DOI 10.1007/s12152-009-9052-x

Does Neuroscience Undermine Deontological Theory?

Richard Dean

Received: 7 July 2009 / Accepted: 12 October 2009 / Published online: 14 November 2009

Abstract Joshua Greene has argued that several lines of empirical research, including his own fMRI studies of brain activity during moral decision-making, comprise strong evidence against the legitimacy of deontology as a moral theory. This is because, Greene maintains, the empirical studies establish that "characteristically deontological" moral thinking is driven by prepotent emotional reactions which are not a sound basis for morality in the contemporary world, while "characteristically consequentialist" thinking is a more reliable moral guide because it is characterized by greater cognitive command and control. In this essay, I argue that Greene does not succeed in drawing a strong statistical or causal connection between prepotent emotional reactions and deontological theory, and so does not undermine the legitimacy of deontological moral theories. The results that Greene relies on from neuroscience and social psychology do not establish his conclusion that consequentialism is superior to deontology.

Keywords Deontology · Consequentialism · Personal moral dilemmas · Joshua Greene · Neuroscience · Harmless wrongs

Joshua Greene's work in neuroscience deserves the considerable attention it has received from moral philosophers. The fMRI studies that Greene and his colleagues conducted have opened a rich discussion of the neural activity associated with different types of moral thinking, and this line of inquiry may well have significant implications for moral theory. However, I think there is good reason to be skeptical of Greene's recent attempt, in "The Secret Joke of Kant's Soul," to draw particular, anti-deontological conclusions from empirical studies [1].

Greene has argued that several lines of empirical research, including his own fMRI studies of brain activity during moral decision-making, comprise strong evidence against both the reliability of particular "characteristically deontological" judgments and against deontological moral theories, which Greene maintains are elaborate rationalizations of these particular deontological judgments [1]. The fMRI studies, Greene maintains, show that deontological thinking arises from areas of the brain more associated with automatic, "emotional" reactions, while utilitarian thinking arises from more "cognitive" areas of the brain like the anterior cingulate cortex and dorsolateral prefrontal cortex. Similarly, social psychology studies of "harmless wrongs" show an emotional or non-cognitive basis for judgments that harmless (but disgusting or offensive) actions are wrong, and other studies show that retributive (and therefore deontological) judgments regarding punishment of wrongdoers have an emotional rather than cognitive basis. Greene argues that all of these lines of research suggest a strong connection between automatic, unreflective emotional responses and deontological moral judgments, and uses this as evidence that deontological moral theory really, "..essentially, is an attempt to produce rational justifications for emotionally driven moral

R. Dean (✉)
California State University Los Angeles,
Los Angeles, USA
e-mail: rdean@calstatela.edu

judgments" [1: 39]. Deontological theory is a "post hoc rationalization" of emotional reactions, and these "characteristically deontological" reactions are unreliable guides to morality, because they evolved in circumstances that are morally different from, and less complicated than, the circumstances that most of us find ourselves in today [1: 61–63]. Although consequentialism also is a "philosophical manifestation" of an underlying pattern of neural activity, consequentialism is based on more "cognitive" rather than emotional processes in the brain, and so allows "highly flexible behavior" that is responsive to important moral considerations, instead of reflexive "alarm" reactions that may pull an agent away from clear reflection on the morally significant features of a situation [1: 64].

I will argue that the empirical evidence Greene offers does not support the conclusion that deontological theory is primarily a post hoc rationalization of morally unreliable emotional reactions, or that it is therefore inferior to consequentialist moral theory. However, even if I am right that Greene does not marshal enough data to undermine deontological theories in general, a separate question is whether he provides good reason to doubt what he calls "characteristically deontological" judgments in a particular range of cases, in which a deontologist and a consequentialist would *prima facie* disagree about whether an action is permissible (because "e.g. Better to save more lives") or impermissible (because "e.g. It's wrong despite the benefits") [1: 69]. Although there is some evidence in favor of Greene's position that these particular "characteristically deontological" judgments may stem from automatic or emotional processes that are less reliable than reflective, cognitively accessible processes, the case is far from conclusive, and faces some stiff challenges from other empirical studies. So, Greene establishes neither that these particular "characteristically deontological" judgments are unreliable nor that deontological theories are unsound because of being rationalizations of the particular judgments.[1]

Greene's Evidence from Neuroimaging

The fMRI studies that Greene and co-researchers have performed on subjects faced with various hypothetical "moral dilemmas" provide the main inspiration for his anti-deontological position. The main point of Greene's studies is to support a "dual process theory" which identifies two types of neural activity involved in moral judgment—"emotional processes" which Greene associates with deontological judgments, and more "cognitive" processes which he associates with consequentialist moral judgments. Greene rightly acknowledges that "cognition" often refers to "information processing in general," so both types of process are "cognitive" in the broadest sense, but he is interested in another, "more restrictive" use of "cognition" which is meant to contrast with "emotion" [1: 40]. Although there is no consensus within neuroscience or cognitive science on the exact criteria for distinguishing emotional processes from the more narrowly defined "cognitive" processes, the working definitions within the discussion of moral processing typically emphasize differences in automaticity versus conscious deliberation, or motivationally neutral versus behaviorally valenced representations.[2] Greene maintains that the neural activity involved in consequentialist judgments is "cognitive" in the narrower sense, in that it is less automatic and more behaviorally neutral [1: 40], and that for this reason it is more flexible and provides a sounder basis for moral theory than the emotional processes associated with deontological theory.

In the initial fMRI study in this line of research, Greene and his colleagues identified two different types of brain processes, and associated each with a particular type of moral thinking [2]. This study did not focus directly on a distinction between deontological and consequentialist theories, but it did focus on many cases in which a deontologist and a consequentialist typically might disagree. The researchers presented subjects with a set of moral dilemmas they classified as "personal" and a set of moral dilemmas that they classified as "impersonal,"

[1] Although I will employ both philosophical arguments and empirical data, I am not qualified to explore some potential problems, such as disputes about the functions of different areas of the brain [28], or claims that Greene offers an inadequate model of moral decision-making because his discussion does not include any computational theory regarding moral perception [29]. I will also pass over some philosophical issues, such as whether Greene's argument really counts against all deontology or only rationalist deontology [30].

[2] For a survey of some background sources on the reason/emotion distinction, see [24].

along with a set of non-moral dilemmas.[3] The study designated a moral dilemma as "personal" if it involved causing direct, serious bodily harm to a particular person or set of people. Otherwise, if a dilemma involved no serious physical harm, or harm only to unspecified victims, or only required diverting some preexisting threat onto different victims rather than initiating the harm oneself, the dilemma was classified as impersonal.[4] A paradigmatic example of an impersonal case, according to Greene and his co-researchers, is a "trolley case," in which one must choose whether to hit a switch to divert a runaway trolley away from five victims toward a different, single victim. Although the single victim is killed, the harm is caused only by diverting a pre-existing threat, and is not inflicted in a direct, personal way. In contrast, a "footbridge" case, in which the only way to stop a trolley headed toward five victims is to personally shove an innocent passerby into the path of the trolley is a paradigmatic personal dilemma, because the victim is specific, and the harm is inflicted directly and personally. The fMRI scans, conducted on subjects who were asked to classify an action (such as hitting the switch or pushing the passerby) as appropriate or inappropriate, indicated that when considering the personal cases, subjects showed increased activity in "brain areas associated with emotion" [2: 2106] and decreased activity in areas associated with "working memory" or other "higher cognition" ([2: 2106], and [1: 43], respectively). In contrast, when they considered the impersonal moral dilemmas, the subjects' patterns of brain activity more closely resembled the pattern of activity when considering non-moral dilemmas, with less emotional activity and more activity in "cognitive" areas of the brain. Subjects' reaction times also were longer in cases requiring personal harm, which Greene took as evidence that in these cases, there was a conflict between an immediate emotional response and a more cognitive calculation of overall effects, and that it took time to exert cognitive control over the emotional response.

Greene has very recently acknowledged that a statistical reanalysis of the results of this first study shows that the study's results are dubitable, because the lower reaction times for impersonal cases were largely due to several morally obvious "no conflict" cases, in which there was neither a consequentialist nor deontological rationale for an obviously wrong action ([3], responding to [4]). But this first study set the basic framework for later studies that Greene still regards as legitimate evidence for a dual process theory of moral judgment (and so as evidence against deontology). In these later studies, Greene and his colleagues turned from a focus on just impersonal versus personal types of situations to the actual, divergent judgments regarding the moral permissibility of inflicting personal harm to promote better overall consequences.

Greene and his colleagues' second fMRI study, and another subsequent study, ignored non-moral dilemmas and impersonal moral dilemmas, and instead divided the personal dilemmas into difficult and easy cases, depending on how long it took subjects to reach a judgment on the permissibility of directly causing harm [5, 6]. In the second study [5], subjects faced with the difficult personal cases showed the same levels of activity in "emotion-related" areas of the brain as in easy personal cases, but they showed more activity in areas of the brain that had been previously identified with "abstract reasoning, cognitive conflict, and cognitive control," such as the DLPFC, ACC, inferior parietal lobes, and posterior cingulate cortex. In addition, more activity was observed in these "cognitive coordination and command" areas in difficult cases in which subjects decided that personal moral violations were acceptable than in cases in which they deemed the violations unacceptable. The study interpreted these fMRI results as evidence that in difficult decisions about inflicting personal harm for the greater good, emotional reactions conflict with cognitive calculation of costs and benefits, and that subjects who showed more cognitive activity tended to overcome this emotional reaction and ultimately judge the sacrifice to be acceptable because of its greater benefits. A "broader implication" that the researchers draw from the experiment is that "the social-emotional responses that we've inherited from our primate ancestors…undergird the absolute prohibitions that are central to deontology" while "the 'moral

[3] The sets of dilemmas used in [2], [5], and [6] are almost identical, with one impersonal moral dilemma ("eyes") omitted from [5] and with some variations among the core set of personal dilemmas in the three papers. The personal moral dilemmas were also used by Koenigs, et al. in [7], with some omissions.

[4] Greene seems to have recently modified his position regarding exactly "what features of an action elicit [emotional] response" [31: 365] but the core idea that deontological judgments are motivated by "negative emotional response" is the same [31: 365].

calculus' that defines utilitarianism is made possible by more recently evolved structures in the frontal lobes that support abstract thinking and high-level cognitive control" [5: 398]. This foreshadows the objection to deontological theory that Greene develops more fully in "The Secret Joke of Kant's Soul" [1].

More recent studies seem to add weight to the case for associating deontology with emotion and consequentialism with cognitive control. Among these is a study in which Greene and coresearchers asked subjects to reach judgments on difficult personal dilemmas while they were subjected to cognitive load (the subjects were asked to perform an additional cognitive task of identifying numbers scrolling across the screen) [6]. The result, that it took longer for subjects to reach positive "utilitarian" judgments about inflicting harm for the greater good, but no longer for them to reach "deontological" judgments that inflicting the harm was inappropriate, was taken as further evidence that cognition selectively supports "utilitarian" thinking. A study by Michael Koenigs and coresearchers adopted Greene's emphasis on difficult personal moral dilemmas, and showed that patients with damage to their ventromedial prefrontal cortex (VFMPC), an area of the brain associated with social emotion, are more likely than neurologically normal subjects to approve of inflicting harm for the greater good in cases of difficult moral conflict [7]. They take this to support the idea that moral judgment is a dual process system, consisting of "intuitive/affective" and "conscious/rational" mechanisms, and that absolute prohibitions on inflicting harm depend on the affective or emotional systems [7: 910]. So, several studies seem to link emotion with "deontological" moral judgments or cognition with "utilitarian" judgments about the permissibility of harming others.

But to move from the results of these fMRI studies to the conclusion that deontological theory is inferior to consequentialism requires further steps.[5] Greene takes it that the fMRI studies establish a strong correlation between automatic, emotional reactions and deontology, and between more strictly "cognitive," conscious control and consequentialism. The "prepotent emotional responses that drive people to disapprove of the personally harmful actions" are "characteristic of deontology, but not of consequentialism," while "the parts of the brain that exhibit increased activity when people make characteristically consequentialist judgments are those that are most closely associated with higher cognitive functions such as executive control" [1: 46]. Greene regards the argument against deontology as "implicit in the empirical material," but adds that it is "worth spelling it out" [1: 68]. Two claims need to be added to fill out the anti-deontological argument. One is that the emotional responses associated with deontology are an unsound basis for moral theory, and the second is that deontology as a moral theory is really just a rationalization of this set of emotional responses. Greene enthusiastically defends both of these claims.

Greene's position in [1] regarding the unreliability of deontological judgments is an extension and development of views originally expressed in [8], that "an improved understanding of where our intuitions come from, both in terms of their proximate neural/psychological bases and their evolutionary histories" will show that some of our moral judgments, namely the consequentialist ones, are "more reliable than others" [8: 848]. The problem with emotional moral responses, according to Greene, can be seen by looking at how they evolved. Our emotional aversion toward causing direct, personal harm to others evolved in the conditions of Pleistocene hunter-gatherers, when direct evolutionary advantages would have been gained by peaceful cooperation with a small set of other humans who shared one's immediate social set and physical environment. But there were no opportunities to "save the lives of anonymous strangers through modest material sacrifices," the type of impersonal moral situation that Greene regards as central to consequentialist moral theory [1: 47, 59, 70–72, 75]. So humans evolved strong emotional aversions to inflicting personal harm on others, but no similar emotional reactions to "impersonal" moral situations. The strength of our aversion to inflicting personal harm arises from the fact that these emotional aversions "help individuals spread their genes within a social context," [1: 59] not because such aversions "reflect deep, rationally discoverable moral truths" [1: 70]. So this "contingent, nonmoral feature of our history" often is "morally irrelevant" [1: 70, 75], and our evolved emotional reactions are "unlikely to track the moral truth." If deontological theory is based on these reactions, then it is based on morally irrelevant factors, and so is unreliable.

[5] For convenience in quoting Greene, I will follow his practice of using "utilitarian" and "consequentialist" interchangeably.

Greene's second step is to argue that deontological theory is "driven" by such emotional judgments, and in fact "…essentially, is an attempt to produce rational justifications for emotionally driven moral judgments" [1: 39]. Taking it as established by the fMRI studies that "characteristically deontological" judgments are generally the result of emotional processes and "characteristically consequentialist" judgments are generally the result of cognitive processes, Greene goes on to point out the likelihood that deontological theory is a post hoc rationalization of these emotion-driven judgments. It is well-established that "humans are, in general, irrepressible explainers and justifiers of their own behavior," and that "when people don't know what they're doing, they just make up a plausible-sounding story" [1: 61]. There is abundant evidence of the human tendency to construct supposedly rational justifications for their intuitive reactions, even specifically in cases of moral judgment [9–11]. If we "put two and two together," that is, if we combine the fact that the characteristically deontological judgments are "driven largely by intuitive emotional responses" and the fact that we are "prone to rationalization" of our non-rational behavior, we should conclude that deontological theory is "a kind of moral confabulation" [1: 63]. Deontological theory is really an attempt to dress up emotional reactions that are an unsound basis for moral judgment.

Problems with Greene's "Rationalization" Argument

Each of the two premises of Greene's argument—that "characteristically deontological" responses to moral dilemmas are based on morally dubious emotional responses and that deontological moral theories are a rationalization of these emotional responses—is potentially controversial. In this section, I will question only the latter claim. So, for the sake of argument, I will grant the most central results of the fMRI studies that Greene cites, namely that in cases of personal moral dilemmas, the "characteristically deontological" judgments that it is wrong to inflict harm for the greater good are based on more automatic, "emotional" neural processes, and the "characteristically consequentialist" judgments that it is acceptable to inflict personal harm for the greater good tend to be based on more consciously accessible "cognitive" brain processes. In the next section, I will turn to the question of whether the characteristically deontological responses to personal moral dilemmas really are unreliable or defective because of being products of emotional alarm reactions.

Even granting, hypothetically, that typically deontological reactions to Greene's personal moral dilemmas are somehow morally unreliable, it still does not follow that deontological theories are less legitimate than consequentialist theories. This is because, contrary to Greene's position in [1], deontology is not fundamentally a rationalization of these reflexive, emotional reactions. There is not evidence of a strong overall correlation between emotional reactions and deontological theories, or between cognition-based reactions and consequentialist theories. It is worth noting that it would not be surprising if the competition between cognitive and emotional processes in a specific range of cases (personal moral dilemmas) fails to reflect a more general dynamic of two competing systems involved in moral judgment. Empirical research provides good reason for caution about generalizing from specific types of cases to conclusions about human thought processes in general. The specific context and content of scenarios can affect subjects' approach to them, and limit the legitimacy of drawing general conclusions ([12], [13: 383], [13:394]). When applied to Greene's conclusions, this implies that even if there is a conflict between cognitive and emotional processes in cases of difficult personal dilemmas, this may not be indicative of two competing systems involved in moral judgment more generally.

Is Deontological Theory Entirely a "Rationalization" of Emotional Reactions?

Greene often seems to make a strong claim, that deontological theories are nothing but a rationalization of emotional reactions. He points out a "…natural mapping between the content of deontological philosophy and the functional properties of alarmlike emotions" and a similarly "…natural mapping between the content of consequentialist philosophy and the functional properties of 'cognitive' processes" [1: 63–64]. Greene himself says that the bulk of "The Secret Joke of Kant's Soul" is devoted to "identifying a factor, namely emotional response, which predicts deontological judgment" [1: 68]. If the emotional reaction against inflicting personal harm really does track the verdicts of deontological theory closely, then the deontologist is left with a huge coincidence to explain, and it is quite plausible to agree with Greene that there is no "naturalistically respectable explanation" for the coincidence that is as likely as Greene's position that deontological theory is a "post hoc rationalization." Greene offers an analogy: If your friend Alice claims that "her romantic judgments are based on a variety of complicated factors," but you notice that over the course of many years she has actually been attracted only to quite tall men, then you are justified in concluding that she is just rationalizing a simple "height fetish" [1: 67].

But if Greene really means to claim that the emotional reaction to inflicting personal harm identified in the fMRI studies strongly predicts all of the verdicts of deontological theories, then this claim is false. Although there is no single, quintessential deontological theory, any actual moral theory must cover a much wider range of cases than just the cases that Greene calls personal dilemmas, which only involve personal harm. Any remotely plausible moral theory, including any plausible deontological theory, would provide some guidance in many other situations that seem morally significant. Greene's own "impersonal" moral dilemmas include cases of taking money from a lost wallet, falsifying information on a resume, and hiring a black-market surgeon to kidnap a stranger and carve out one of his eyes in order to give you a transplant.[6] And even the Ten Commandments, a simple but influential version of deontology, include only one commandment against direct personal violence ("Thou Shalt Not Kill"). The Ten Commandments also include religious prohibitions, a duty of sexual fidelity, duties to one's family, a requirement of honesty (not to bear false witness), and an injunction against stealing. More sophisticated deontological theories follow this trend of including a wide variety of duties, encompassing not only prohibitions of direct physical harm to others, but typically also some requirements of honesty, duties of station (obligations related to one's profession or family), and often even a requirement of some sort of beneficence or a duty to help others. If this last sounds more consequentialist than deontological, it is worth noting that deontologists from Kant to WD Ross have endorsed it [14–16]. Although the duty to promote human well-being is not the sole priority of deontological theories, it is a duty according to most of them, which underscores the point that focusing only on cases in which promoting the greater good conflicts with deontological requirements is unlikely to capture the real nature of deontology.

The very ease with which the strict "personal harm reactions predict all of deontological theory" interpretation of Greene's argument can be refuted suggests that it must be uncharitable. One alternative reading of Greene's position is that perhaps he means that many different reflexive emotional reactions can be involved in moral judgment, with the emotional aversion to inflicting personal harm being just one of them. In fact, Greene does regard disgust and an emotion-based desire for revenge as other reactions that are encompassed by deontological theory's rationalizations [1]. And one might think there are other emotional reactions available to fuel most deontological judgments. So, to revise the analogy of Alice, it is not that Alice always ends up with tall men, but that we notice that she always ends up with tall men, or grey-eyed men, or poets. She provides rationalizations of her preferences in other terms, regarding compatibility, intellect, kindness, and the like, but really she has a set of factors that drive her romantic choices without her knowing it.

This more charitable interpretation of Greene's position, enfolding a wider variety of emotional reactions as the driving force behind deontological theories, is more plausible than the reading based just on reactions to personal harm. In fact, the more encompassing version of the "rationalization" claim has some empirical support. Greene cites studies that show that feelings of disgust do affect moral judgments in at least some circumstances [1: 58]. And Jana Schaich Borg and coresearchers found that the doctrine of double effect, which they take to be a deontological factor in moral judgment, is correlated with neural activity in emotion-related areas of the brain [17: 814–815]. It is, at any rate, quite possible that some number of unreflective emotional responses contribute to some "deontological" moral judgments that an action is wrong.

But that is far from conclusive evidence that all or most verdicts of deontological theories are driven by various kinds of emotional reactions. In fact, there are substantial obstacles to this modified version of Greene's position. Some of the obstacles are empirical. The same study by Schaich Borg et al. which correlated the (deontological) factor of the doctrine of double effect

[6] The "eyes" story appears in [2] and [6], but is missing from [5].

with activity in emotion-related areas of the brain also correlated another deontological factor, the distinction between acting to harm and merely allowing harm, with activity in cognitive areas of the brain. The study reports that "…when consequences were held constant, moral deliberation about action versus inaction invoked activity in areas of the brain dedicated to high level cognitive processing and suppressed activity in areas associated with socioemotional processing" [17: 813], and an overall conclusion of their study is that "…In contrast to the speculations of Greene, Nystrom, et al. (2001), our data show that some deontological responses (such as the DDA-implied intuition to refrain from action) can be mediated by reason" [17: 815]. A recent study by Greene and Joseph Paxton actually provides further evidence against the idea that deontological theories generally are a rationalization of emotional reactions [18]. The study, which is not directly related to Greene's "dual process theory" line of research, examines neural activity and reaction times involved in subjects' decisions of whether to tell the truth or lie when reporting their success at predicting the outcome of coin flips. One of Greene and Paxton's main conclusions is that the fMRI results do not show signs that honest people exert increased cognitive control activity when they "forgo opportunities for dishonest gain" [18: 12508]. But in addition, whole-brain fMRI analysis showed no increase in any "other processes" either, whether "in the control network or elsewhere," when honest people chose to tell the truth instead of lying for profit [18: 12508].[7] So, besides the main points of the study, one result is that there were not signs of emotional alarm reactions that contributed to the decision not to lie. This means that at least one kind of typically deontological proscription (against lying) does not depend on emotional alarm reactions, since the opportunity to lie did not elicit any such reaction.

This empirical data reinforces an idea that is intuitively plausible, that not all deontological duties are associated with emotional reactions. Of course, many violations of "personal harm" restrictions evoke strong emotional reactions. But many other obligations recognized by many deontological theories, involving lying, tax evasion, breaking promises, contributions to charity, and the like, seem unlikely to elicit alarmlike emotional reactions developed early in human evolutionary history. Considering both the empirical data and the overall, intuitive picture, it seems unlikely that most deontological theories give weight only to reflexive emotional responses. Then deontological theories are not just "rationalizations" of a set of emotional reactions.

Could Emotional Reactions Nevertheless "Drive" Deontological Theory?

A more charitable possible reading of Greene's anti-deontological argument deserves consideration. Although Greene often puts his main argument in terms of emotion "tracking" or "predicting" the verdicts of deontological theory, which makes it sound as if deontological theory is nothing but a rationalization of a particular set of emotional reactions, he also sometimes puts it in other terms, which suggest a more promising version of his argument. He often says that emotional reactions are what really "drive" deontological theory [1: 69–72]. This could mean, not that deontological theories recognize and give weight only to emotional reactions, but rather that emotional reactions are the only thing that sometimes "drive" deontological theories away from an otherwise default process of cognitive, and consequentialist, moral reasoning. This does not imply that deontologists must ignore moral judgments that are based on cognitive processes—deontological moral theories may be a hodgepodge, constructed to give weight to both cognitive and emotional judgments, instead of filtering out the emotional, reflexive reactions as a strictly consequentialist theory would. This reading fits well with Greene's dual-process theory of moral judgment. On this theory, ordinary people, and most moral philosophers, routinely engage in cognitive thinking about moral situations (which tends to lead to consequentialist judgments) and also have emotional reactions to moral situations (which tend to lead them to deontological judgments). The deontological moral philosopher's mistake, on this version of Greene's argument, is not that she completely ignores or rejects cognitive moral processes and their consequentialist results, but that she fails to reject the distorting influence of emotional, deontological reactions and so does not give cognitive processes and consequentialist theory their proper primacy in the moral domain. To modify the story of Alice: It's not

[7] Also see the supplementary material, p. 2, at www.pnas.org/cgi/content/full/0900152106/DCSupplemental

that Alice only chooses tall men or men with grey eyes or poets, it’s that she often seems to make her choices using perfectly rational, reasonable selection processes that lead to sound relationships, but she also displays a pattern of every once in a while choosing to pursue a disastrous relationship with a partner who is tall, grey-eyed, or a poet.

This picture of baseline moral reasoning that is cognitive and consequentialist, disrupted by some irrational, morally unjustified deontological constraints, will of course be appealing to philosophers with consequentialist sympathies. The terminology of calling some duties “side constraints” on the pursuit of the good, which has become common in discussions of consequentialism, seems based on such a picture. But, of course, the picture by itself will not convince deontological moral theorists. Many deontologists would deny that the consequentialist concern with outcomes has any “default” normative priority over other moral considerations. For example, WD Ross treats the duty of beneficence as one prima facie duty among many, and Immanuel Kant argues that there is one basic, non-consequentialist principle of morality and that it must be applied to human circumstances to derive various specific duties, including duties to assist others [14–16]. Greene can be viewed as providing empirical evidence in favor of the consequentialist-friendly picture. He presents evidence that some moral scenarios elicit stronger emotional responses than others, that subjects who make consequentialist-like judgments in these emotionally laden scenarios show signs of higher cognitive activity than those who make deontological-like judgments, and that it takes longer for subjects to reach consequentialist-like judgments when they are under cognitive load. All of these results fit with the picture of cognitive moral thinking that is basically consequentialist, but which sometimes faces disruptions caused by emotional reactions. But the issue is not yet settled. Even granting, for the sake of argument, that cognition is allied with consequentialism and emotion is allied with deontology in Greene’s “difficult personal cases,” in which some harm must be done for the sake of the greater good, it remains an open question whether these cognitive/consequentialist and emotional/deontological alliances hold throughout the rest of the domain of moral thinking. Studies of the difficult personal cases are pertinent evidence, but not comprehensive or conclusive evidence.

It is important, then, to consider other moral situations, or moral decision-making in general, and examine the evidence for and against the claims that cognitive thinking is generally consequentialist, and that non-consequentialist judgments are generally correlated with emotion. Some of the points mentioned above, in disputing whether emotional reactions predict the content of deontological theories, are relevant to this question. At least one deontological factor, the doctrine of double effect, does seem related to activity in emotion-related areas of the brain [17]. But the same study shows that another deontological factor, the distinction between doing and allowing, involves more activity in areas of the brain related to cognition, casting doubt on an overall correlation of cognition and consequentialism or emotion and deontology. Greene and Paxton’s study on lying also failed to show an emotional basis for the typically deontological resistance to lying [18].

A look at Greene’s own examples of “impersonal” dilemmas provides some additional *prima facie* reason to doubt the picture of baseline cognitive moral reasoning that is consequentialist, subject to occasional disruptions from emotional, deontological reactions. One of Greene’s main claims about the impersonal cases is that they are characterized by more activity in cognitive areas of the brain, and less in emotion-related areas, than the personal cases. The signs of emotional “alarm reactions” that characterize the personal cases are absent. Then we should expect, on the face of it, that since the subjects’ thinking was more cognitive, most subjects would reach judgments about the cases that are consistent with consequentialism. But according to supplemental materials for the study, this was not always so [6]. Admittedly, in many of the dilemmas, it was not clear what a consequentialist would decide—as Greene says, “For real-life consequentialism, everything is a complex guessing game, and all judgments are revisable in light of additional details” [1: 65], and in other dilemmas a majority did reach judgments consistent with consequentialism [6]. But in two of the cases, a clear majority of respondents gave answers that seem strongly inconsistent with consequentialism. In the case of the “lost wallet,” the description specifies that you, the finder of a lost wallet, need the money in it much more than the obviously wealthy owner does, and yet 84% of respondents said it was not appropriate to keep the money. And in a case in which you could do great good by donating a small amount of money to charity, 64% of respondents said it was acceptable not to give. Since Greene does not report the relative amounts of activity in different brain areas for each case separately, but only average results for the category of impersonal dilemmas overall, we can not be sure that fMRI studies would show these are cases of highly cognitive but non-consequentialist moral judgments. But, thanks to Koenigs et.al., we do know that the case of the wallet and the donation are not very emotional according to one standard [7]. Koenigs and his coresearchers employed a subset of Greene's impersonal cases, including "lost wallet" and "donation," in their study of patients with VMPFC damage, and they had an independent group of subjects rate each dilemma for "emotional salience" on a scale of 1 to 7. "Lost wallet" and "donation" rated low in emotionality, with a 2.9 and a 1.1, respectively (the average for the impersonal cases was 3.0). So there is at least some reason to think they are cases in which the subjects' approach is more cognitive than emotional, and yet most subjects gave non-consequentialist responses. In addition, the two impersonal cases that Koenigs's subjects rated highest in emotional salience, the trolley case and a case of diverting fumes from one hospital room to another to minimize deaths, were rated just as highly in emotional salience as some of the personal cases (5.3 and 5.5, respectively, while the average for personal cases was 5.9) and yet received overwhelmingly pro-consequentialist responses (82% of respondents approved of diverting the trolley to kill fewer people, 76% approved of diverting the fumes) [6, 7]. These cases at least raise doubts about whether Greene's impersonal cases fit a model in which cognitive moral reasoning supports consequentialism and emotional reactions support deontology.

Looking at more general statistical trends associated with moral decision-making, one might expect that if emotional reactions serve mainly to provide deontological constraints on pursuit of good consequences, then increased activity in emotion-related areas of the brain would be correlated with an increase in judgments that actions are morally unacceptable. But Schaich Borg et al. found that "…emotional activation in the paralimbic system [which includes the brain regions Greene identifies as emotion-related] is not associated with an increased frequency of judgments that something is morally wrong" [17: 816]. This fits a general conclusion of their study, that deontological factors can be associated either with cognitive or emotion-related neural activity.

Another implication of Greene's position that consequentialist approaches to moral judgment are more cognitive and deontological approaches are based more on emotion is that people who think more cognitively or have more cognitive capacities should tend to make more consequentialist judgments, while more intuitive people should tend to make deontological judgments. There is some evidence in favor of these correlations, but more evidence against them. In favor of the cognitive/consequentialist and emotional/deontological links, a study by Daniel Bartels found that subjects inclined to be "intuitive" thinkers were more likely to make characteristically deontological judgments than subjects who employed more "deliberative" thinking styles [13]. However, Bartels also notes that a prior "endorsement of deontological principles" also helps predict moral judgment, so sometimes "deontological preference may be principled" instead of emotion-based [13: 390, 391].

A study by Adam Moore and coresearchers casts even greater doubt on the position that cognitive thinkers are inclined toward consequentialist moral judgment. The study examines the relationship between working memory capacity and decisions about various hypothetical moral scenarios [19]. The researchers identified working memory capacity (WMC) as an indicator of individuals' "cognitive" abilities, in the form of both "controlling emotion and engaging deliberative processing" [19: 550]. This is consistent with Greene's assumption that the dorsolateral prefrontal cortex, an area of the brain heavily involved in working memory, is also associated with cognitive control in moral judgment. Moore et al. presented subjects with twenty four moral scenarios, all involving the death of one person in order to benefit more people, but systematically varying in other ways, including whether the death was caused in a personal or impersonal manner, and whether the one person's death was inevitable (whether the person would end up dead regardless of which option the subjects chose). One of Moore's findings was that subjects with higher WMC (more "cognitive" subjects) did not approve of "consequentialist" choices (to inflict direct harm in order to promote a greater good) more often than other subjects, except in cases in which death was inevitable. The overall lack of correlation between higher WMC and consequentialist choices is straightforward evidence against Greene, but it may be less obvious that even the cognitive thinkers' more "consequentialist" choices about inevitable deaths do not support any connection between cognitive thinking and consequentialism. It is true that cognitive thinkers more often chose to sacrifice one for the sake of many when the one's death was inevitable, but this may not reflect any consequentialist reasoning. To see why, think of a particular example from the study, in which you are asked to imagine you are a paramedic riding in a helicopter along with several injured people. The helicopter malfunctions, loses some power, and you know that it will crash unless one injured person is thrown off. You can not jump off yourself, or else the injured people would die from lack of medical treatment. A consequentialist would choose to throw one injured person off the helicopter for a particular reason, namely that one life must be weighed against many. The fact that the one person who will die is also one of the very same "many" who will die anyway is morally irrelevant, from a consequentialist standpoint—the one who is sacrificed to save many just as well be a bystander who would otherwise survive, rather than one of the many who would die if you do nothing. Since subjects who think more cognitively give weight to inevitability, but not to more general consequentialist calculation, according to Moore's study, inevitability apparently is a cognitive but non-consequentialist factor in moral thinking.

There is nothing paradoxical about the claim that cognitive moral thinkers are no more prone to consequentialist calculation than less cognitive moral thinkers. It would only be paradoxical if the areas of the brain that Greene identifies as "cognitive" were devoted only to numerical and quantitative calculation. But no one, including Greene, claims this—the brain areas also involve activities like planning, abstract reasoning, working memory, and deductive reasoning ([5: 390, 396], [1:40, 46]). So there is abundant conceptual space to suppose it is possible to engage in cognitive, but non-consequentialist, moral deliberation, and in fact there is significant empirical evidence that such deliberation is common.

Without strong empirical evidence that deontological theory is largely motivated by rationalizing unreflective emotional alarm reactions, there is no obvious reason to discount the prima facie evidence that the motivations for developing or defending any moral theory, including a deontological one, appear to be varied. Although some moral philosophers give great weight to particular cases,[8] others start from more general considerations. Theoretical elegance and consistency with the non-moral aspects of a philosophical system have been powerful forces shaping the moral theories of philosophers from Plato to Kant, and beyond. Kant, whom Greene takes to be an arch-deontologist [1], provides a striking illustration of how unlikely it is that deontologists really are mainly motivated by emotion-driven intuitions about a few specific cases. Kant's own stated strategy, the relationship between his moral theory and the rest of his critical philosophy, and the general timeline of the development of different aspects of his moral theory, do not lend themselves easily to Greene's "rationalization" account of deontology. In *Groundwork of the Metaphysics of Morals*, Kant specifically disavows, and condemns, the strategy of starting from intuitions about particular cases, an approach that he calls "popular philosophy" [15: 210–211]. Instead, he starts from more general intuitions about the nature or concept of morality (such as that moral requirements are commands, and must apply to everyone) and tries to derive the content of the categorical imperative from these widely accepted conceptual claims. The nature of the Categorical Imperative, or supreme moral principle, is fairly analogous to the role of the Categories in Kant's theoretical philosophy. Basic moral principles are internal or supplied by each rational being herself, but we can see that they are inescapable because they are necessary preconditions of an unavoidable activity of human reason (the activity of deliberating about what to do). This mirrors the status of the Categories, which Kant argues are rules or organizing concepts that we supply as necessary preconditions for coherent perception and theoretical thinking. The analogy between moral principles and theoretical organizing principles is no coincidence, since Kant says in the *Critique of Practical Reason* that the ultimate aim of his metaphysical and epistemological system is to make room for rational belief in God, freedom, and morality [20: 28–29]. His theoretical philosophy may be, in part, a rationalization designed to support a moral system, but it does not appear that either his theoretical or moral philosophy is a rationalization of intuitions about specific moral dilemmas. It is not until literally years into the development of his overall philosophical system that he begins deriving many

[8] Frances Kamm is a striking example.

specific moral duties. His statement of his strategy in moral philosophy came in 1785 [15], and (although four examples of particular duties appear in that work) his more extensive examination of specific duties comes 12 years later [14]. It is hard to make all this fit a model of deontological theorizing as being driven by intuitions about particular cases.

Are Particular "Characteristically Deontological" Judgments Unreliable?

Above, I granted for the sake of argument that conflicting intuitions work roughly the way Greene says they do in cases of personal moral dilemmas—that in these cases of conflicting reactions, deontological reactions are based on emotion-related neural processes, consequentialist reactions are based on more strictly cognitive neural activity, and the deontological reactions are less reliable because of being based on less flexible, more automatic emotional responses. In this section, I examine whether these claims are actually justified.

The question is important, in two ways. First, there is some evidence that Greene really is mainly concerned with showing that particular, "characteristically deontological" judgments are morally unreliable, at least in personal cases in which they are "at odds with consequentialism" [21:116]. Although he often targets deontological theories, he also says, in a response to a critique by Mark Timmons, that his real concern is "ground-level deontology" rather than deontology as a "metaethical" theory [21:116–117]. Greene says that if one starts with a "would-be deontological foundation" which ends up justifying only characteristically consequentialist judgments, then "we are left with a ground-level utilitarian philosophy" and "as long as starving children get helped and people get shoved in front of speeding trolleys, that's all I care about." [21:117]. It is clear that particular, characteristically deontological judgments (that it is wrong to perform some action even though performing it will maximize net good consequences) are at least a part of Greene's target—"The moral arguments presented here cast doubt on the moral intuitions in question, regardless of whether one wishes to justify them in abstract theoretical terms" [1: 75]. And even if they are not his main target, it is worth considering whether his arguments show that such intuitions are morally unreliable. Even if he does not show that deontological theories overall are flawed, it would be a significant achievement in moral philosophy to show that certain kinds of much-discussed cases are resolvable in favor of consequentialism.

And of course, much of Greene's evidence focuses on exactly these cases in which characteristically consequentialist and characteristically deontological moral judgments conflict. The experiments by Greene and his coresearchers, along with recent studies by others [7, 22] show a steady path of development that Greene regards as support for his claim that characteristically deontological intuitions are dubious because of their correlation with activity in areas of the brain "associated with emotional response and social cognition" [1: 43]. But to see exactly what the data shows about these cases, it will be helpful to employ some terminological caution. To separate the question about the reliability of particular judgments from questions about the larger implications for deontological and consequentialist moral theories, the labels "characteristically deontological judgment" and "characteristically consequentialist judgment" should be replaced by terms focusing more exclusively on the specific judgments that it is wrong or not wrong to inflict personal harm for the sake of promoting the greater good. I will use the phrase "harm is wrong" judgment to replace "deontological" judgment, and "harm is acceptable" judgment to replace the term "consequentialist" judgment, when discussing cases in which one could harm one person to benefit many. This terminology also will be a reminder to avoid begging an important question, since it is not clear that "deontological" alarm reactions against inflicting harm are the only source of the neural activity in emotion-related areas of the brain in personal harm cases.

So, suppose that Greene and others have succeeded in showing an association between emotion-related neural activity and "harm is wrong" judgments when harming one person will provide great benefits, and a similar association between "harm is acceptable" judgments and activity in areas of the brain involved in problem-solving and cognitive control. In what sense does this show that the "harm is wrong" judgments are less morally reliable than the "harm is acceptable" judgments?

The type of emotional reaction that Greene finds particularly problematic in cases of moral dilemmas are the prepotent alarm reactions that are "subserved by processes that in addition to being valenced, are quick and automatic…" [1: 41]. Although not all emotional reactions are quick (there can be stable moods over time), and not all quick or automatic responses are emotional,[9] the literature on the issue of emotion and cognition in "personal dilemmas" generally follows Greene in focusing on quick and automatic emotional reactions as the ones of interest.[10] This automaticity connects to a non-technical, "common sense" idea of one way in which moral judgments may be defective. It is widely thought that reflective, thoughtful moral judgments are more reliable than hasty judgments. Stereotypical "automatic" moral judgments, not subject to rational revision through deliberation, would include reactions of moral disgust to mixed marriage, homosexuality, stem cell research, and the like, and it is plausible enough to discount such reactions. But the data on reaction times (RT) in personal harm cases do not overall support the idea that "harm is wrong" judgments are less reflective or more impulsive in the everyday sense than "harm is acceptable" judgments. Admittedly, the RT data does provide some prima facie support for the position that "harm is wrong" judgments are quicker and less reflective. Greene et al.'s initial study [2] showed a longer RT for giving a "harm is acceptable" response than a "harm is wrong" response in cases of personal harm, which they take as evidence that cognitive processes must override prepotent emotional responses when a subject judges that harm is acceptable for the greater good [2]. A later study by Greene and his colleagues found that placing subjects under cognitive load causes subjects to take longer to reach "harm is acceptable" judgments in difficult personal cases but that cognitive load does not affect RT for reaching "harm is wrong" judgments in these cases [6]. Greene takes this as "evidence for the influence of controlled cognitive processes in moral judgment, and utilitarian moral judgment more specifically" [6: 1144]. But the same cognitive load study showed, contrary to the results in [2], that in cases of no cognitive load—that is, in more usual circumstances for moral judgment—RT for reaching "harm is acceptable" judgments and "harm is wrong" judgments were virtually identical in difficult personal cases [6]. This result is confirmed by another, earlier study by Greene and a study by other researchers ([5: 396], [19: 555]). And RT for personal cases overall, whatever judgments subjects reach, are high enough that it does not seem like the RT data supports the idea of the judgments being hasty or automatic. Greene himself says, regarding high-conflict personal dilemmas, that "the RT data raise doubts about moral judgment as unreflective, as our participants routinely exhibited RTs over 10s, and in some cases over 20s…" [5: 397]. In addition, Moore et al.'s study on working memory capacity (WMC) and decisions in difficult personal cases showed that more "cognitive" subjects (those with higher WMC) actually took longer to reach "harm is acceptable" judgments than subjects with lower WMC, contrary to the faster response that would be expected if reaching "harm is acceptable" judgments is a result of cognitive processes overriding emotional alarm reactions against killing [19: 556]. On balance, RT data provides more evidence against than for the position that "harm is wrong" judgments in personal cases are problematically unreflective or thoughtless.

Neither do fMRI results provide clear support for the position that "harm is wrong" judgments are defective, or inferior to "harm is acceptable" judgments, in personal cases. The activity in different areas of the brain does not, without a contentious interpretive story, even clearly correlate emotional activity with "harm is wrong" judgments or cognitive activity with "harm is acceptable" judgments. Judgments that it is wrong to inflict personal harm for the greater good actually are not more "emotional" according to the fMRI results, than decisions that such harm is not wrong. Greene's own finding is that activity in areas of the brain related to social emotions are similar in all cases of "personal" moral dilemmas, whether the cases are easy or difficult, and whether the subjects in difficult cases approve or disapprove of inflicting harm for the greater good ([5: 392], [1: 45]). Greene's position is not that "harm is wrong" judgments in difficult cases are the result of higher levels of emotion-related brain activity, but rather that "harm is acceptable" judgments in these cases involve more neural activity in "cognitive control and command" related areas of the brain, and that this cognitive control overrides prepotent emotional responses [5, 6]. So, is the problem with the "harm is wrong" judgments in difficult cases that they do not involve enough cognitive activity? Although Greene's

[9] For a sophisticated discussion of a possible analogy between an innate, automatic, but not emotional linguistic faculty and a similar faculty for morality, see [32].

[10] For a survey of the literature, see [24].

studies did show less cognitive activity for such judgments in tough cases than for “harm is acceptable” judgments in tough cases, the “harm is wrong” judgments in these difficult cases nevertheless involved more cognitive activity than (uncontroversial) moral judgments about easy personal cases, judgments which Greene does not question [5: 396]. So it is not that Greene suggests that “harm is wrong” judgments in difficult personal cases simply fail to meet some minimum level of cognitive control. Just in terms of activity in different areas of the brain, there is no obvious reason to regard these “harm is wrong” judgments as especially defective—they involve more cognitive activity than, and only as much emotion-related activity as, many moral judgments that Greene does not question.

Instead, Greene casts doubt on the “harm is wrong” judgments with an overall picture constructed from the fMRI data. The picture is that activity in areas of the brain related to social emotion reflects a potentially distorting influence in personal moral dilemmas, but that in many cases (most of the easy personal cases) there is no conflict between the emotional reactions and cognitive cost/benefit reasoning, so there is no occasion for cognitive, executive control to become engaged and override emotional reactions. But in other cases (the difficult personal cases in which emotional reactions and cost/benefit reasoning conflict), cognitive, executive control attempts to override these emotional reactions. On this picture, cognitive control and decision-making processes, centered in cognitive areas of the brain such as the DLPFC, ally themselves with the “abstract reasoning” or cost/benefit calculation that leads to consequentialist “harm is acceptable” judgments, so “voxels showing significantly greater activity associated with a utilitarian response reflect primarily the successful engagement of cognitive control in support of that response” [5: 396]. This assumes that greater activity in the DLPFC in difficult personal dilemmas mainly or solely reflects an attempt to support cost/benefit reasoning and override a particular kind of emotional reaction, and this assumption is based on a picture in which only one kind of emotional reaction (a prepotent alarm reaction to inflicting harm) and one type of cognitive reasoning (cost/benefit calculation) are competing.

But there is substantial evidence against this simple picture. There are other emotions that may be involved in such cases, and some cognitive processes in these cases seem to support moral reasoning other than consequentialist cost/benefit calculations. The emotional reactions in cases of difficult personal dilemmas may be the result of feeling torn or conflicted, rather than being reactions to personal harm violations [23–25: 102–103]. Jana Schaich Borg and her co-researchers specifically suggest, based on their fMRI results from more specifically sub-divided cases of moral dilemmas, that the emotional activity that Greene et al. identified in cases of personal dilemmas may be a reaction to a conflict of values (promoting greater good versus avoiding direct harmful action) rather than merely a reaction to the possibility of taking direct harmful action [17]. They also suggest that the posterior cingulate cortex, an emotion-related area, may be activated in Greene’s personal dilemmas not because of any “deontological or emotional moral processings” [17: 813], but because the scenarios involve danger or harm to family members or close friends. Not only are a possible multiplicity of emotional reactions involved in Greene’s personal moral dilemmas, but some of them may not be reactions that reasonable people would ignore—Greene himself notes that we may not want to discard commitments that are based on emotion [1: 76] even though we should “rethink at least some” [1: 75]. There is also evidence, as noted above, that the cognitive processes involved in considering personal moral dilemmas include more than just consequentialist calculation. Schaich Borg et al. identify the doing/allowing distinction as involving activity in brain areas dedicated to “high-level cognitive processing” [17: 813]. Moore et al. conclude that the inevitability of someone’s death is a factor in personal moral dilemmas that is processed cognitively, though it is not a consequentialist factor. They also suggest in their conclusion that “automatically and more deliberately processed variables”—what Greene calls automatic emotional reactions and more cognitive moral processes—“may influence the formation of moral judgments when they converge, and not only when they conflict” [19: 556].[11] Schaich Borg et al. not only conclude that deontological responses can be mediated either by “reason” or “emotion” but also

[11] Moore et al. [19] does not involve fMRI neuroimaging, but nevertheless has results relevant to discussion of the role of different types of neural activity.

emphasize that "…a next step for moral researchers is to use network modeling methods to delineate how the regions of the brain identified in these first fMRI studies cooperate and interact" [19: 816]. The variety of neural activity involved in consideration of personal moral dilemmas, in both emotion-related and cognition-related areas of the brain, and the varied roles they may play in contributing to "harm is wrong" or "harm is acceptable" judgments, suggest that it is hasty to assume that the main role of cognitive, executive control is to override emotional responses in order to make "harm is acceptable" judgments possible. As Moore et al. conclude, the cognitive activity involved in difficult personal dilemmas may not serve just to compete with and override "prepotent emotional responses," but rather may be a "selectively engaged, voluntary reasoning system" that engages in "deliberative reasoning" to take account of relevant emotional and cognitive factors and reach a moral judgment [19: 556]. Then "harm is acceptable" judgments may not be more "cognitive" or rationally justified than "harm is wrong" judgments.

Greene's Other Evidence

Besides relying on neuroscientific studies, Greene cites other evidence for the claim that deontology is based on emotion and is therefore less sound than consequentialism. The main additional arguments are based on social psychology studies of two topics: reactions to cases of "harmless wrongs," and subjects' intuitions about retributive versus consequentialist approaches to punishment of criminals [1: 50–58].

The study of harmless wrongs on which Greene primarily relies is a study by Haidt, Koller, and Dias [26], which presented a number of cases of "harmless wrongs" to subjects in Brazil (in an affluent city, Porto Allegre and a poor city, Recife) and in Philadelphia, and asked them to answer questions about the scenarios, including "Is it very wrong, a little wrong, or is it perfectly OK for [specific act description]?" The scenarios included actions that were offensive or disgusting, but did not harm anyone, such as a man masturbating into a chicken carcass before cooking and eating it, a son breaking his promise to his dying mother to visit her grave, a woman cleaning her toilet with the national flag, and a family eating its dog. The study found that the responses varied depending on location (fewer respondents in Philadelphia than in Recife thought the actions were wrong), age (adults were less likely than children to regard the actions as wrong) and socioeconomic status (high SES subjects were less likely to regard the actions as wrong). In order to use these results as support for his position, Greene adds two additional claims. First, he suggests that each of the variables (age, location, and SES) is related to how "cognitive" the respondents are—the older, the more "Westernized," and the higher SES a subject is, the more likely she is to approach moral issues in a cognitive manner [1: 56–7]. Second, Greene connects judging a harmless but offensive action wrong to deontology, and judging that such an action is acceptable to consequentialism [1: 57]. Putting these ingredients together, Greene concludes that the study shows that more cognitively oriented subjects tended to make consequentist judgments, supporting his overall position that consequentialist judgments are more cognitive, and thus more reliable, than deontological judgments.

But the study Greene cites does not support either of the two premises Greene adds in order to reach his anti-deontological conclusion. Instead, it is more or less at odds with both of Greene's additions.

Far from endorsing an idea that degree of westernization or SES are any kind of absolute, reliable indicators of more cognitive (as opposed to emotional) approaches to morality, the Haidt study instead questions the universality of the then-standard "Cognitive-Developmental View" proposed by Piaget and Kohlberg, which "limited the domain of morality to actions that affect the material or psychological well-being of others" [26: 614]. In opposition to this view that harm to others provides a more cognitively developed, transcultural standard for the realm of moral concern., Haidt et al. conclude in their "discussion" section that "the relationships among moral judgment, harm, and affective reactions may be culturally variable" [26: 625]. The basic point of the study is to show that previous studies, which were performed only on subjects in westernized countries, overemphasize one dimension of moral concern, namely harm to others, and underemphasize two other dimensions that are common in other cultures, namely "the ethics of community" (which has to do with a person's "social role") and the "ethics of divinity" (which has to do with a spiritual attempt to "avoid pollution and attain purity and sanctity") [26: 614]. They do not conclude that an exclusive concern with harm or consequences is a sign of a more cognitive approach to moral judgment, but rather that cultural influences shape one's conception of the realm of moral concern. They certainly do not endorse any idea that a more westernized approach is more cognitive or reliable. Neither does the study suggest that higher-SES subjects have a more cognitive or reliable approach to moral judgment. Of course, there may well be some senses in which high SES contributes to cognitive development. Good nutrition is a necessary condition for maximal brain development, and high-SES children are likely to have access to more formal education, so they possess some skills and some types of knowledge to a greater degree than people of low SES. But that is not to say that low-SES subjects are likely to exhibit fundamentally different, less cognitive, neural processes than high-SES subjects. If Greene means to maintain that income and social status play a large role in determining the neural processes involved in moral judgment, then this is at the very least a highly controversial thesis, touching on longstanding debates about nature versus nurture in a particularly volatile way. It is at least as plausible to accept the spirit of the actual study [26], that in making moral judgments, differences in the amount of emphasis on harm versus on social roles or feelings of disgust are largely a product of cultural influences, instead of a measure of how "cognitive" subjects' moral judgments are according to some absolute standard. All in all, Greene's attempt to correlate a more cognitive approach to moral judgment with westernization, or with high socioeconomic standing, is at odds with the study on which he mainly relies.

It is similarly problematic for Greene to append to the study any claim that judging harmless actions to be morally acceptable relies on a fundamentally "consequentialist" approach, or that judging them wrong relies on a deontological approach. Greene says, "In this study, the connection between a reluctance to condemn and consequentialism is fairly straightforward" [1: 57]. But it is not. Haidt et al. do sometimes describe the traditional Kohlbergian approach to defining morality as depending on "personal harmful consequences" or on "acts that have 'intrinsically harmful' consequences to others" [26: 614]. But they clarify, on the same page, that the emphasis on harm is connected to the "ethics of autonomy," which emphasizes not just overall consequences but "harm, rights, and justice." According to the ethics of autonomy, harm to others and violations of rights are the only types of wrong actions, because apart from those, people have a moral right to control their own lives. So the study is not proposing that subjects who deny that harmless actions are wrong must be relying on consequentialist thinking. Instead, the study assumes that they are relying on a fundamentally deontological, autonomy-based approach that counts harm to others as the essential distinguishing feature of the scope of morality. To describe the "not wrong" responses as consequentialist then distorts the position of the study itself. And if "no" responses to the question of whether the action is wrong do not reveal consequentialist thinking, then neither do "yes" responses reveal deontological thinking. Greene himself admits that, "The connection between the tendency to condemn harmless action and deontology is, however, less straightforward and more questionable" than between consequentialism and lack of condemnation [1: 57]. Greene says that the study's scenario of breaking a promise to one's dead mother is "downtown deontology"—a violation that most deontological moral theories would condemn—but provides no overall reason to associate moral judgments that harmless actions are wrong with deontological theory. He mentions that "commonsense moralists" as well as adherents of any deontological theory will tend to condemn some harmless wrongs [1: 55], but the arguments I have given above, in the section on "Problems with Greene's 'Rationalization' Argument", suggest that it is difficult to strongly correlate any particular deontological moral theory with all and only emotional reactions, and that it may therefore be easier to connect non-philosophical, "commonsense" judgments to these emotions. So Greene fails to show that judgments that harmless actions are wrong depend especially on deontological theory, as well as failing to show that judgments that harmless actions are acceptable depend on consequentialist moral theory.

A third main argument that Greene offers for his position that deontology is driven by emotional reactions is based on the connection between emotional reactions to wrongdoing (anger and outrage) and retributive judgments about punishing the wrongdoers. Greene argues that psychological studies show that

> People endorse both consequentialist and retributive justifications for punishment in the

> abstract, but in practice, when faced with more concrete hypothetical choices, people's motives appear to be predominantly retributivist. Moreover, these retributivist inclinations appear to be emotionally driven. [1: 51]

Greene's idea is that since retributive inclinations toward punishing wrongdoers are fueled by emotion, the retributive (and therefore non-consequentialist, or deontological) theories that accommodate these impulses are driven by emotion rather than cognition, providing another example of how deontological theory is an unreliable rationalization of unreflective emotional reactions. Greeen does provide significant evidence that ordinary people's judgments about punishment are more retributive than consequentialist [1: 51–53]. He also describes several studies showing that in many cases, subjects' retributive reactions to real or hypothetical cases of wrongdoing correlate with emotional reactions like anger or moral outrage [1: 54–55]. Greene says less about how exactly the argument against deontology is supposed to proceed from there. But presumably, it would roughly follow the model of his main anti-deontological argument that is based on difficult personal moral dilemmas. So, retributive theories must be driven by these angry, retributive reactions to wrongdoing, and for that reason they are less reliable than consequentialist approaches to punishment, which rely on more cognitive processes of calculating harms and benefits to society. It would be uncharitable to suppose that Greene thinks that non-consequentialist theories of punishment are solely rationalizations of angry reactions—in fact he even says that most retributive theories of punishment allow that "prevention of future harm provides *a* legitimate justification for punishment," and that the distinguishing feature of non-consequentialist theories is that they maintain that "such pragmatic considerations are not the *only* legitimate reason for punishment, or even the main ones" [1: 50]. So the defect of retributive theories apparently is that they diverge from purely cognitive, consequentialist approaches to punishment, in order to accommodate emotion-based retributive intuitions.

This position would parallel Greene's argument from neuroscience against deontological theories. Not surprisingly, it faces parallel challenges. More evidence is needed to show that emotions selectively support retributive reactions, and that consequentialist reactions to wrongdoing are the result of cognitive processes. Without these correlations, retributive and even consequentialist theories of punishment might be driven by a mix of various cognitive and emotional processes, with no clear implications regarding which type of theory is more rationally justified. Of course, Greene's anti-retributivist line of argument is understandably less developed than his main line of anti-deontological argument based on neuroscience. But it is worth noting some points that warrant further examination. One important feature of most retributive theories, which at first glance appears to depend neither on anger-based desire for revenge nor on consequentialist reasoning about consequences for society, is that most retributive theories and most actual legal systems serve as much to limit anger-based retribution as to legitimize or endorse it. One main point of attempts to insure impartiality and to follow an established set of legal procedures is to limit the extent to which victims' or others' outrage and anger directly dictate the ways in which accused wrongdoers are punished. Even the most famous slogan of retributive legal punishment, "An eye for an eye," is an attempt to limit punishment to an amount proportionate to the crime, rather than a demand for bloody vengeance—that is, if someone puts out your eye, you can demand only that he lose his eye, not that he be killed, tortured, or the like. A fuller account of the bases of different aspects of retributive theories would need to explain how the demand for proportionality fails to square with desire for revenge. One possibility is that social emotions actually limit the amount of retributive harm to be inflicted on wrongdoers. A study of difficult personal moral dilemmas by Koenigs et al. [7] associated damage to the VMPFC with both a lessening of social emotions and with an increase in consequentialist judgments in difficult cases, but another study by the same primary investigator [27] showed that patients with the same kind of VMPFC damage acted more vengefully (or emotionally?) in retaliating against unfair offers in the ultimatum game. This raises questions about the exact role of social emotions in both kinds of decisions, with one commentary suggesting that the VMPFC is involved with general pro-social emotions, so that damage to the VMPFC makes subjects less reluctant to inflict harm [28]. If this is correct, then it suggests that the role of emotions in retributive judgments is more complex than just driving subjects toward harming wrongdoers. It is plausible enough, though certainly not proven, that emotions may even contribute to consequentialist judgments in dealing with wrongdoers, by drawing attention to benefits for society. Similarly, more research is needed on the role of cognitive processes in judgments about punishment. Studies of the neural activity associated with various decisions about punishment would be useful, as would studies of possible correlation between subjects overall thinking styles (cognitive or intuitive) and their tendencies in making decisions about how to deal with wrongdoing. Without a good deal more data, Greene's proposal that retributive (deontological) theories of punishment are driven by unreflective emotion is possible, but not at all obvious.

In "The Secret Joke of Kant's Soul," Greene relies on several lines of empirically based arguments, and says, "Any one of the results and interpretations described here may be questioned, but the convergent evidence assembled here makes a decent case for the association between deontology and emotion…" [1: 59]. But I have suggested that there is serious room for doubt about Greene's three main lines of argument, and this in turn undermines his picture of several converging bodies of evidence. Whatever the many fruitful results of the recent boom in empirical research on moral judgment, Greene has not yet shown that the results of this research provide reason to favor consequentialist over deontological moral theory.

References

1. Greene, Joshua. 2008. The secret joke of Kant's soul. In *Moral psychology, volume 3, the neuroscience of morality: Emotion, brain disorders, and development*, ed. Walter Sinnott-Armstrong, 35–79. Cambridge: MIT.
2. Greene, Joshua, Brian R. Sommerville, Leigh Nystrom, John Darley, and Jonathan D. Cohen. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293: 2105–2108. Supplementary descriptions of dilemmas available at: www.sciencemag.org/cgi/content/full/sci;293/5537/2105/DC1.
3. Greene, Joshua. 2009. Dual-process morality and the personal/impersonal distinction: a reply to McGuire, Langdon, Coltheart, and Mackenzie. *Journal of Experimental Social Psychology* 45: 581–584.
4. McGuire, Jonathan, Robyn Langdon, Max Coltheart, and Catriona Mackenzie. 2009. A reanalysis of the personal/impersonal distinction in moral psychology research. *Journal of Experimental Social Psychology* 45: 577–580.
5. Greene, Joshua, Leigh E. Nystrom, Andrew D. Engell, John M. Darley, and Jonathan D. Cohen. 2004. The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44: 389–400. Supplementary description of dilemmas available at: www.neuron.org/cgi/content/full/44/2/389/DC1.
6. Greene, Joshua, Sylvia A. Morelli, Kelly Lowenberg, Leigh E. Nystrom, and Jonathan D. Cohen. 2008. Cognitive load selectively interferes with utilitarian moral judgment. *Cognition* 107: 1144–1154. Supplementary description of dilemmas available at: https://mcl.wjh.harvard.edu/materials/Greene-CogLoadSupMats.pdf.
7. Koenigs, Michael, Liane Young, Ralphs Adolphs, Daniel Tranel, Fiery Cushman, Marc Hauser, and Antonio Damasio. 2007. Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature* 446: 908–911. Supplemental description of dilemmas available: doi: 10.1038/nature05631.
8. Greene, Joshua. 2003. From neural 'is' to moral 'ought': what are the moral implications of neuroscientific moral psychology? *Nature Reviews, Neuroscience* 4: 847–850.
9. Gazzaniga, Michael and Jonathan Le Doux. 1978. *The Integrated Mind*. New York: Plenum.
10. Haidt, Jonathan. 2001. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review* 108(4): 814–834.
11. Nisbett, Richard and Timothy Wilson. 1977. Telling more than we can know: verbal reports on mental processes. *Psychological Review* 84: 231–259.
12. Tetlock, Philip, Randall Peterson, and Jennifer Lerner. 1996. Revising the value pluralism model: Incorporating social content and context postulates. In *Ontario symposium on social and personality psychology: Values*, ed. C. Seligman, J. Olson, and M. Zanna. Hillsdale: Earlbaum.
13. Bartels, Daniel. 2007. Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition* 108: 381–417.
14. Kant, Immanuel. 1996. *The metaphysics of morals*. (Trans: Mary Gregor, from *Die Metaphysik der Sitten*, vol. vi of *Kant's Gesammelte Schriften*, 203–491). Cambridge: Cambridge University Press.
15. Kant, Immanuel. 2002. *Groundwork of the metaphysics of morals*. (Translated and edited by Thomas Hill, Jr. and Arnulf Zweig, from *Grundlegung zur Metaphysik der Sitten*, vol. iv of *Kant's Gesammelte Schriften*, 387–463). Oxford: Oxford University Press.
16. Ross, William David. 2003. *The right and the good*. New York: Oxford University Press.
17. Schaich Borg, Jana, Catherine Hynes, John Van Horn, Scott Grafton, and Walter Sinnott-Armstrong. 2006. Consequences, action, and intention as factors in moral judgments: an fMRI investigation. *Journal of Cognitive Neuroscience* 18(5): 803–817.
18. Greene, Joshua and Joseph Paxton. 2009. Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences USA* 106(30): 12506–12511. Supplementary materials available at: http://www.pnas.org/cgi/content/full/0900152106/DCSupplemental.
19. Moore, Adam, Brian Clark, and Michael Kane. 2008. Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science* 19(6): 549–557. Supplementary description of dilemmas is available at www.uncg.edu/~mjkane/memlab.html.

20. Kant, Immanuel. 1965. *Critique of pure reason* (Trans: Norman Kemp-Smith from *Kritik der Reinen Vernunft*). New York: St. Martin's.
21. Greene, Joshua. 2008. Reply to Mikhail and Timmons. In *Moral psychology, volume 3, the neuroscience of morality: Emotion, brain disorders, and development*, ed. Walter Sinnott-Armstrong, 105–117. Cambridge: MIT.
22. Valdesolo, Piercarlo and David DeSteno. 2006. Manipulations of emotional context shape moral judgment. *Psychological Science* 17(6): 476–477.
23. Luce, Mary, James Bettman, and John Payne. 1997. Choice processing in emotionally difficult decisions. *Journal of Experimental Psychology. Learning, Memory, and Cognition* 23: 384–405.
24. Monin, Benoît, David Pizarro, and Jennifer Beer. 2007. Deciding versus reacting: conceptions of moral judgment in the reason-affect debate. *Review of General Psychology* 11 (2): 99–111.
25. Tetlock, Philip, Orie Kristel, Beth S. Elson, Melanie Green, and Jennifer Lerner. 2000. The psychology of the unthinkable: taboo trade-offs, forbidden base rates, and heretical counter-factuals. *Journal of Personality and Social Psychology* 78: 853–870.
26. Haidt, Jonathan, Silvia Helena Koller, and Maria G. Dias. 1993. Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology* 65 (4): 613–628.
27. Koenigs, Michael and Daniel D. Tranel. 2007. Irrational economic decision-making after ventromedial prefrontal damage: evidence from the ultimatum game. *Neuroscience* 27: 951–956.
28. Moll, Jorge and Ricardo de Oliveira-Souza. 2007. Moral judgments, emotions, and the utilitarian brain. *Trends in Cognitive Science* 11(8): 319–321.
29. Mikhail, John. 2008. Moral cognition and computational theory. In *Moral psychology, volume 3, the neuroscience of morality: Emotion, brain disorders, and development*, ed. Walter Sinnott-Armstrong, 81–91. Cambridge: MIT.
30. Timmons, Mark. 2008. Toward a sentimentalist deontology. In *Moral psychology, volume 3, the neuroscience of morality: emotion, brain disorders, and development*, ed. Walter Sinnott-Armstrong, 93–104. Cambridge: MIT.
31. Greene, Joshua, Fiery Cushman, Lisa Stewart, Kelly Lowenberg, Leigh Nystrom, and Jonathan Cohen. 2009. Pushing moral buttons: the interaction between personal force and intention in moral judgment. *Cognition* 111(3): 364–371.
32. Hauser, Marc. 2006. *Moral minds: How nature designed a universal sense of right and wrong*. New York: Echo/Harper Collins.