Paper accepted by *Synthese* (2018)

**Deontology Defended**[[1]](#endnote-1)

Nora Heinzelmann

Faculty of Philosophy, Ludwig Maximilian University, Munich, Germany

nora.heinzelmann@lmu.de

*Abstract. Empirical research into moral decision-making is often taken to have normative implications. For instance, in his recent book, Joshua Greene (2013) relies on empirical findings to establish utilitarianism as a superior normative ethical theory. Kantian ethics, and deontological ethics more generally, is a rival view that Greene attacks. At the heart of Greene’s argument against deontology is the claim that deontological moral judgments are the product of certain emotions and not of reason. Deontological ethics is a mere rationalization of these emotions. Accordingly Greene maintains that deontology should be abandoned.*

*This paper is a defense of deontological ethical theory. It argues that Greene's argument against deontology needs further support. Greene's empirical evidence is open to alternative interpretations. In particular, it is not clear that Greene’s characterization of alarm-like emotions that are relative to culture and personal experience is empirically tenable. Moreover, it is implausible that such emotions produce specifically deontological judgments. A rival sentimentalist view, according to which all moral judgments are determined by emotion, is at least as plausible given the empirical evidence and independently supported by philosophical theory. I therefore call for an improvement of Greene's argument.*

## **1 Introduction**

Empirical investigations of moral behavior and decision-making have recently gained much attention in moral philosophy. Although interdisciplinary work in the cognitive sciences and moral psychology tends to take a descriptive approach, it seems plausible that scientific insights might have normative implications as well.

Work by Joshua Greene and his colleagues has been a prominent example for such an approach. In a seminal study, Greene and colleagues (2001) provided evidence for the claim that the emotional processing of a moral dilemma determines participants' responses crucially. More recently, Greene (2013, 2007a) has drawn on these and other empirical findings to argue in favor of utilitarianism, a normative approach in moral philosophy, and against rivals like contractualism and deontology (Greene 2013, p. 333). This paper focuses on Greene's critique of non-utilitarian moral theories, in particular deontology. The following second part presents his central argument. The third part discusses more closely his critique of deontology. The fourth and final part develops a reply to this critique.

## **2 Utilitarianism As Greene's Superior Meta-morality**

## Greene (2013, ch. 1) argues that, in order to solve humanity’s most pressing problems today, we have to overcome the multitude of moral outlooks that exist in our globalized world. More precisely, we are confronted with a so-called ‘tragedy of commonsense morality’: for any supporter of a moral theory, it is most advantageous to insist in his or her theory instead of even considering rival views (p. 26). But if each and every one of us behaves like this, we thereby gravely endanger the understanding and cooperation which is required between supporters of rival moral theories (‘moral tribes’) to solve the global problems humanity faces. In order to successfully confront the tragedy of commonsense morality, a meta-morality is required that supporters of all moral theories can adhere to. But how can such a meta-morality be found? Just turning to one's own morality cannot solve the problem; we need to find a moral basis that everyone has in common (pp. 25–7).

Greene turns to science for an answer and in particular to dual-process theory. Roughly, according to dual-process theory, there are two psychological systems in the brain (Kahneman 2011, Stanovich and West 2000). System 1 operates without effort, fast, automatically, and is presumably evolutionarily prior. System 2 operates with effort, slowly, and deliberately. Yet it is able to perform demanding computations. Greene (2013, p. 133) uses the metaphor of a camera as an analogy for the dual-process model: system 2 functions like a camera’s manual mode whereas system 1 corresponds to its automatic settings. For the processing of morals in particular, the automatic settings are moral emotions and the manual mode is a capacity for practical reasoning (2013, p. 15).

Manual-mode moral thinking, Greene argues (p. 172), yields utilitarian moral judgments. Automatic settings in moral thinking, in contrast, yield any other kind of moral judgments, like deontological ones. For instance, ‘trolleyology’ research investigates how humans react to various ‘trolley dilemmas’, and why they do so. There are two paradigm cases[[2]](#endnote-2). In the first (call it ‘switch’), a runaway trolley is headed towards a track, threatening to kill five workmen. You are in a position to operate a switch that would cause the trolley to be diverted onto a sidetrack, where it would kill one workman. In the second case (call it ‘footbridge’), the trolley is again headed towards the five workmen. You are standing on a footbridge together with another workman who is wearing a heavy backpack. By pushing this person off the footbridge, you could stop the trolley. It would kill the person with the rucksack but not the five workmen.

From a utilitarian perspective, it can be argued that one morally ought to both push the person in the footbridge case and divert the trolley in the switch case, as this saves the greatest number of lives in both cases. However, most participants’ responses seem to concur in the switch case only, where the majority favors diverting the trolley. In the footbridge case, people tend to indicate that one ought not push the backpacker. Thereby they seem to conform to a deontological norm which prohibits ‘using’ a human being as a means to an end (i.e., using the lone workman as a ‘trolley stopper’ to save five others).

In a seminal fMRI[[3]](#endnote-3) study (Greene et al. 2001), participants undergoing brain scanning responded to moral dilemmas like the footbridge and switch cases. Greene and his colleagues found that dilemmas like the footbridge case were likely to induce strong activity in brain regions associated with emotions, whereas dilemmas like the switch case triggered higher activity in brain regions linked to reasoning. According to Greene (2013, pp. 124-8), the former regions are part of system 1 and the latter are part of system 2. He therefore hypothesizes that the footbridge case engages system 1 more than system 2. This, in turn, leads to the high percentage of deontological responses. In particular, the strong emotional abhorrence of killing a fellow human being (instead of letting five others die) is more pertinent in the footbridge case than in the switch case[[4]](#endnote-4).

This link between deontological responses to the footbridge case and engagement of system 1 is supported by further evidence: system-1 activation correlates positively with judgments of aversiveness and negatively with judgments of moral acceptability in response to footbridge-style dilemmas (Shenhav and Greene 2014). Patients with damage to system-1 regions are more likely to make utilitarian decisions (Ciaramelli et al. 2011, Koenigs et al. 2007). Moreover, activity in system-1 regions correlates with emotional appraisals and moral judgments in healthy adults (Hutcheson et al. 2015). Lastly, neural correlates of activation for dilemmas like the footbridge case (as compared to less personal or difficult cases) have been reported in system 1 (Greene et al. 2004).

According to Greene (2013), a further difference between system 2 and system 1 is normatively relevant. System 2 thinking is universal – as he puts it, “all humans have more or less the same manual mode machinery” (p. 194). In contrast, automatic settings (or system 1) are to a great extent determined by one’s cultural and personal experiences (pp. 142–143). For instance, as a member of Western culture, “you know in your heart—that is, in your amygdala—that swastikas and men in pointy white hoods are bad news” (p. 143). In short, the moral judgments which are the outcome of manual-mode thinking are comprehensible and acceptable to very nearly everyone. This is not so for moral judgments produced by automatic settings. It is thus highly unlikely to reach a consensus on the latter. Outcomes of automatic settings are, according to Greene's camera metaphor, analogous to non-utilitarian moral judgments, whereas outcomes of manual mode are analogous to utilitarian moral judgments. Utilitarianism, Greene concludes (pp. 15–16, cf. ch. 8), is thus the meta-morality we need to solve the pressing problems humanity faces today and in the future.

To sum up, Greene relies on empirical evidence from neuroscience to make a normative argument: utilitarianism is superior to rival moral theories because it can produce moral agreement supported by a universal psychological system in the brain. In this paper, I shall criticize Greene's approach. I shall focus on just one crucial argument that he makes against deontology. Without this argument, though, Greene’s critique of non-utilitarian moral theories is no longer plausible. Or so I shall argue.

## **3 Greene's Critique Of Deontology**

Greene criticizes all non-utilitarian moral theories on the grounds that they cannot serve as the meta-morality needed to resolve the tragedy of commonsense morality. However, his critique is most developed for deontology and even more specifically for Kant’s ethics (cf. Greene 2007a), which has been regarded as a prime example of deontology. I shall therefore focus on Greene’s critique of Kant’s deontology, although what I say applies, mutatis mutandis, for his critique of non-utilitarian moral theories more generally.

To briefly clarify Greene’s terminology: deontology and consequentialism are, in his view “philosophical manifestations of two dissociable psychological patterns, two different ways of moral thinking” (Greene 2007a, p. 37, cf. 2014, pp. 699–700). Focusing on the functional roles of the two different ethical theories, Greene defines characteristically deontological judgments as those justified in terms of rights, duties, and similar deontological notions whilst characteristically consequentialist judgments are justifiable in terms of impartial cost-benefit analysis or other consequentialist conceptions (Greene 2014, pp. 699–700, 2007a, p. 39). For instance, a judgment that one ought to turn the trolley in the switch case is characteristically consequentialist or utilitarian whereas a judgment that one ought to not push the backpacker in the footbridge case is characteristically deontological (Greene 2014, p. 700, 2007a, p. 42).

Note that this contrasts with the received view that ‘deontology’ denotes a family of normative ethical theories that assign a central role to moral duties (δέον: ‘what is binding’, ‘duty’). They differ in how they spell out these duties. In this vein, consequentialism could be described as a deontological theory[[5]](#endnote-5) which claims that one has a duty to do what has the best consequences. However, deontology is usually regarded as a foil to consequentialism. Sometimes, deontology is even identified with non-consequentialism. It is thus a family of moral theories which claim that it is not only the outcome of an action that determines whether it ought to be performed (Kamm 1999). Kant's ethics is a prominent example of a deontological theory. Its fundamental principle is the so-called categorical imperative. In one famous version, it requires that a person act only in accordance with that maxim which they can at the same time will to become a universal (Kant, G 4:421). Maxims that conform to this requirement are moral permissions or duties. Kant provides the rational agent with a decision procedure to identify what she ought to do, or what is permissible for her to do.

Greene’s terminology is thus non-standard and has been challenged on those grounds (Kahane and Shackel 2010, Klein 2010, Kamm 2009). However, we shall accept it in this paper, bearing in mind that Greene raises a challenge that targets deontology and deontological judgment understood from his point of view, and not the received view in philosophy[[6]](#endnote-6). Let us now take a closer look at it.

In a nutshell, Greene’s critique is the following. He claims that Kant used “manual-mode thinking to *justify* our automatic settings” (Greene 2013, p. 329, cf. pp. 300–301). In other words, Kant engaged in practical reasoning merely to justify moral emotions of system 1. From Greene’s perspective, Kant’s categorical imperative merely serves as post-hoc rationalization of certain emotional reactions. As a normative moral theory, Kant's deontology should thus be rejected.

More precisely, Greene criticizes deontology by giving a debunking explanation for it[[7]](#endnote-7). There is an ongoing debate about (evolutionary) debunking arguments in the literature (Mogensen 2015, Kahane 2011, Street 2006), and various versions of such arguments have been suggested. For brevity's sake, I shall not enter this debate here but merely present and discuss the debunking argument that Greene gives.

According to Greene, deontological moral judgments result from emotional responses. These judgments are then rationalized by (deontological) argument. The deontologist is merely a rationalizer. To criticize this rationalization process, Greene takes the following two steps: first, he identifies a factor which predicts deontological judgments: emotional responses. Second, he argues that this factor is not related to the reasons the rationalizer gives in order to justify his judgment, i.e., deontological moral truth. Taken together, both steps support the claim that it is the rival predictor (the emotions) rather than the reasons given by the rationalizer (moral truth), which determines the moral judgment. The explanation given by the rationalizer is thus debunked.

More formally, this argument can be stated as follows (cf. Greene 2007a)[[8]](#endnote-8):

(P1) “emotional response […] predicts deontological judgment” (p. 68).

(P2) It is false that “emotions somehow track the rationally discoverable deontological moral truth” (p. 69).

(C) Hence, “deontological moral intuitions […] are […] unlikely to track the moral truth” (p. 70).

If (C) is true, then deontological theories are plausibly mere rationalizations of emotional responses. They do not tell us anything about the moral truth. We should reject deontology.

In this paper, I shall argue against (P1) only. Greene provides empirical evidence in its support, to which we shall turn in the following section.

As for (P2), Greene sees no need to argue for this claim. Instead, he maintains that the burden of proof lies on the deontologists' side (Greene 2007a, pp. 68–71): they have to show that the relation between certain kinds of emotions and deontological judgments is genuine and not merely coincidental. They also have to explain how these emotions track moral truths although they are evidently influenced by morally irrelevant factors. Given that it is unlikely that the deontologists can provide reasons for why this should be so, Greene argues, his rival debunking interpretation should be adopted by inference to the best explanation.

## **4 A Critique Of Greene's Argument**

As we have seen, Greene's debunking argument against deontology relies heavily on the claim that emotional responses predict deontological judgments (P1). I am now going to criticize this claim. In doing so, I am not taking a range of related concerns into account that can be and have been raised for Greene. Before turning to my own critique, I shall therefore briefly mention and set aside some of those concerns in order to delineate my case.

The fMRI study that Greene and colleagues published in 2001 has been seminal in that it applied neuroimaging to the study of moral decision-making. Many legitimate questions have been raised about the methodology, epistemology, and philosophical relevance of neuroimaging in general, and in the context of moral philosophy in particular (see Klein 2010 for an overview). Specifically, various authors have pointed out that the data reported in Greene et al. (2001) do not provide evidence for a dual-process morality (Klein 2011) or for consequentialist moral reasoning being more likely to produce correct moral judgments (Kamm 2011, p. 331 et passim), and by themselves have no normative implications at all (Berker 2009). Indeed, the 2001 paper itself does not even mention the word ‘deontology’ and is thus arguably not concerned with it. However, in some of his other writing, Greene does rely on the fMRI studies as well as other evidence to make normative claims. In the current paper, we are foremost concerned with a novel characterization of emotions that he uses in his recent monograph *Moral Tribes* (2013) to make a negative case against non-utilitarian ethics. We shall focus on the example of deontology that has been outlined in the previous section.

In the following section, we are going to take a closer look at Greene’s evidence for the first premise (P1) of his debunking argument against deontology.

### **4.1 Greene’s Evidence Against Deontology**

Greene relies on at least six strands of research to support (P1). First, he invokes his own work (Greene et al. 2001, cf. 2013, ch. 4), which suggests that activity in regions known to process emotions correlates with deontological judgments whilst activation in regions associated with cognitive processing is correlated with utilitarian ones.

Second, according to kin selection and reciprocal altruism, individuals are prone to care for their relatives because they share similar genes, and because reciprocal aid increases evolutionary fitness. Both social phenomena rely heavily on emotions, which in turn shape moral intuitions.

Third, the more of a victim's individuality is conveyed (e.g., by pictures or descriptions) or the closer she is, the more likely and generously people tend to help (Schelling 1968, Singer 1972, Unger 1996, Small and Loewenstein 2003, Nordgren and McDonnell 2011). Greene explains both effects by the stronger emotional responses that identifiable and close-up victims elicit, which in turn generate a felt moral obligation to help.

Fourth, Greene maintains that emotional reactions may explain moral dumbfounding (cf. Haidt et al. 1993): people tend to condemn harmless actions that elicit negative sentiments such as disgust. For instance, incest is often seen as a moral transgression even when no negative results are to be expected.

Fifth, Greene draws on research into confabulation, people’s tendency to invent and give plausible reasons for one's own actions when they are unaware of the true causes (cf. Gazzaniga and Le Doux 1978). Confabulation has been found in split-brain and amnesiac patients as well as healthy persons (Wheatley and Haidt 2005). Confabulation seems to arise from the felt need to give reasonable explanations for one’ actions.

Finally, Greene suggests that emotional reactions linked to deontological moral judgments are specific in two ways: first, a “sort of ‘alarm bell’ emotion” drives deontological but not consequentialist judgment (Greene 2013, pp. 245–254, 2007a, p. 63). “The content of deontological philosophy and the functional properties of alarmlike emotions” map (p. 63). In Greene’s view, the latter are primarily responses to ‘up close and personal’ harms, have a long evolutionary history, are innate and primitive (p. 43). In contrast, “consequentialist moral judgment is not driven by emotion, or at least it is not driven by” such an alarm-like emotion (p. 63). If emotions play any role at all, they are subtle and ‘currency-like’. But consequentialist judgment is essentially determined by cognitive processing. Greene maintains “that there is a natural mapping between the content of consequentialist philosophy and the functional properties of ‘cognitive processes’” (pp. 63–4). Note that, by conceding that consequentialist judgment may after all be sensitive to an emotional response, albeit a very specific one, Greene retracts to some degree the view that consequentialist or utilitarian judgments are purely cognitive. We shall return to this point in section 4.2.

Greene also offers conceptual support for the hypothesis that deontology (or non-utilitarianism more generally) and utilitarianism (or consequentialism more generally) are concerned with and informed by different kinds or features of emotions. Deontology is “a kind of moral confabulation” intended to make sense of “strong feelings that tell us in clear and uncertain terms that some things *simply cannot be done* and that other things *simply must be done*” (Greene 2007a, p. 63). It stipulates inviolable rights, such as the right to not be pushed to death off a footbridge. Deontology thereby justifies those strong emotions. Such rationalization, which supports tribal views by rational argument, is “the great enemy of moral progress” (Greene 2013, p. 301). Consequentialism, in contrast, is “by its very nature systematic and aggregative” and “fundamentally actuarial” (Greene 2007a, p. 64) because it weighs competing interests against each other. Emotions play a role in utilitarian calculus only insofar as they influence our happiness, which is the ‘common currency’ that all affected parties care about: “our concern for happiness and suffering lies behind nearly everything else” (Greene 2013, p. 291, cf. p. 161).

A second feature or kind of emotions specific for non-utilitarian moral judgments is their relativity. This point is novel and central in Greene’s recent *Moral Tribes* (2013). Greene proposes that utilitarianism provides a ‘common currency’ for all moral codes (part III) , fit to resolve intertribal conflict. This is so because utilitarian judgments are produced by ‘manual-mode thinking’, which is universal. In contrast, non-utilitarian judgments, including deontological ones, are produced by ‘automatic mode’, i.e., certain emotional responses. Whilst these may have universal neural correlates (i.e., VMPFC, amygdala, or insula), they “work best when they have been ‘manufactured’ based on lessons learned from past experience” (p. 142). Yet such experiences are not purely universal. They depend on cultural learning as well as personal experience. In short, then, this context-relativity of ‘alarm-like’ emotions is both specific to non-utilitarian judgment and problematic because it fuels tribalism.

Greene offers surprisingly little evidence for cultural relativity of moral emotions that drive non-utilitarian judgments. The only research he cites (2013, p. 369) is by Olsson and Phelps (2004, 2007). They report that Pavlovian fear responses can be acquired through social learning: when participants observed someone receiving electric shocks following certain stimuli, they subsequently showed a fear response to those very stimuli themselves (Olsson and Phelps 2004). This mechanism has also been found in mice and primates, and amygdala activity correlates with learned fear responses (Olsson and Phelps 2007). Evidence that ‘alarm-like’ fear responses can be acquired thus supports Greene’s claim that experience and cultural learning may influence such responses.

### **4.2 Evidence Revisited**

We shall now turn to a critique. In a nutshell, my worry is that the evidence Greene relies on does not establish that emotion-based moral intuitions yield specifically deontological judgments. But if this is true, then Greene is not justified in claiming that emotional response predicts deontological moral judgment (P1). In what follows, I shall focus on two points: first, Greene's evidence in favor of (P1) is not specific enough to corroborate his critique of deontology. Second, evidence given for (P1) is open to at least one alternative interpretation. I shall take each objection in turn, beginning with the former here before turning to the latter in section 4.3.

It is remarkable that most research strands Greene relies on in order to establish (P1) do not specifically pertain to deontology. Evolutionary theories explain solely in what way reactions to social challenges are motivated by emotions and how this mechanism enhances evolutionary fitness. Evidence that people are more likely to help victims that are close-by or identifiable does not show that moral judgments advocating such behavior tend to be deontological. Research into confabulation and dumbfounding merely shows that individuals are likely to give plausible justifications for their actions even in cases where they do in fact not know the causes of their behavior. Yet it has not been shown that these justifications invoke duties, rights, and other deontological terms.

Consider moral dumbfounding in greater detail. Although moral dumbfounding may be a widely observed phenomenon, it seems far from clear that post-hoc rationalizations are mostly deontological. I am not aware of any empirical evidence supporting this claim. There are two further reasons to be doubtful about it. First, Greene himself admits that the relation he establishes between moral dumbfounding and deontological moral judgments is not straightforward and questionable (Greene 2007a, p. 57). Second, Haidt, whose study on dumbfounding Greene refers to, does not make such a connection at all. On the contrary, he argues that all moral reasoning is generally post-hoc rationalized and often influenced by emotions (Haidt 2001, 2012). Neither certain kinds of feelings nor confabulations are thus specifically linked to deontological judgments. Empirical evidence for moral dumbfounding does not support the claim that emotional reactions prompt deontological rationalizations.

Yet the alleged specificity of emotions prompting deontological judgments and Greene’s own fMRI study have been explicitly linked to deontological moral judgments or deontology. Consider first the fMRI study by Greene and colleagues. Note that, as mentioned earlier, this study does not discuss the moral theory of deontology. Only in Greene 2007a and later writings, he relates his empirical findings to deontology. Here, Greene maintains that his data show that people have an inclination toward consequentialist judgments in situations that trigger weak emotional responses and tend to make deontological judgments in cases in which the emotional response is strong (Greene 2007a, p. 43). Greene thus relies on the assumption that it is not emotional responses in general but specific emotions that determine deontological judgments. We shall turn to this claim about specificity in the next paragraph. What is relevant for the interpretation of the fMRI study, though, is that this study does not independently support the claim that emotions predict *deontological* judgments[[9]](#endnote-9). As Klein (2010, pp. 192-3, pp. 194-5, cf. Berker 2009) notes, the fMRI data and further evidence from neuroscience are equally consistent with rival accounts, such as a view according to which emotions can track moral reasons (Moll et al. 2005, 2007). The perspective one takes on the empirical evidence, then, depends substantially on further and independent philosophical assumptions. One need not share Greene’s philosophical outlook.

Consider now the alleged specificity of emotions that prompt deontological judgments. According to Greene, those emotions are specific in two ways: they are ‘alarm-like’, and they are relative to cultural and personal experience. In contrast, a utilitarian judgment is, if at all, concerned with ‘currency-like’ emotions. The manual-mode thinking that produces it is universal. Greene’s premise (P1) is thus, strictly speaking, no longer the claim that “emotional response […] predicts deontological judgment” (Greene 2007a, p. 68). Rather, it is a claim like

(P1.1) Specific emotional response predicts deontological judgment

or

(P1.2) Specific kinds of emotional response predict deontological judgment.

We shall set aside the question of whether it is a certain *kind* of emotion or a specific *feature or property* of emotion that prompts deontological judgment, and the further question of whether the two are co-extensive. In what follows, we shall focus on whether either or both of the two features or kinds of emotion do indeed specifically predict deontological judgment.

Regarding the ‘alarm-like’ nature, feature, or kind, Greene does not provide any empirical evidence for this characterization of emotions. Nor does he explain why one of the two kinds of emotions should be normatively superior to the other (cf. Gill and Nichols 2008, pp. 157–9). Indeed, it appears consistent with the empirical evidence that an alarm-like emotion elicited by the need of five innocent victims may motivate an agent to pull the lever in the switch case and also prompt a utilitarian judgment justifying the decision. We simply do not know what the neural correlates of alarm-like emotions, as compared to currency-like ones, should be. Stipulating that they are the same as those of Greene’s automatic settings (e.g., the amygdala) would be an illegitimate instance of reverse inference. Such alarm-like emotion is merely “*hypothesized* to be involved” in the footbridge but not in the switch case (Greene 2007a, p. 64, my emphasis).

Moreover, Greene’s claims about a supposedly natural mapping of deontological ethics and alarm-like emotions on the one hand and consequentialist and currency-like emotions on the other do not seem very compelling.

First, regarding deontology, he offers little more support for this view than the debunking explanation he aims to establish: deontology’s appeal to rights and duties is a confabulation driven by a strong emotional impulse. Greene does not directly engage with the conceptual foundations of deontological ethics but merely dismisses Kant’s attempt to derive moral truth from first principles as a failure: “not one moral controversy has ever been resolved with a proof from first principles” (Greene 2013, p. 332), and “no one has ever managed to transform Kant’s flawed arguments into rigorous moral proofs” (p. 333).

Second, that deontological rights and duties should “match” strong emotional responses better than consequentialist or utilitarian ones does not appear convincing. Obeying the duty to tell the truth, even if it comes at a high price, is plausibly not prompted by but rather contrary to a strong emotion, and telling a polite lie to get out of a difficult situation might not involve any alarm-like emotions at all (Kahane et al. 2012, cf. Paxton et al. 2013, Kahane 2014). It remains equally unclear why utilitarianism should be more “cognitive” than deontology. Reasoning about whether the workman’s right on the footbridge is trumped by the duty to save the lives of five other victims or not may be a cognitively demanding task. In contrast, a simple utilitarian calculus (comparing the numbers of one life to that of five lives) is less cognitively taxing. Conversely, agents might make deontological judgments without involvement of ‘alarm-like’ emotions. One study (Liao et al. 2011) found this for decisions made about a variant of the switch case (Thomson 1985, p. 1402): here, the side track loops back to the main track, so that hitting the bystander on that side track is required to prevent the trolley from hitting five other victims. A decision not to push the button seems deontological (as in the footbridge case) but it seems as little driven by emotional aversion towards up-close and personal harm as the analogous decision in the switch case.

Regarding the cultural and personal specificity of emotions supposedly prompting deontological responses, the only empirical support Greene refers to is evidence that fear responses can be acquired, for instance from observation of others (Olsson and Phelps 2004, 2007). However, first, whether this finding generalizes from fear to other emotions in other circumstances remains to be seen.

Second, even for emotional responses that are learned from personal experience or cultural upbringing, it is far from clear whether these are specifically associated with non-utilitarian or deontological judgments[[10]](#endnote-10). It is entirely possible that humans acquire from experience or observation a stronger aversion towards greater harm than towards smaller harm, and it may prompt them to judge greater harm to be morally worse than smaller harm. This, it seems, would be a utilitarian judgment on Greene’s definition.

Third, even on the assumption that emotions relative to culture and individuals do determine non-utilitarian judgments it is a further question whether this casts doubt on the normative validity of such judgments. Greene argues from pragmatic considerations that moral judgments that are not universally shared are normatively inferior because they hinder compromise between different ‘moral tribes’ (2013, ch.1 et passim). But it is not clear why a member of one tribe should not be able to respect that members of another tribe have different values, without ceasing to appreciate her own ones. In addition, rather than disregarding all but those values which are universal, it may be pragmatically advantageous to broaden one’s own spectrum of values by taking onboard those of other tribes. Indeed, it may be morally advisable that we take moral values into account even if we ourselves do not endorse them (Haidt 2012, ch. 12). For instance, a liberal’s outlook focuses on harm and fairness norms. But even a liberal's psychology is by its very nature more diverse (Haidt 2012, part II). It is this broad psychological make-up that enables all of us to accommodate moral views radically opposed to our own. It is the prerequisite for respectfully “getting along” with people who hold those views (Haidt 2012, p. xi, p. 364, p. 371, et passim).

In sum, without further evidence it is not plausible to claim that emotional responses shaped by culture and personal experience specifically prompt non-utilitarian judgments.

In short, then, we have reviewed the evidence Greene relies on in order to establish (P1), i.e., the claim that “emotional response […] predicts deontological judgment” (Greene 2007a, p. 68). However, it has turned out that this evidence does not sufficiently establish (P1). We are now going to turn to just one[[11]](#endnote-11) possible rival interpretation of the evidence that seems at least as plausible as (P1) and that does not cast doubt on deontological ethical theory.

### **4.3 An Alternative Re-interpretation Of the Evidence**

According to a plausible rival approach, emotion predicts *all* moral judgments, may they be deontological or utilitarian. This view is commonly associated with the moral sentimentalist tradition (Hume 1998 [1751], Hutcheson 2004 [1725], 2002 [1728], Shaftesbury 2001 [1714]), according to which moral judgments express sentiments of approval or disapproval. The approach has more recently gained renewed traction both in classical and in empirically inspired philosophical research.

Consider the literature on moral responsibility[[12]](#endnote-12). Reactive attitudes, which comprise emotions like gratitude, anger, or resentment, seem to ground our widespread practice of holding others morally responsible (Strawson 1980 [1960]). Empathy and the ability to care about those affected by one’s actions may be a prerequisite for membership in the moral community (Shoemaker 2007). It may even be the case that we are morally responsible for emotional commitments that are without or contrary to our own reason (Shoemaker 2011). At the very least, though, it seems that practices of holding others morally responsible “*are* expressions of our moral attitudes” (Strawson 1980 [1960], p. 27). Along similar lines, norm-expressivism states that moral judgments express non-cognitive attitudes towards rules or norms, notably moral sentiments such as guilt or resentment (Gibbard 1990). The source of moral concern, Gibbard writes (1990, pp. 255–256), are feelings that have a (moral) rationale. Hence, following Strawson, Gibbard, and others, if moral sentiments[[13]](#endnote-13) are essential for all of our moral judgments and practices, then the question arises whether it is sensible, or indeed possible, as Greene recommends, to disregard all but those emotions that feature in utilitarian judgment.

That emotion predicts all moral judgments regardless of whether they are deontological or not is a claim well supported by empirical evidence, as philosophers have noted (cf. Strawson 1980 [1960], p. 26, Gibbard 1990, part I). There is strong evidence from anthropology for a universal, *emotional* currency common to all human beings and moral tribes. Cross-cultural studies have shown that moral violations of autonomy, community, and purity elicit universal emotional reactions of contempt, anger, and disgust, respectively (Rozin et al. 1999, Shweder et al. 1999). Building on this work, psychologists have more recently found that the moral values of care, fairness, loyalty, respect, and sanctity are universal and each characterized by specific emotions: compassion, anger, pride, fear, and disgust, respectively (Haidt 2012, ch. 6, Graham et al. 2011, Haidt and Joseph 2004).

More specifically, there is evidence that emotions determine judgments that could be described, using Greene’s terminology, as characteristically utilitarian or consequentialist. A primary source is behavioral and neuroeconomics, which has shown that emotions are an integral part of all decision-making, may it be moral or non-moral, and can in turn be explained and justified using utilitarian conceptions (Loewenstein 2000, Frank 1988). For one thing, emotions are involved in financial decisions (Bossaerts 2009) as well as in decisions under risk, where they are associated with system-1 regions (Bechara et al. 1997).

More relevantly to our debate, system-1 encoded emotions are known to determine *utilitarian* pro-social judgments (Glimcher and Fehr 2003). For instance, a judgment that money should be distributed equally, even at a cost to oneself, is plausibly justified by impartial welfare concerns and therefore utilitarian according to Greene’s definition. In economic ultimatum games, such choices have been shown to involve emotion as well as system-1 brain regions like the insula (Sanfey et al. 2003). Moreover, charitable donations are justified or even required by utilitarian judgment (Singer 1972) and therefore characteristically utilitarian on Greene’s definition. Such donations are associated with compassion and increased activity in system-1 regions (Moll et al. 2006, Rand et al. 2012).

Note that the existence of universal emotions, universal moral codes and an involvement of emotion in utilitarian judgment do not by themselves challenge Greene’s approach. Greene and the sentimentalist differ over whether emotions play the *dominant* role in *all* of our moral judgments. The sentimentalist affirms this claim, Greene denies it. He assigns a special status to utilitarian judgments, which he regards as *determined* by reasoning, even though they might concern ‘currency-like’ emotions that influence happiness. Whilst a sentimentalist can allow for a normative role of strong or ‘alarm-like’ emotions in resolving pressing moral issues, Greene denies it, assigning it instead to universal and utilitarian reasoning.

But this view is ill supported by science as well as by argument. For instance, for social cooperation aimed at resolving moral issues, emotions are a prerequisite. This includes seemingly irrational and ‘alarm-like’ emotions. Social cooperation is impossible without unreasonable emotions that prompt individuals to punish free riding, to retaliate against rule breakers, or to stick to a promise or demand of reputation at all costs (Frank 1988). An individual may find it impossible to justify these actions by utilitarian reasoning alone. Nevertheless, they are beneficial to the social community. This is so because some behavior that is destructive to a social community can only be prohibited at a cost to the individual or institution that prohibits it. More recently, empirical research has demonstrated the essential need for altruistic punishment, which comes at a cost for the punisher but enforces pro-social behavior (Fehr and Gächter 2002). Importantly, negative emotions towards a defector motivate such altruistic punishment. Not only is reasoning insufficient for cooperation, cooperation requires emotion – the very kind of emotion Greene regards as socially destructive.

A response that could be given on Greene’s behalf is the following[[14]](#endnote-14): moral judgments based on weak emotions or weakly based on emotions are normatively superior to those related to strong emotions (Gill and Nichols 2008, pp. 157–9). This might be so because a strong but not a weak influence of emotion on decision-making impairs successful reasoning. However, it has not been shown that the mere strength of an emotion, or the mere degree to which a judgment is based on emotion, determines the normative status of that judgment (Gill and Nichols 2008, pp. 157–9). On the contrary, some evidence suggests that the absence of emotion is normatively problematic. For instance, sociopaths (or, as they are more commonly called, psychopaths) are known to exhibit both emotional and social deficits (American Psychiatric Association 2013), and the former probably causes the latter (Anderson et al. 1999, Blair et al. 2005, ch. 3).

Greene’s fMRI data, too, can be interpreted from a sentimentalist perspective as supporting the view that emotions predict moral judgments, regardless of whether those judgments are utilitarian or deontological. People may solve moral dilemmas by calibrating conflicting rules of different emotional strengths, which determine moral decision (Prinz 2016, pp. 56–60, 2007, pp. 24–5). For example, in the footbridge dilemma the rules ‘do not kill!’ and ‘save lives!’ conflict. Normally, the former rule is obeyed because it applies higher emotional pressure to the decision-maker. Cognitive processing plays no crucial role. The initial fMRI study (Greene et al. 2001) revealed not only significantly stronger activation for system 1 regions for footbridge-type scenarios but also for switch-type cases, as compared to non-moral choices. Consistent with a sentimentalist approach, activation was larger for trolley cases that can be expected to engage emotions more.

Moreover, whilst neural correlates for the *type* of response – utilitarian or deontological – can be dissociated, they do not map onto the distinction between system 1 and system 2 (Greene et al. 2004). For one thing, DLPFC (system 2) activation has not been specifically associated with utilitarian judgment (Greene et al. 2004) but non-specifically with low-conflict switch-style as well as non-moral cases (Greene et al. 2001). Even more importantly, Greene and colleagues find that brain areas associated with emotions predict utilitarian judgment, and accordingly take a sentimentalist stance: “like David Hume (Hume, 1978), we suspect that all action, whether driven by ‘cognitive’ judgment or not, must have some affective basis” (Greene et al. 2004, p. 397).

Lastly, the VMPFC, a system-1 region, has more recently been found not to predict deontological judgment specifically but moral judgment more generally (Shenhav and Greene 2014). The same study has associated utilitarian judgment with activation in regions known to process emotions, such as posterior cingulate cortex and insula. These findings corroborate previous evidence that VMPFC activity is associated with fair and equal allocations of money (Tricomi et al. 2010, Tabibnia et al. 2008), that charitable giving is determined by system 1 (Rand et al. 2012), and that fairness is hedonically valued (Tabibnia et al. 2008).

Taken together, this strand of research again supports the view that emotions are involved in all kinds of moral judgments, including utilitarian ones.

Along similar lines, utilitarian judgments may be interpreted as determined by decreased aversion to harm rather than by reasoning (Liao 2016, p. 32). This is supported by evidence that patients with damage to the VMPFC are more likely to make utilitarian judgments (Koenigs et al., 2007, Ciaramelli et al., 2007), that they experience emotions like fury (Koenigs and Tranel 2007), and that they perform worse in reasoning tasks (Fellows and Farah 2007, Nicolle and Goel 2013). Whilst evidence from VMPFC patients is consistent with dual-process theory (Greene 2007c), a sentimentalist interpretation is equally strong and more ontologically parsimonious because it stipulates a single process instead of two (Moll and de Oliveira-Souza 2007). Moreover, lack of empathy, such as in sociopaths, is associated with a greater likelihood of making utilitarian decisions (Bartels and Pizarro 2011, Young et al. 2012, Côté et al. 2013). Lastly, pharmaceutical enhancement of cognitive control in system-2 regions does not promote utilitarian responses in the ultimatum game; instead, enhancement of emotional sensitivity in system-1 regions does have this effect (Crockett et al. 2010).

In short, it might not be the dominance of reasoning but rather a lack of concern for others that determines utilitarian judgment.

Not only, then, does it seem plausible and entirely consistent with recent evidence that emotion may predict all kinds of moral judgments. Moreover, those judgments may be rationalized by deontological, utilitarian, or altogether different argument. Psychologist Jonathan Haidt interprets Greene et al.’s (2001) study along those lines: “the areas of the brain involved in emotional processing activate almost immediately, and high activity in these areas correlates with the kinds of moral judgments or decisions that people ultimately make” (Haidt 2012, p. 78). Haidt thus treats all moral judgments alike, whereas Greene maintains that only some of them (viz. the utilitarian judgments) are unique in that they correlate with increased activity in the DLPFC, a brain region associated with controlled cognition (Greene 2013, p. 127). The evidence seems consistent with either view. At times, Haidt and Greene even contradict each other directly, for instance on the question of cognitive load. If participants perform a simultaneous cognitively taxing task whilst making moral decisions, and if this selectively interferes with their utilitarian but not their deontological judgments, then this would support Greene’s view, according to which utilitarian but not deontological judgments involve cognitive processing. Greene (2013, p. 127, cf. Greene et al. 2008) claims that empirical evidence shows precisely this but Haidt (2012, p. 42) maintains that it does not. Future research may conclusively settle this debate; for now it seems advisable to not invoke findings on cognitive load to argue for one position or the other.

However, beyond neuroscience, there is independent evidence for Haidt’s social intuitionist model according to which all (not just deontological) moral judgments are determined by intuitions and post hoc rationalized by reasoning (Haidt 2001, 2012, pp. 55–56). Intuitions comprise but are not limited to emotional responses. All moral judgments prompted by emotional response are rationalized but so are moral judgments prompted by non-emotional intuitions. The model is supported by evidence from the dumbfounding and confabulation studies already mentioned above. As I have argued, this evidence does not show that made-up justifications of moral judgments draw on deontological theory. It merely shows that agents readily provide moral reasons to justify a previously made judgment. What is more, these reasons need not be inappropriate or absurd, they may on the contrary be quite adequate (Liao 2011).

To briefly summarize my critique: Greene aims to reject rival ethical views by providing a debunking explanation for moral judgments justified by those views. In particular, his debunking argument against deontology relies on the premise (P1) that emotional reaction predicts deontological judgment, or, more accurately, that some *specific* emotional reaction does so. I have argued that the empirical evidence Greene relies on does not unanimously favor this premise. It is highly questionable whether the judgments predicted by certain emotional responses are specifically deontological. There is no evidence which makes it plausible that certain emotions are relative to culture and personal experience whereas reasoning is not, or that there is a meaningful distinction between ‘alarm-like’ and ‘currency-like’ emotions. What is more, there is no conclusive evidence that either of the two kinds or features of emotion determines deontological judgments. A rival explanation of moral judgments that is not hostile to deontological ethical theory is equally consistent with the evidence and enjoys independent support from philosophical research.

**6 Conclusion**

Greene (2013) claims that utilitarianism is superior to rival ethical theories, notably deontology. He justifies this, very roughly, as follows: whilst a universal psychological system of cost-benefit reasoning produces utilitarian judgments, non-utilitarian judgments are heavily affected by emotional processes, which are relative to culture and personal experience. Therefore, persons with opposing moral outlooks can more easily reach an agreement on the former than on the latter kind of judgments, which is in turn needed to overcome the pressing problems our multipolar world faces today.

This paper has focused on Greene’s debunking argument against non-utilitarian ethics (and thus against deontology), which relies on the premise that certain emotions predict deontological judgment. I have examined the empirical evidence for two features or kinds of emotions that are supposed to be specifically associated with deontological judgment. On the one hand, Greene hypothesizes that they have a strong or alarm-like nature. On the other hand, he maintains that they are relative to culture and personal experience. I have argued that the empirical evidence does not suffice to establish either of those characterizations. In addition, I have made plausible that alternative accounts like sentimentalism are equally consistent with the empirical data but do not target deontological ethics. In conclusion, as it stands, Greene’s debunking argument is false because one of its premises is untenable. Deontology has been successfully defended.[[15]](#endnote-15)

## **References**

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders*, fifth edn, American Psychiatric Association, Arlington (VA).

Bartels, D. and Pizarro, D. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition* *121*(1), 154–161.

Bechara, A., Damasio, H., Tranel, D. and Damasio, A. (1997): Deciding advantageously before knowing the advantageous strategy. *Science 275*, 1293–1295.

Berker, S., (2009). The normative insignificance of neuroscience. *Philosophy and Public Affairs*, 37(4), 293-329.

Blair, J., Mitchell, D., and Blair, K. (2005). *The psychopath : emotion and the brain*. Malden (MA): Blackwell.

Bossaerts, P. (2009). What Decision Neuroscience Teaches Us About Financial Decision Making. *Annual Review of Financial Economics* *1*(1), 383–404.

Campbell, R. (2011). Moral Epistemology. In: E. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Spring 2014 Edition).

Ciaramelli, E., Muccioli, M., Làdavas, E. and di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social cognitive and affective neuroscience*, *2*(2), 84–92.

Côté, S., Piff, P. and Willer, R. (2013). For whom do the ends justify the means? *Journal of Personality and Social Psychology*, *104*(3), 490.

Crockett, M., Clark, L., Hauser, M. and Robbins, T. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences*, *107*(40), 17433–17438.

Fehr, E. and Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*(6868), 137–140.

Fellows, L., and Farah, M. (2007). The role of ventromedial prefrontal cortex in decision making. Cerebral Cortex 17(11), 2669–2674.

Foot, P. (1978). *Virtues and vices*. Berkeley: University of California Press.

Frank, R. (1988). *Passions within reason: the strategic role of the emotions*. New York: Norton.

Gazzaniga, M. and Le Doux, J. (1978). *The integrated mind.* New York: Plenum.

Gibbard, A. (1990). *Wise choices, apt feelings*. Cambridge, MA: Harvard University Press.

Gill, M. and Nichols, S. (2008). Sentimentalist pluralism: moral psychology and philosophical ethics. *Philosophical Issues,* 18, 143–163.

Glimcher, P. and Fehr, E. (eds.). (2013). *Neuroeconomics: Decision making and the brain*. Academic Press.

Graham, J., Nosek, B., Haidt, J., Iyer, R., Koleva, S., and Ditto, P. (2011). Mapping the moral domain*.* *Journal of Personality and Social Psychology* 101, 366.

Greene, J. (2007a). The secret joke of Kant’s soul. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 3: The Neuroscience of Morality* (pp. 35–79). Cambridge, MA: MIT Press.

Greene, J. (2007b). Reply to Mikhail and Timmons. In W. Sinnott-Armstrong (Ed.), *Moral Psychology,* *Volume 3: The Neuroscience of Morality* (pp. 105–117). Cambridge, MA: MIT Press.URL: http://www.wjh.harvard.edu/ jgreene/GreeneWJH/Greene-KantSoul.pdf

Greene, J. (2007c). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in cognitive sciences*, *11*(8), 322–323.

Greene, J. (2010). Notes on ‘The Normative Insignificance of Neuroscience’ by Selim Berker. Unpublished manuscript.

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

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

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

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

Greene, J. (2013). *Moral Tribes*. London: Atlantic Books.

Greene, J. (2014). Beyond point-and-shoot morality: why cognitive (neuro)science matters for ethics. *Ethics* 124(4), 695–726.

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

Haidt, J. (2013). *The Righteous Mind*. London: Penguin.

Haidt, J., and Joseph, C. (2004). Intuitive ethics. *Daedalus* 133(4), 55–66.

Haidt, J., Koller, S., and Dias, M. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, 65, 613–628.

Hume, D. (1978 [1739/40]). *A Treatise of Human Nautre*. Edited by L. A. Selby/Bigge and P. H. Nidditch. Oxford: Oxford University Press.

Hume, D. (1998 [1751]). *An Enquiry concerning the Principles of Morals*. Edited by T. L. Beauchamp. Oxford/New York: Oxford University Press.

Hutcherson, C., Montaser-Kouhsari, L., Woodward, J. and Rangel, A. (2015). Emotional and utilitarian appraisals of moral dilemmas are encoded in separate areas and integrated in ventromedial prefrontal cortex. *Journal of Neuroscience*, *35*(36), 12593-12605.

Hutcheson, F. (2004 [1725]).An Inquiry Into the Original of Our Ideas of Beauty and Virtue in Two Treatises. W. Leidhold (Ed.). Indianapolis: Liberty Fund.

Hutcheson, F. (2002 [1728]), An Essay on the Nature and Conduct of Passions, With Illustrations on the Moral Sense. A. Garrett (Ed.). Indianapolis: Liberty Fund.

Kahane, G. (2014): Intuitive and counterintuitive morality. In: D'Arms, J., and Jacobsen, D. (eds): *The Science of Ethics: Moral Psychology and Human Agency*, Oxford: Oxford University Press, pp. 9–39

Kahane, G., and Shackel, N. (2010). Methodological issues in the neuroscience of moral judgement. *Mind and Language*, *25*(5), 561–582.

Kahane, G., Wiech, K., Shackel, N., Farias, M., Savulescu, J. and Tracey, I. (2011): The neural basis of intuitive and counterintuitive moral judgment. *Social cognitive and affective neuroscience 7*(4): 393–402.

Kahane, G. (2011). Evolutionary debunking arguments. *Noûs*, *45*(1), 103–125.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kamm, F. (1999). Nonconsequentialism. In: H. LaFollette (Ed.), *The Blackwell Guide to Ethical Theory*. Oxford: Blackwell.

Kamm, F. (2009). Neuroscience and moral reasoning: a note on recent research. *Philosophy and Public Affairs*, 37(4), 330–345.

Kant, I. (1968). *Werke*. Berlin: De Gruyter. (Original work published in 1902).

Klein, C. (2010). Philosophical issues in neuroimaging. *Philosophy Compass*, 5(2), 186–198.

Klein, C. (2011). The dual track theory of moral decision-making: A critique of the neuroimaging evidence. *Neuroethics*, 4(2), 143–162.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V. and Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, *314*(5800), 829–832.

Koenigs, M. and Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: evidence from the Ultimatum Game. *Journal of Neuroscience*, *27*(4), 951–956.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M. and Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, *446*(7138), 908.

Liao, S. (2016). Morality and neuroscience. In: S. Liao (Ed.), *Moral Brains* (pp. 1–42). New York: Oxford University Press.

Liao, S (2011). Bias and Reasoning: Haidt's Theory of Moral Judgment. In: T. Brooks (Ed.), *New Waves in Ethics* (108–128). London: Palgrave Macmillan.

Liao, S., Wiegmann, A., Alexander, J. and Vong, G. (2012). Putting the trolley in order: Experimental philosophy and the loop case. *Philosophical Psychology*, *25*(5), 661–671.

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *The American Economic Review 90*(2), 426–432.

Louden, R. (1996). Towards a genealogy of 'deontology'. *Journal of the History of Philosophy*, 34, 571–592.

Mikhail, J. (2007). Moral cognition and computational theory. In W. Sinnott-Armstrong (Ed.), *Moral Psychology* (pp. 81–91). Cambridge, MA: MIT Press.

Mogensen, A. (2015). Evolutionary debunking arguments and the proximate/ultimate distinction. *Analysis*, *75*(2), 196–203.

Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F. and Grafman, J. (2005). Opinion: the neural basis of human moral cognition. *Nature Reviews Neuroscience*, *6*(10), 799.

Moll, J., Krueger, F., Zahn, R., Pardini, M., de Oliveira-Souza, R., and Grafman, J. (2006). Human fronto–mesolimbic networks guide decisions about charitable donation. *Proceedings of the National Academy of Sciences*, *103*(42), 15623–15628.

Moll, J., and de Oliveira-Souza, R. (2007). Response to Greene: Moral sentiments and reason: friends or foes?. *Trends in Cognitive Sciences*, *11*(8), 323–324.

Moll, J., Oliveira-Souza, R., Garrido, G., Bramati, I., Caparelli-Daquer, E., Paiva, M., Zahn, R. and Grafman, J. (2007). The self as a moral agent: linking the neural bases of social agency and moral sensitivity. *Social Neuroscience*, *2*(3-4), 336-352.

Nichols, S., and Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition*, 100, 530–542.

Nicolle, A. and Goel, V. (2013). What is the role of ventromedial prefrontal cortex in emotional influences on reason? In: I. Blanchette (Ed.), *Emotion and reasoning* (p. 154). Hove: Psychology Press*.*

Nordgren, L., and McDonnell, M. (2011). The scope-severity paradox: why doing more harm is judged to be less harmful. *Social Psychological and Personality Science*, 2, 97–102.

Olsson, A. and Phelps, E. (2004). Learned fear of “unseen” faces after Pavlovian, observational, and instructed fear. *Psychological Science*, *15*(12), 822-828.

Olsson, A. and Phelps, E. (2007). Social learning of fear. *Nature neuroscience*, *10*(9), 1095.

Paxton, J., Bruni, T., and Greene, J. (2013): Are ‘counter-intuitive’ deontological judgments really counter-intuitive? An empirical reply to Kahane et al. (2012). *Social cognitive and affective neuroscience* *9*(9):1368–71.

Prinz, J. (2007). *The emotional construction of morals*. New York: Oxford University Press.

Prinz, J., (2016). Sentimentalism and the moral brain. In M. Liao (Ed.), *Moral Brains: The Neuroscience of Morality* (pp. 45–73). New York: Oxford University Press.

Putnam, H. (1979). The meaning of ‘meaning’. In H. Putnam (Ed.), *Mind, Language, and Reality. Philosophical Papers, Volume 2*, Cambridge: Cambridge University Press. (Original work published in 1975).

Rozin, P., Lowery, L., Imada, S. and Haidt, J., (1999). The CAD triad hypothesis: a mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of personality and social psychology*, *76*(4), 574.

Sanfey, A., Rilling, J., Aronson, J., Nystrom, L., and Cohen, J. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, *300*(5626), 1755–1758.

Schelling, T. (1968). The life you save may be your own. In S. Chase (Ed.), *Problems in Public Expenditure Analysis* (pp. 127–176). Washington: Brookings Institute.

Shaftesbury, Third Earl of (2001 [1714]). An Inquiry Into Virtue and Merit. In D. Den Uyl (Ed.). *Characteristicks of Men, Manners, Opinions, Times, Vol.2* (pp. 1–100). Indianapolis: Liberty Fund.

Shenhav, A. and Greene, J. (2014). Integrative moral judgment: dissociating the roles of the amygdala and ventromedial prefrontal cortex. *Journal of Neuroscience*, *34*(13), 4741–4749.

Shoemaker, D., 2007. Moral address, moral responsibility, and the boundaries of the moral community. *Ethics*, *118*(1), 70–108.

Shoemaker, D., 2011. Attributability, answerability, and accountability: Toward a wider theory of moral responsibility. *Ethics*, *121*(3), 602–632.

Shweder, R., Much, N., Mahapatra, M. and Park, L. (997). The ‘‘big three’’of morality (autonomy, community and divinity) and the ‘‘big three’’explanations of suffering. In A. Brandt and P. Rozin (Eds.), 2013. *Morality and Health* (pp. 119-169). Oxford: Routledge.

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

Small, D., and Loewenstein, G. (2003). Helping a victim or helping the victim: altruism and identiﬁability. *The Journal of Risk and Uncertainty*, 26, 5–16.

Stanovich, K., and West, R. (2000). Individual differences in reasoning: implications for the rationality debate. *Behavioral and Brain Sciences*, 23, 645–65.

Strawson, P. (2008 [1960]). Freedom and resentment. In his *Freedom and Resentment and Other Essays* (pp. 1–28). London: Routledge.

Street, S. (2006). A Darwinian dilemma for realist theories of value. *Philosophical Studies*, *127*(1), 109–166.

Thomson, J. (1985). The trolley problem. *The Yale Law Journal*, *94*(6), 1395-1415.

Unger, P. (1996). *Living High and Letting Die: Our Illusion of Innocence*. New York: Oxford University Press.

Wheatley, T. and Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological science*, *16*(10), 780–784.

Young, L., Koenigs, M., Kruepke, M. and Newman, J. (2012). Psychopathy increases perceived moral permissibility of accidents. *Journal of abnormal psychology*, *121*(3), 659.

1. ### **Notes**

 [↑](#endnote-ref-1)
2. The switch case was first introduced by Philippa Foot (1978), the footbridge case by Judith Jarvis Thomson (1985). [↑](#endnote-ref-2)
3. Functional magnetic resonance imaging (fMRI) is a species of scientific neuroimaging. It implicitly assumes that active brain regions consume a comparatively high amount of oxygen, causing a change in blood flow, which in turn can be visualised through magnetic fields. [↑](#endnote-ref-3)
4. Greene's initial hypothesis was that the two classes of dilemmas could be distinguished by whether they were ‘personal’ or ‘impersonal’. A personal dilemma leads to serious bodily harm of a particular person in such a way that it is not a result of deflecting a threat onto a third party (Greene and Haidt 2002, p. 519). An impersonal dilemma does not involve such a harm. However, the personal-impersonal distinction has been criticized by various authors (e.g., Nichols and Mallon 2006, Mikhail 2007) and finally been questioned by Greene himself (Greene 2007b, p. 108). [↑](#endnote-ref-4)
5. Ironically, the utilitarian Jeremy Bentham seems to have been one of the first philosophers to use ‘deontology’ (Louden 1996, pp. 573–579). He applied it far more broadly than we do nowadays, sometimes as a synonym for ‘ethics’, sometimes as a synonym for ‘utilitarianism’. [↑](#endnote-ref-5)
6. Greene would claim that, by targeting the former, he also refutes the latter. Whether this claim is correct is a question that has been addressed elsewhere. [↑](#endnote-ref-6)
7. This argument is most developed in Greene 2007a, an earlier paper to which Greene refers in his book (2013, p. 381) for further detail. [↑](#endnote-ref-7)
8. I thank an anonymous reviewer for clarificatory advice. [↑](#endnote-ref-8)
9. I thank an anonymous reviewer for Synthese for helping me see this implication. [↑](#endnote-ref-9)
10. A view on which the strength to which emotions influence a moral judgment determines its normative status would also have to explain how relativity to culture or personal experience should determine this influence on moral judgment. Thanks to an anonymous reviewer for this point. [↑](#endnote-ref-10)
11. There are others but I do not have the space to discuss them here. For instance, Nichols and Mallon (2006) take the view that moral decision-making is determined by rule-based assessments, regardless of whether these rules concern consequentialist or deontological considerations. Rules are supplemented by emotion and cost-benefit analysis. [↑](#endnote-ref-11)
12. I thank an anonymous reviewer for Synthese for bringing the relevance of this literature to my attention. [↑](#endnote-ref-12)
13. Strawson laments that “it is a pity that talk of the moral sentiments has fallen out of favour” (1980 [1960], p. 26), as he regards it as quite apt to describe reactive attitudes. [↑](#endnote-ref-13)
14. I thank an anonymous reviewer for suggesting this objection and reply. [↑](#endnote-ref-14)
15. This paper has greatly benefitted from discussions with Gunnar Björnsson, Josh Greene, Richard Holton, Gina Rini, and audiences in Riga, Granada, and Mainz, as well as from a debate between Rae Langton and Joshua Greene in Cambridge. I also thank two anonymous reviewers for their extremely helpful feedback, which has to a great degree shaped and improved this paper. [↑](#endnote-ref-15)