large person in *Footbridge*. If our deontological judgment about *Footbridge* is based merely on personal harm, Greene argues, then it is poorly grounded and we should give it up.

Greene argues next that the debunking of intuitions about *Footbridge* undermines deontological moral theory. *Footbridge* is generally thought to support deontology over consequentialism, because the judgment that it is wrong to push the large person off the footbridge is consistent with deontology and inconsistent with consequentialism. Greene argues, however, that because our intuitions about *Footbridge* are ill grounded, as his research shows, a piece of evidence in support of deontology is undercut. Of course, many other considerations are relevant to the evaluation of deontology and consequentialism. Nevertheless, if Greene's reasoning is sound, he has produced an empirical reason to favor consequentialism over deontology.[6]

Before shifting to critical assessment of Greene, we want to take note of a crucial, hidden assumption that he does not identify or defend. The conclusion Greene draws in the argument from morally irrelevant factors supports an argument against deontological moral theory only with the aid of a *moral-epistemological assumption*. Roughly, the assumption is that the principal evidence for moral theories is our first order intuitions about concrete cases. One moral theory is more justified than another principally insofar as it better explains and systematizes our first order intuitions. Thus, Greene may argue, debunking intuitions about *Footbridge* removes a principal source of support for deontology over consequentialism. Without assuming that moral theories are justified principally by intuitions about cases, Greene's argument would be incomplete. For it would be open to his opponents to claim that deontology is justified principally on other grounds, besides the intuitions it captures, e.g., a priori justification for deontological principles. In our view the moral-epistemological assumption is defensible, and we won't challenge it here. It is needed, now, to complete Greene's argument and, later, to uncover the actual normative implications of his and others' experimental research.[7]

3. Berker's Objection

For now, we are interested in scrutinizing the alleged normative implications of
Greene's research. So, we will for the sake of argument accept empirical premise
(1) about the influence of personal harm on responses to *Footbridge*. Berker (2009,
pp. 305–313) challenges the validity of Greene's research, and Greene (2010) has
offered a careful and lengthy reply, but we will not attend to the empirical grounds of
Greene's argument until section 6. Berker's most promising normative objection is,
we argue, unsuccessful.

Berker argues that even if it is true that our deontological judgments are sensitive
to morally irrelevant factors and therefore unwarranted, consequentialist judgments
can likewise be debunked:

> Suppose we deem some of the features triggering deontological intuitions to,
> intuitively, be morally irrelevant, thus granting [(2)]. This is a strike against
> deontological intuitions. However, we can only conclude that consequentialist
> intuitions should be privileged over deontological intuitions if a parallel case
> cannot be made against consequentialist intuitions. Moreover, it is open to the
> defender of deontology to reply that, intuitively, the faculty eliciting consequen-
> tialist reactions is also responding to morally irrelevant factors, or failing to
> respond to morally relevant ones. For example, a deontologist could contend that
> the neural processes giving rise to consequentialist judgments are failing to respond
> to morally relevant factors by ignoring the separateness of persons, or by treating
> people as vats of well-being, or by assuming that all value is to-be-promoted, or by
> making morality incompatible with integrity, [etc.]. So basically we have just
> recapitulated the same old battle of intuitions over the plausibility of consequen-
> tialism versus deontology in our evaluation of which sorts of factors are and are not
> morally relevant. (2009, pp. 324–325)

Berker's objection fails for two reasons. Conceding in the passage above that Greene
has produced a "strike against deontological intuitions," Berker wonders whether
consequentialist judgments might likewise be ill grounded. This depends on
"whether a parallel case [can] be made against consequentialist [judgments]." But
what parallel case does Berker have in mind? Absent any empirical basis for this
thought, the idea of a parallel case is no more than speculation. Sure, perhaps
consequentialist judgments about *Bystander* are based on morally irrelevant factors,
but as a mere possibility this does not seriously impugn those judgments. What
Berker must offer—and does not offer—is some positive empirical reason to think
this possibility is actual. Insisting that Berker meet this standard is only fair since,
after all, it is the very standard that Greene applies in his own argument.

Berker's objection fails for a second and more important reason. It is question
begging. Greene's normative premise is that personal harm is a morally irrelevant
factor. Berker agrees that this is an uncontroversial premise and therefore that
Greene's argument does not beg the question against deontology. But the morally
relevant factors that Berker has in mind are extremely controversial in the context of
a debate between consequentialists and deontologists. This is not, despite what Berker
says, "the same old battle of intuitions over . . . which sorts of factors are and are not

morally relevant." Deontologists, Greene is right, must concede that whether or not harm is personal or impersonal is morally irrelevant. However, consequentialists of course will not concede that one fails "to respond to morally relevant factors by ignoring the separateness of persons, or by treating people as vats of well-being, or by assuming that all value is to-be-promoted, or by making morality incompatible with integrity."[8] Thus, although Berker suggests a "parallel case" can be made against consequentialism, the normative premises he offers as part of that case beg the question.[9]

Let us sum up. Greene attempts to undermine deontology by arguing that the judgment about *Footbridge* does not in fact support deontology over consequentialism, because it is based merely on the presence of personal harm. Berker's normative objection against Greene's argument fails. His concrete suggestions about how to debunk, in turn, characteristically consequentialist judgments about *Bystander* reflect question begging normative assumptions. In the next section we present a better and decisive objection. Greene's finding concerning personal harm, we argue, does not supply even a prima facie challenge to judgments about *Footbridge*, nor a fortiori to deontology.

4. Moral Consistency Reasoning

In order to develop our objection to Greene's argument from morally irrelevant factors, we must first describe in detail a form of moral reasoning that we call "moral consistency reasoning," or "consistency reasoning" for short (Campbell & Kumar, forthcoming). Consistency reasoning is, roughly, treating like cases alike—exposing and resolving inconsistencies among judgments about concrete cases. Opposing judgments are inconsistent if the cases are morally similar by one's own lights. Consistency reasoning is common in everyday moral discussion. It also features in applied ethics and the law. It has, unfortunately, received insufficient theoretical attention in the literature.[10] Elsewhere we introduce and defend a psychological model of consistency reasoning and its effect on moral judgment and emotion-based moral intuitions (Campbell & Kumar, forthcoming; see also Campbell, 2009, unpublished manuscript; Campbell & Woodrow, 2003). On the basis of that model we develop a normative account of moral justification that captures consistency reasoning—arguing on empirical grounds that other, more intellectualized accounts fail to do so.

Although we have some ideas about the psychological processes that are at play in consistency reasoning, that won't matter here. However, it will be important, for what follows, to have a clear sense of the form of argument in consistency reasoning. In other kinds of moral arguments, described for example in Rawls' model (Rawls, 1971, pp. 20–21, 48–51) of how to reach reflective equilibrium, we discover an inconsistency between an abstract moral principle and a concrete case moral judgment. For example, a principle permits what seems to be an impermissible action. In consistency reasoning, by contrast, we discover an inconsistency between a

pair of opposing concrete case moral judgments. In the one case we judge that an action is morally wrong, while in the other we judge that an action is morally okay. Inconsistency arises because we perceive no morally relevant differences between the two cases. To restore consistency we must revise the less tenable judgment, or else identify a heretofore unseen morally relevant difference.[11]

Consider, by way of illustration, a consistency argument taken from Thomson's (1971) celebrated defense of abortion. Thomson assumes for the sake of argument that a fetus is a person, a being that possesses full moral rights. Many who oppose abortion believe that killing a fetus is morally wrong even in cases of rape because it violates the fetus' right to life. Thomson, however, asks us to consider another case. You are kidnapped, and against your will your circulatory system is connected to a famous violinist who will die if she is disconnected from you before the end of a nine month term. Here we are inclined to judge that it is not wrong to disconnect yourself. But what is the difference between this case and the case of pregnancy due to rape? In both cases, a person has against her will found herself in a situation in which her right to decide what happens in and to her body conflicts with another person's right to life. Possibly, there is a hidden morally relevant difference between the two cases. Otherwise, either our judgment about the violinist case or our judgment about abortion (of persons) in cases of rape must be revised, on pain of inconsistency.[12]

With consistency reasoning in mind, it is time now to return to Greene's argument from morally irrelevant factors:

(1) The deontological judgment about *Footbridge* is a response to personal harm.
(2) Personal harm is a morally irrelevant factor.

Therefore,

(3) The deontological judgment about *Footbridge* is unwarranted.

(2) is the critical premise, if as before we accept (1) for the sake of argument. Greene claims, and Berker agrees, that it is an uncontroversial normative claim. It is not. (2) is either false or, if modified so that it is true, does not support Greene's conclusion.[13]

Let's begin with the first horn of the dilemma. (2), as stated, is false. That is, personal harm is a morally *relevant* factor. Because harm is morally relevant, the additional fact that the harm is personal does not make it irrelevant. If a person inflicts personal harm, that is a perfectly reasonable basis for moral disapproval of her action. To see this clearly, consider another example Greene (2010, p. 9) discusses in the course of attempting to explain how empirical evidence could be relevant to normative questions. Suppose that researchers find juries' decisions are affected by the race of the defendant. This finding debunks juries' decisions (to some extent, depending on what else affects their decisions) if combined with the following uncontroversial normative premise: race is a morally irrelevant factor in assessing a defendant's guilt. Notice that personal harm is not morally irrelevant in the same way that race is. If I observe Bob stab Jill unprovoked, it is perfectly reasonable for me to base my judgment that Bob has done something morally wrong on the fact that he

has inflicted personal harm on Jill, and perfectly unreasonable to base my judgment on Bob's skin color.

On the first horn of the dilemma, the argument from morally irrelevant factors fails because (2) is false. Of course, there is something to this premise, but it must be revised. Greene might avoid the first horn by arguing that personal harm is a morally irrelevant factor in this sense: judging that it is impermissible to push the large person in *Footbridge* because the harm is *personal rather than impersonal* is incorrect. Yes, but here comes the second horn. It is equally true that judging that it is permissible to flip the switch in *Bystander* because the harm is *impersonal rather than personal* is incorrect. Being swayed morally by the "personalness" of harm is unreasonable, but so is being swayed by the "impersonalness" of harm.

Premise (2) in Greene's argument states that personal harm is a morally irrelevant factor, and this is false. More precisely, *the difference between personal harm and impersonal harm is not a morally relevant difference.* That is, while personal harm *is* a reasonable basis for moral evaluation, the difference between personal and impersonal harm is *not* a reasonable basis for drawing moral distinctions. And now, with the revised premise (in italics above) in place of (2), the argument from morally irrelevant factors is invalid. What follows given the revised normative premise is that our different responses to personal harm in *Footbridge* and to impersonal harm in *Bystander* do not track a morally relevant difference. This *difference* in moral response is, therefore, unwarranted. Greene's empirical findings thus imply that one of the two judgments about trolley cases must be revised, but they do not tell us which. Let us make the steps in consistency reasoning explicit. Call this the "argument from morally irrelevant differences":

    (4) The deontological judgment about *Footbridge* is a response to personal harm.
    (5) The consequentialist judgment about *Bystander* is a response to impersonal harm.
    (6) *The difference between personal harm and impersonal harm is morally irrelevant.*

Therefore,

    (7) Either the deontological judgment about *Footbridge* or the consequentialist judgment about *Bystander* is unwarranted.

Once the normative premise is revised and made plausible, Greene's argument from morally irrelevant factors does not by itself threaten the deontological judgment about *Footbridge*. It casts doubt only on the attempt to reconcile opposing judgments about trolley cases.

This is not what Greene wants, of course, but it is something. Greene's research yields empirical claims (4) and (5) that can be deployed in "debunking consistency reasoning." Philosophers seeking principles that license our opposing judgments about *Footbridge* and *Bystander* should entertain doubts about that project in light of Greene's findings. For example, some philosophers believe that the doctrine of double effect is justified in part because it accounts for intuitions about what actions are permissible in trolley cases. According to Greene's research, however, these

intuitions are in fact based on a morally irrelevant difference between personal and impersonal harm. Therefore, the doctrine of double effect cannot be supported by these intuitions insofar as they are based on this difference. Because the cases are similar, we ought to paint *Footbridge* and *Bystander* with the same brush, on pain of inconsistency. Either we ought to side with consequentialists and say that it is right to sacrifice one for five in both cases; or we should say that it is wrong to knowingly inflict harm in both cases. Greene's findings, therefore, do not undermine deontological moral theory, in particular deontological theory that does not endorse the doctrine of double effect (e.g., Scanlon, 2008). His findings, if sound, would undermine a deontological theory only to the extent that it endorsed the doctrine based on making different judgments about trolley cases.[14]

Greene may be tempted to respond to this criticism as follows. ''Yes,'' he might concede, ''the behavioral findings by themselves show only that either the judgment about *Footbridge* or the judgment about *Bystander* is unwarranted. However, this inconsistency is best resolved by revising the judgment about *Footbridge*, because it is just obvious that one ought to flip the switch in *Bystander*.'' Obvious though it may seem to consequentialists, it is highly controversial which judgment is more plausible. Thomson (1976, pp. 206–208), for one, thinks that, intuitively, it is clear that one is not permitted to push the large person in *Footbridge* and that, perhaps, but not obviously, one is permitted to flip the switch in *Bystander*. Naïve subjects similarly favor the judgment about *Footbridge*. Those who are given *Footbridge* first are significantly more likely to say that it is wrong to flip the switch in *Bystander*, but those who are given *Bystander* first are not significantly more likely to say that it is permissible to push the large person in *Footbridge* (Petrinovich & O'Neill, 1996; Schwitzgebel & Cushman, forthcoming).

Greene's research, we have shown, *is* normatively significant, contrary to Berker (at least if his empirical premises are true, more on which in section 6). The argument from morally irrelevant differences, (4)–(7), derives a normative conclusion from Greene's work. There is no illicit jump of the is-ought gap here: (4) and (5) have normative implications only in combination with a plausible normative premise (6). The conclusion (7) concerns not the truth of a normative position but its warrant. Greene's research indicates not that there is no morally relevant difference between *Footbridge* and *Bystander*, but that our responses to the cases do not track a morally relevant difference. There may be, that is, some other morally relevant difference between *Footbridge* and *Bystander*, though it does not influence our actual responses to the cases. However, the moral-epistemological assumption described in section 2 militates against this possibility. On that assumption, the principal justification for moral theories is our first order intuitions about concrete cases. If the assumption is true, our knowledge of morally relevant differences is grounded in the differential responsiveness of our intuitions. That is to say, a principle that something is (or is not) a relevant difference is justified because it makes sense of distinctions we intuitively make (or refuse to make). It follows that our best way of discovering what morally relevant differences might exist among trolley cases is to examine what actually drives our responses to them (and other

cases). Thus, the moral-epistemological assumption presupposed by Greene's original argument also makes (4)–(7) defeasibly compelling.

Our objection to Greene is different from Berker's objection discussed in section 3. In defense of deontology, Berker claims that the consequentialist judgment about *Bystander* might be sensitive to some other morally irrelevant factor, or fail to respond to some morally relevant factor. We argued the morally relevant or irrelevant factors cited in Berker's objection beg the question in a debate with consequentialists. Furthermore, the objection looks too far afield. Berker is right that the problem is the issue of morally relevant differences between *Footbridge* and *Bystander*, but the problem for Greene is not some other difference: the very difference Greene cites between personal and impersonal harm does not count, contrary to Berker's quick concession, as a "strike" against deontological judgments.

The argument from morally irrelevant factors is the main part of Greene's attack on deontology. The rest of Greene's attack is in his (2008), and also appears in Greene's most recent piece (2010) in a form that lends itself to a response to our objection. Next, we examine these other debunking arguments and explain why they too are unsatisfactory.

5. Greene's Other Debunking Arguments

Greene and colleagues (Greene & Haidt, 2002; Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Greene et al., 2004; Koenigs et al., 2007) have gathered behavioral and neuroscientific evidence for a "dual process" model of moral judgment. According to this model, two cognitive systems guide moral evaluation. "System 1" is domain-specific, modular, quick, automatic, unconscious, and emotion-driven. "System 2" is domain-general, holistic, slow, controlled, conscious, and reasoning-driven. On Greene's controversial version of a dual process model, system 1 is responsible for characteristically deontological judgments about moral dilemmas; system 2 is responsible for characteristically consequentialist judgments about moral dilemmas.[15]

If Greene's experimental findings are accurate, either our deontological judgment about *Footbridge* or our consequentialist judgment about *Bystander* is unwarranted. Greene's dual process model entails that system 1 produces the judgment about *Footbridge*, while system 2 produces the judgment about *Bystander*. Are there any reasons to think that system 1 is less reliable than system 2? Greene (2008, pp. 62–77, 2010, pp. 18–24) thinks there are empirical reasons. System 1, he argues, recruits emotional processing rather than reasoned processing and, as an evolutionary adaptation, exists because it improved fitness, not because it revealed moral truths. Built for efficiency, system 1 employs simple, inflexible heuristics, and is therefore likely to get things wrong in many cases. Thus, because our judgments about *Footbridge* and *Bystander* are jointly unwarranted, and because the source of our judgment about *Footbridge* is unreliable, we should reject the latter judgment and with it a piece of evidence that supports deontology.

This is a condensed summary of Greene's very long discussion. However, the problem with the discussion can be put succinctly before more detail is admitted. Greene argues that system 1 is epistemically suspect because (i) it involves emotional processing, (ii) it is an adaptation, and (iii) it employs simple and inflexible heuristics. However, Greene nowhere presents a compelling case that (i)–(iii) reflect normatively appropriate criteria of evaluation (Berker, 2009, pp. 314–321). He does not tell us why emotional processing is worse than reasoned processing, or why fitness is unlikely to be correlated with moral truth, or why the simplicity and inflexibility of a rule impugns its content when applied to trolley cases. In other words, Greene wants us to accept further debunking arguments, in addition to his argument from morally irrelevant factors, without this time offering us the (seemingly) uncontroversial normative premises needed to make those arguments valid. Greene does not persuasively defend the crucial normative premise below, (9):

> (8) System 1 involves emotional processing (is an adaptation; employs heuristics).
> (9) *Emotional processing (products of natural selection; heuristics) is/are unreliable.*

Therefore,

> (10) System 1 is unreliable.

Let's now look more carefully at Greene's claim that some feature of system 1's operation or history makes it epistemically suspect. One of Greene's arguments is that evolutionary processes unlikely to track moral truth have been responsible for shaping system 1. Other philosophers have developed similar evolutionary debunking arguments (Joyce, 2006; Ruse, 1986; Street, 2006), and one of us has responded at length to these kinds of arguments in print elsewhere (Campbell, 1996; see also Campbell, 2011, section 4). We shall focus here only on the central argument in Greene's most recent piece, though analogous criticisms can be made of his other arguments.

Greene begins by drawing an analogy between system 1 and system 2 and the automatic and manual settings on a camera. Like a camera's automatic settings, system 1 is "highly efficient, and not very flexible" (Greene, 2010, p. 21). System 1 gives us "dispositions to have intuitive emotional responses, gut reactions to actions, people, and other objects of moral evaluation" (Greene, 2010, p. 22). Like a camera's manual settings, system 2 is highly flexible, and not very efficient. It 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" (Greene, 2010, p. 22).

Greene says next:

> 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. (2010, pp. 22–23)

Sounds good. When should we override rather than trust the intuitions we get from system 1? Greene says "we can expect [system 1] to serve us well when our moral problems are familiar and to lead us astray when our moral problems are fundamentally new" (2010, p. 23; see also Allman & Woodward, 2008). Not unreasonable.[16] The problem is that Greene nowhere argues that *Footbridge* is a fundamentally new kind of problem rather than a familiar one. Greene does say with some plausibility that moral problems associated with global warming and global poverty are new problems—i.e., problems we shouldn't trust system 1 to handle. What about the kind of moral problem instantiated by trolley cases? The problem of harming some to help others seems to be old and familiar—not fundamentally new. The burden is on Greene to argue otherwise and he simply doesn't have anything to say. The following argument may be valid but Greene gives us no reason to believe premise (12) is true:

> (11) System 1 employs automatic settings.
> (12) *Automatic settings are unreliable when applied to trolley cases.*

Therefore,

> (13) System 1 is unreliable when applied to trolley cases.

6. General Lessons

Our immediate aim in this essay has been to develop a naturalistic critique of Greene's attempt to debunk deontology. However, our discussion has also been constructive. We argued in section 4 that Greene's findings, if accurate, do have normative significance: they undermine a pair of intuitions that together support the doctrine of double effect. But are Greene's findings accurate? Greene's empirical premise in his argument from morally irrelevant factors is that the deontological judgment about *Footbridge* is a response to personal harm. In our argument from morally irrelevant differences, we add the important counterpart empirical premise that the consequentialist judgment about *Bystander* is a response to impersonal harm. However, it is not clear that this is a complete explanation of subjects' responses. Studies by Cushman and colleagues (Cushman, Young, & Hauser, 2006; Hauser, Cushman, Young, Jin, & Mikhail, 2007) indicate that subjects make a deontological judgment about *Footbridge* in part because the harm is intended; they make a consequentialist judgment about *Bystander* in part because the harm is foreseen but unintended. In short, subjects' responses are also influenced by the distinction embodied in the doctrine of double effect, or something like it (see also Mallon & Nichols, 2010). More recent studies by Greene et al. (2009; see also Huebner, Hauser, & Pettit, 2011) indicate an interaction effect: whether harm is intended, rather than merely foreseen, influences subjects' judgments only if the harm is personal.[17]

It is too early to decide what explains responses to trolley cases. The argument from morally irrelevant differences we offered to Greene, (4)–(7), may be unsound if the empirical premises are false when construed as a complete explanation of

subjects' responses. In any case, however, the argument is valid. We have shown, therefore, that empirical research can be relevant to normative matters, even if Greene's original studies are not. Specifically, our discussion reveals the possibility of *debunking consistency arguments.* Empirical studies can indicate that what accounts for our divergent responses to apparently similar cases does not justify those responses, and therefore that we should withhold from drawing a moral distinction between the cases.

We will formulate our methodological prescriptions abstractly before we illustrate with examples. Consistency reasoning is central to philosophical ethics and plays a major role in a number of different ethical debates (e.g., Singer, 1976, 1993; Unger, 1996). Typically, ethicists engaging in consistency reasoning advance a moral position by arguing as follows. Some action is wrong (or right), contrary to widespread opinion, because it is relevantly similar to another action that is more strongly felt to be wrong (or right). Like cases are not being treated alike, in short, and to achieve consistency the less tenable judgment must be revised. Critics often respond by searching for morally relevant differences between the cases. If there is a morally relevant difference, then the cases can be treated differently without inconsistency.

Empirical research can be of service here. The research cannot of course tell us whether a proposed difference is morally relevant. What it can tell us is which of the usually many differences between the cases is driving the divergent responses. By constructing cases that differ in only one respect—so-called ''minimal pairs''— psychologists can identify the *psychologically efficacious difference.*[18] If the psychologically efficacious difference is not morally relevant, uncontroversially, then the reason we treat the cases differently is not a good reason. Thus, prima facie, we should instead treat the cases alike and reflect on which of our responses is less tenable. The form of debunking consistency arguments can be represented as follows:

(14) The judgment about case A is a response to F.
(15) The opposing judgment about similar case B is a response to G.
(16) The difference between F and G is morally irrelevant.

Therefore,

(17) Either the judgment about case A or the judgment about case B is unwarranted.

As before, the conclusion concerns the epistemic status of a moral claim, not its truth. It is possible that intuitive reactions to the cases have so far been blind to a morally relevant difference and, once a sharp critic identifies it, our intuitions will thereafter be influenced by that difference. The empirical argument is not decisive, therefore. But arguments in ethics (or elsewhere, for that matter) are not bad arguments because they are defeasible.

Let's now make the idea of debunking consistency arguments more concrete by exploring how they could be put to work in two different ethical debates. Singer's (1972) famous case for famine relief employs consistency reasoning. Many people

believe that aid to people in impoverished nations, though admirable, is not something we in the first world are obligated to provide. We have no duty to save starving children in the third world. Singer, however, asks us to imagine the following case. Suppose you are walking next to a pond and you notice a young child drowning. You can save him, but to do so would require diving into the pond and ruining your new suit. Intuitively, you are obligated to save the child. But what is the difference between the case of the drowning child and the (actual) case of starving third world children? In both cases one could save a life at little personal cost. On pain of inconsistency, therefore, we must revise our judgment about one of these cases. Singer thinks, quite plausibly, that less tenable is our judgment about our duties to starving children in the third world.

We need not revise our judgments, however, if there is in fact a morally relevant difference between the cases. So, one might argue that it matters that the drowning child is part of one's community while the starving child isn't, or that you are the only one who can save the drowning child while there are many others in a position to help starving children in the third world, etc. Experimental research has the ability to cast doubt on this type of response. If it turns out that what is driving our divergent responses to the cases is, uncontroversially, a morally irrelevant difference, then it is likely that Singer is right to treat the cases alike. Either we must revise our judgment about famine relief or revise our judgment about our duties to drowning children that can be easily rescued. Experimental research would suggest, in this case, that the search for a morally relevant difference is probably misguided.

Musen (2011) has recently conducted research on Singer's cases. As he notes, there are a number of differences between the case of the drowning child and that of starving children. Unlike starving children, the drowning child (i) is close in proximity, (ii) is directly observed to be in need, (iii) requires non-monetary aid, (iv) faces a one-time emergency, and (v) cannot be helped by anyone else. Constructing pairs of cases that differ in only one of these respects, Musen is able to probe what effect each difference has on intuitions about Singer's original cases.

Musen finds that physical proximity has by far the largest effect on subjects' responses. Whether or not a moral patient is near or far, we are willing to venture, is not a morally relevant difference. So, it appears that the original pair of opposing judgments Singer isolates is jointly unwarranted. One of them must be revised:

(18) The judgment about the drowning child case is a response to the victim's nearness.
(19) The judgment about starving children is a response to the victim's farness.
(20) Physical distance is a morally irrelevant difference.

Therefore,

(21) Either the judgment about the drowning child case or the judgment about starving children is unwarranted.

This suggests we should treat the cases alike. As in all debunking arguments, the conclusion is defeasible. There may yet be a morally relevant difference between the

cases, either one that has so far escaped our notice, or a confound in Musen's apparent minimal pairs. Absent any defeaters, however, the two judgments are together unwarranted, and it is difficult to disagree with Singer's resolution of the inconsistency: we are obliged to save drowning children and also to save starving children in the third world.

Let's now turn to an instance of consistency reasoning discussed already in section 4. Thomson employs reasoning in her defense of abortion that is often interpreted as consistency reasoning, comparing abortion to other activities that we think are morally permissible. Killing a fetus is permissible in cases of rape because it is relevantly similar to the action of disconnecting oneself from and thus killing a famous violinist. Some critics protest that the cases are not at all the same. Indeed, there are many differences and it may be that one or more is morally relevant.

Enter the experimental psychologist. By constructing minimal pairs and presenting them to subjects, she may discover what differences affect their responses. If the differences clearly are not morally relevant then we have good, albeit defeasible reason to accept Thomson's claim that the cases should be treated similarly—just as we might have reason to treat Singer's cases similarly and trolley cases similarly. Empirical research can thus be used to support Thomson's argument: she wants us to treat the cases alike and it may turn out that we treat them differently for no good reason. Of course, this leaves open which of our judgments should be revised, but empirical research would be doing normative work here if it constrained the range of defensible moral positions.

As it happens, new research by Bradner and colleagues (Bradner, Weekes Schroer, & Chin-Parker, unpublished manuscript) explores the factors that influence responses to Thomson's original cases and others like them. Bradner et al. find that subjects are influenced by the presence of a familial relation between mother and fetus, a feature that is absent in the violinist case. When presented with a variation on the violinist case in which one's body is connected to a half-sibling, subjects were more likely to respond that disconnecting oneself is morally wrong. Now, is it morally relevant whether someone is a relative? Arguably, it is. According to many philosophers, we have "special obligations" to our friends and family that we do not have to strangers. Thus, it seems, using Bradner et al's research one *cannot* empirically debunk the original opposing responses to Thomson's abortion and violinist cases. Their studies suggest that a confounding variable in those cases, potentially morally relevant, accounts for our opposing responses. When the confound is eliminated—the case in which one is connected to one's half-sibling instead of a violinist—ordinary people tend to think that is wrong to disconnect oneself.

In this section of the essay we have described, with illustrations, a form of debunking argument that can inform ethical debate. We have shown that Musen's experimental research lends support to Singer. What drives opposing judgments about his drowning and starving child cases is mere physical proximity, and this is morally irrelevant. We treat the cases differently, therefore, for no good reason. Bradner et al.'s experimental research, by contrast, does not lend support to

Thomson, for it fails to show that the opposing judgments about her abortion and violinist cases are responding to a morally irrelevant difference. Arguably, anyway, whether or not someone is a family member affects one's duties to help him or her. Those sympathetic to Thomson might have hoped that treating the cases differently is unwarranted, but Bradner et al.'s research offers no encouragement.

It remains to be seen what kind of normative significance future experimental research will have. We have focused on two prominent consistency arguments in the literature, but there are many others. Sumner (2011), for instance, argues for the permissibility of euthanasia on the grounds that there is no morally relevant difference between euthanasia and other, widely accepted ways that patients end their own lives, when they no longer desire continued life. Consistency reasoning has also been central in recent debates about the moral and legal acceptability of gay marriage, about animal rights, and about the permissibility of torture. As naturalists, we claim that ethical debate that turns on consistency of conflicting moral judgments can profit from experimental investigation of the bases of the intuitions that drive these judgments. Background assumptions about whether differences are morally relevant will always be needed to draw normative conclusions, but the former can be uncontroversial and, together with experimental findings, can justify moral change.

7. Conclusion

Greene's most promising debunking argument, the argument from morally irrelevant factors, contains a normative premise that is either false or does not support the rejection of characteristically deontological judgments about *Footbridge*. However, our modified version of Greene's argument, the argument from morally irrelevant differences, suggests that the judgment about *Footbridge* and the judgment about *Bystander* are jointly unwarranted. If so, then the doctrine of double effect is also unwarranted, to the extent that it is grounded in judgments about trolley cases and these judgments are in fact responses to the differences Greene cites. The empirical premises in this argument may not be supportable in the end, but the viability of this form of argument suggests that analogous approaches based on solid empirical evidence can advance debates in ethics concerning the plausibility of consistency arguments.

Greene's other debunking arguments target the operation or history of the intuitive system that he thinks produces deontological judgments. These arguments require normative premises that are controversial and that Greene has offered critics no good reason to accept. Greene, we conclude, fails to debunk deontological judgments about *Footbridge*, fails to debunk the application of system 1 to trolley cases, and thus fails to debunk the deontological moral theory for which they provide support.

Debunking consistency arguments, we have shown, have a defensible form and can incorporate plausible normative assumptions. If empirical research indicates that we arrive at a moral distinction on the basis of a morally irrelevant difference, then we

have reason to give up that distinction. Debunking consistency arguments show how experimental moral psychology can have normative significance and thus can be a rational basis for moral change.

Acknowledgements

Notes

[1]   Berker (2009, pp. 325–329) argues, more specifically, that neuroscientific research, including Greene's, is not directly relevant to normative ethics. As will become clear, neuroscientific research is inessential to Greene's main argument, as Greene (2010, pp. 14–15) says, but contributes to crucial supporting arguments (see section 5).

[2]   Many philosophers will want to take issue with Greene's classification of moral judgments as characteristically deontological and characteristically consequentialist. We are sympathetic to this worry. Although we are open to Greene's idea of defining ''deontological thinking'' and ''consequentialist thinking'' by appeal to paradigm cases of the judgments associated with each (alleged) mode of moral thinking, we doubt that any such attempt will produce a classification that fits Greene's aims. (Thanks to Selim Berker for discussion.) Greene's argument fails, we will show, even if his classification of judgments about *Footbridge* and *Bystander* is granted for the sake of argument.

[3]   In fact, Greene et al.'s (2009) most recent study on *Footbridge*, *Bystander*, and other trolley cases indicates that the distinction embodied in the principle of double effect between intended harms and unintended but merely foreseen harms (or something like this distinction) also accounts for the variation among subjects' responses to *Footbridge* and *Bystander*. More on this in section 6.

[4]   A rather significant assumption in experimentally based debunking arguments is that philosophers' intuitions have the same bases as the intuitions of psychological subjects. Otherwise it would not follow that philosophers' intuitions in particular are called into question. We won't discuss this here, except to note that it merits attention.

[5]   Berker and Greene express the conclusion of the argument from morally irrelevant factors differently. Berker says, ''so, deontological intuitions, unlike consequentialist intuitions, do not have any genuine normative force'' (2009, p. 321). Greene concludes that deontological judgments about *Footbridge* are ''unreliable'' (2010, p. 16). These are explanations for what we take to be the main issue in a debunking argument: that the relevant judgments are epistemically challenged and so one must withhold judgment. Thus, in our formulation of the conclusion, we say that the deontological judgments are unwarranted.

[6]   In unguarded moments, Greene misrepresents his argument as giving decisive reason to reject deontology in favor of consequentialism. Evaluation of trolley cases, a single class of test cases for deontology and consequentialism, does no such thing. The matter is of course more complicated than this and many other considerations bear on the relative evaluation of consequentialism and deontology, as Greene is aware. His more grand ambition is to debunk all of the intuitions upon which he thinks deontology rests (Greene, 2010, p. 21).

[7] Greene does not offer arguments for the moral-epistemological assumption per se. He argues not for the philosophical claim that the principle evidence for deontology is intuitions, but for the psychological claim that, in general, deontological commitments are based on unconsciously generated, emotion driven intuitions (Greene, 2008). Drawing on Baron (1994), Haidt (2001), and his own research, Greene argues that the arguments and principles people offer in support of their deontological beliefs are post hoc rationalizations and therefore causally downstream from those beliefs. So, whereas we have insisted on the philosophical claim that deontology is justified principally (if at all) by intuitions, Greene argues instead for the psychological claim that deontologists base their beliefs principally on intuitions. In our view, the philosophical claim is needed to rebut the charge of "genetic fallacy." For someone might argue that although intuitions are the initial cause of deontological moral theorizing, more venerable reasoning later sustains and justifies deontology.

[8] About the final possibility, consequentialists might put the point more finely: that integrity, properly conceived, is compatible with maximizing the good.

[9] Berker's "most pressing worry" (2009, p. 325) in his essay is that neuroscience does no work in Greene's argument. Greene, as we explain in section 5, appeals to neuroscientific findings to argue that characteristically deontological judgments are generated by emotional processing in system 1, while characteristically consequentialist judgments are generated by cognitive processing in system 2, Berker argues that this hypothesis about the neural mechanisms underlying moral evaluation is not germane to the argument from morally irrelevant factors, as these systems are defined in section 5 below. Whether Greene has impugned the epistemic status of our deontological intuitions, Berker says, is "purely a function of what sorts of features out there in the world they are each responding to" (2009, p. 325) and not whether the intuitions derive from emotional systems in the brain.

Greene says, and in our view he is right, that "this shouldn't be Berker's most pressing worry" (2010, p. 14). By itself, Berker's objection claims only that the empirical premise in the argument from morally irrelevant factors is supported by Greene's behavioral findings but not by his neuroscientific findings. Greene can simply concede, as he does, that neuroscience is "not essential to [his main] normative argument" (2010, p. 14). Greene does insist that his neuroscientific research, though inessential, supports his dual process model, and so is indirectly relevant to the empirical premise in his argument. In any case, Berker's main worry does not even purport to challenge the soundness of Greene's argument from morally irrelevant factors. If Greene's argument against deontology fails, it cannot be simply because neuroscience is irrelevant to that argument.

However, another interpretation of Berker's main worry is available in light of the fact that there are two versions of Greene's argument from morally irrelevant factors. Our focus has been on the most recent incarnation: the judgment about *Footbridge* is based on personal harm, a morally irrelevant factor, and is therefore unwarranted. Greene's original argument, according to Berker, is broader: deontological judgments in general are based on personal harm, a morally irrelevant factor, and are therefore unwarranted. Berker argues that experimental moral psychology (he says "neuroscience") does no work in the argument because Greene provides no way of defining deontological and consequentialist judgments independently of what he thinks they are responses to. So, Berker thinks, the claim that deontological judgments are responses to personal harm is a question-begging armchair stipulation and not an empirical claim—not supported by any of Greene's research. In our view, this objection fails too. Greene's approach is to identify paradigm cases of deontological and consequentialist judgments without already deciding what is essential to them, and then argue that when empirically examined each has in common that they are responses to one of two contrasting factors. Now, in our view Greene is not in fact right about what each has in common (Campbell & Kumar, forthcoming). But the claim can be empirical and not merely armchair.

[10]   Universalization tests, which Hare (1981) thinks define moral evaluation, are discussed extensively in the literature. ''Moral judgments,'' Hare says, ''are universalizable in . . . that they entail identical judgments about all cases identical in their universal properties'' (1981, p. 108). Arguably, consistency reasoning is the application of a more demanding standard, that cases be identical in all their morally relevant properties.

[11]   In Campbell and Kumar (forthcoming) we argue that the inconsistency at play here is often practical, rather than logical. One emotional response is practically inconsistent with another in the sense that they motivate incompatible behavior.

[12]   Thomson's argument is very often seen as consistency reasoning, as in the context of the research cited in section 6, and for that reason we use it to illustrate this important form of argument. To be fair, it can, and perhaps should, be interpreted otherwise, as giving a counterexample to a general moral principle, such as ''saving an innocent person's life always takes moral priority over personal inconvenience,'' since saving the life of the violinist in the example appears not to have that moral priority.

[13]   It's not entirely clear which version of premise (2) Greene accepts in his (2010) work. Sometimes he says things that suggest the version above, sometimes the version we will defend below. How exactly to interpret Greene does not matter much to us. What matters is that in either case his argument fails.

[14]   A deontological moral theory that endorsed the doctrine of double effect on the basis of its ability to capture intuitions about cases other than trolley cases would not be subject to this debunking argument.

[15]   In Campbell and Kumar (forthcoming) we endorse a ''minimalist'' dual process model of moral judgment but argue that this pattern among ''deontological'' and ''consequentialist'' judgments is illusory. The former do not derive solely from system 1; the latter do not derive solely from system 2. For similar criticisms see also Kahane et al. (forthcoming).

[16]   However, what would it be to successfully deploy this criterion? To claim that a moral problem is familiar or new is to stake out a certain normative position, a position that needs to be made explicit. This is an important issue, but we won't pursue it here.

[17]   Greene has so far paid little philosophical attention to this complication in his published work. He argues, briefly, that the intuitive distinction from which philosophers construct the doctrine of double effect is untrustworthy because, rather than being embodied in an internal norm, it falls out of the structure of our action planning system (Greene, 2010, p. 17). This argument runs afoul of the same kind of problems we discussed in section 5. A crucial normative assumption is required: that the way in which a distinction between means and side effects is represented in our cognitive system makes it likely to be unreliable. Again, we are given no reason to think that is true.

[18]   Sometimes researchers construct minimal pairs in their stimuli, but not in the subject's representation of the stimuli. That is, the stimuli are most easily interpreted in a way that leads to further differences. An important methodological virtue is thus the production of psychologically effective minimal pairs.

References

Allman, J., & Woodward, J. (2008). What are moral intuitions and why should we care about them? A neurobiological perspective. *Philosophical Issues*, *18*, 164–185.

Baron, J. (1994). Nonconsequentialist decisions. *Behavioral and Brain Sciences*, *17*, 1–42.

Berker, S. (2009). The normative insignificance of neuroscience. *Philosophy and Public Affairs*, *37*, 293–329.

Bradner, A., Weekes Schroer, J., & Chin–Parker, S. (unpublished manuscript). When the violinist is your half-sibling: An experimental study of Thomson's classic thought experiment.

Campbell, R. (1996). Can biology make ethics objective? *Biology and Philosophy*, *11*, 21–31.

Campbell, R. (2009). The origin of moral reasons. In L. Johansson, J. Osterberg, & R. Sliwinsky (Eds.), *Logic, ethics, and all that jazz: Essays in honour of Jordan Howard Sobel* (pp. 67–97). Uppsala, Sweden: Department of Philosophy, Uppsala University.

Campbell, R. (2011). Moral epistemology. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved 1 January 2012 from http://plato.stanford.edu/archives/win2011/entries/moral-epistemology/.

Campbell, R. (unpublished manuscript). Naturalizing moral inconsistency: Two models and a new direction.

Campbell, R., & Kumar, V. (forthcoming). Moral reasoning on the ground. *Ethics*.

Campbell, R., & Woodrow, J. (2003). Why Moore's open question is open: The evolution of moral supervenience. *Journal of Value Inquiry*, 37, 353–372.

Cushman, F., Young, F., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science*, 17, 1082–1089.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5, 5–15.

Greene, J. (2003). From neural "is" to moral "ought": What are the moral implications of neuroscientific moral psychology? *Nature Reviews Neuroscience*, 4, 847–850.

Greene, J. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral Psychology: Vol. 3* (pp. 35–80). Cambridge, MA: MIT Press.

Greene, J. (2010). Notes on "The normative insignificance of neuroscience" by Selim Berker. Retrieved 1 January 2011 from http://www.wjh.harvard.edu/~jgreene/GreeneWJH/Greene-Notes-on-Berker-Nov10.pdf

Greene, J., Cushman, F., Stewart, L., Lowenberg, K., Nystrom, L., & Cohen, J. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, 111, 364–371.

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Science*, 6, 517–523.

Greene, J., Morelli, S., Lowenberg, K., Nystrom, L., & Cohen, J. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, 107, 1144–1154.

Greene, J., Nystrom, L., Engell, A., Darley, J., & Cohen, J. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44, 389–400.

Greene, J., Sommerville, R., Nystrom, L., Darley, J., & Cohen, J. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293, 2105–2108.

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, 108, 814–834.

Hare, R.M. (1981). *Moral thinking: Its levels, method, and point*. Oxford: Clarendon Press.

Hauser, M., Cushman, F., Young, L., Jin, R., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind and Language*, 22, 1–21.

Huebner, B., Hauser, M., & Pettit, P. (2011). When are intentional, means based harms morally permissible? Folk–psychological intuitions confirm philosophical intuitions. *Mind and Language*, 26, 210–233.

Joyce, R. (2006). *The evolution of morality*. Cambridge, MA: MIT Press.

Kahane, G., Wiech, K., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (forthcoming). The neural basis of intuitive and counterintuitive judgments. *Social Cognitive and Affective Neuroscience*.

Kamm, F. (1996). *Morality, mortality: Vol. 2. Rights, duties, and status*. Oxford: Oxford University Press.

Kamm, F. (2007). *Intricate ethics*. Oxford: Oxford University Press.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., & Hauser, M. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature*, 446, 908–911.

Mallon, R., & Nichols, S. (2010). Rules. In J. Doris & the Moral Psychology Research Group (Eds.), *Moral psychology handbook* (pp. 297–320). Oxford: Oxford University Press.

Musen, J. (2011). Psychology of obligations to help those in need. Unpublished honors thesis, Harvard University.

Petrinovich, L., & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology*, *17*, 145–171.

Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.

Ruse, M. (1986). *Taking Darwin seriously*. Oxford: Blackwell.

Scanlon, T. (2008). *Moral dimensions: Permissibility, meaning, blame*. Cambridge: Basic Books.

Schwitzgebel, E., & Cushman, F. (forthcoming). Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind and Language*.

Singer, P. (1972). Famine, affluence and morality. *Philosophy and Public Affairs*, *1*, 229–243.

Singer, P. (1976). *Animal liberation*. London: Cape.

Singer, P. (1993). *Practical ethics*. Cambridge: Cambridge University Press.

Street, S. (2006). A Darwinian dilemma for realist theories of value. *Philosophical Studies*, *127*, 109–166.

Sumner, L.W. (2011). *Assisted death: A study in ethics and law*. Oxford: Oxford University Press.

Thomson, J. (1971). A defense of abortion. *Philosophy and Public Affairs*, *1*, 47–66.

Thomson, J. (1976). Killing, letting die, and the trolley problem. *The Monist*, *59*, 204–217.

Thomson, J. (1985). The trolley problem. *Yale Law Journal*, *94*, 1395–1415.

Unger, P. (1996). *Living high and letting die: Our illusion of innocence*. New York: Oxford University Press.