DOI:
10.1007/s12152-013-9191-y

Document Version
Peer reviewed version

THE SCIENCE OF MORALITY AND ITS NORMATIVE IMPLICATIONS

Authors: Tommaso Bruni, Matteo Mameli (corresponding author), Regina A. Rini

Abstract: Neuromoral theorists are those who claim that a scientific understanding of moral judgment through the methods of psychology, neuroscience and related disciplines can have normative implications and can be used to improve the human ability to make moral judgments. We consider three neuromoral theories: one suggested by Gazzaniga, one put forward by Gigerenzer, and one developed by Greene. By contrasting these theories we reveal some of the fundamental issues that neuromoral theories in general have to address. One important issue concerns whether the normative claims that neuromoral theorists would like to make are to be understood in moral terms or in non-moral terms. We argue that, on either a moral or a non-moral interpretation of these claims, neuromoral theories face serious problems. Therefore, neither the moral nor the non-moral reading of the normative claims makes them philosophically viable.

Keywords: Morality, Normativity, Ethics, Neuroscience, Psychology

THE MORAL JUDGMENT MACHINERY AND NEUROMORAL THEORIES

The Moral Judgment Machinery—MOJUM for short—is the neuropsychological machinery underpinning the *formation* (the mental tokening) and the *processing* of moral judgments. Here 'moral judgment' is taken to refer to a kind of mental state rather than a kind of linguistic utterance. 'Processing' comprises all the ways in which moral judgments are used within the mind *after* their formation, including the production of overt behavior. MOJUM is being studied by an increasingly large number of psychologists, neuroscientists, and researchers in related disciplines, such as psychiatry, behavioral economics, etc.

We will call 'neuromoral theories' all those theories according to which the science of MOJUM can be used to improve or enhance human moral thinking. The literature displays a range of attempts to justify drawing normative conclusions from empirical premises. In this paper we focus on an aspect of these attempts that has received relatively little attention. It concerns the *normative grounding* of these conclusions. This normative grounding can be either moral or non-moral in nature. If a moral normative grounding is adopted, then certain forms of moral thinking are to be recommended because they are morally better, where 'morally better' appeals to distinctively moral concerns, rather than instrumental contributions to contingently-shared goals. We can call this the 'meta-moral' view. On the other hand, it might be that the grounding of the normative claims put forth by neuromoral theories is *not* moral. The most plausible and natural interpretation of a non-moral grounding is what we call the 'collective usefulness' view. According to this view, certain forms of moral thinking are to be recommended because they serve instrumentally to further widely shared goals, such as a reduction in conflict, or an increase in social cohesion.[1]

Our main aim in this paper is to show that the neuromoral theories we examine face a common dilemma concerning the interpretation of their normative claims. These claims can be understood *either* in moral *or* in non-moral terms. We argue that on either interpretation, defending these normative claims will be difficult.

To our purposes, it is important to underscore that neuromoral theorists can have different views about the following questions:

> ***(i)*** What does it mean to say that the human ability to make good moral judgments can be *improved*?

[1] This view will be highly objectionable to certain moral theorists, particularly those of a Kantian orientation [1]. The view appears to entail that moral reasoning can be treated as an instrumental means to some other, non-moral, end. Moreover, there is a possible view on which 'collective usefulness' and 'meta-moral' criteria will coincide, such as if one believes that (some of) the moral truths just *are* whatever best serves a particular set of social aims. Views in this vicinity are defended by Boyd [2] and Copp [3].

(ii) How exactly can the scientific understanding of MOJUM have a positive impact on this ability? How exactly can science help us make good moral judgments?

Consider a neuromoral theorist who is also a moral realist. She will claim that the ability to form good moral beliefs consists (at least partly) in the ability to acquire moral beliefs that are *true*, that is, moral beliefs that accurately represent the moral facts. On views of this kind, improving moral thinking through the scientific study of MOJUM consists in using the science of MOJUM for improving the human ability to form *true* moral beliefs. An account of this sort seems to fit well with the meta-moral view: it is the moral facts themselves, together with facts about the reliability of different mental processes at representing such facts, that give us the normative standard by which we can judge different forms of moral reasoning.

A neuromoral theorist need not be a moral realist though. She may believe that there are no moral facts: she may be an error theorist. An error theorist *can* make a distinction between good moral beliefs and bad ones, and thereby can make sense of the idea that our ability to form good moral judgments (and to avoid forming bad ones) can be improved through the scientific study of MOJUM. What she *cannot* do is appeal to truth or accuracy in drawing this distinction, since she is committed to the view that all moral beliefs are false. How can an error theorist make a distinction between good and bad moral beliefs? Here is one way. Some error theorists are fictionalists, in that they maintain that moral beliefs are fictions. On this view, moral beliefs, while false, can be useful fictions. One could argue, for example, that moral beliefs are useful fictions when they help us engage in rewarding forms of cooperation [4, 5]. An error theorist of this sort can argue that good moral beliefs are those whose features make them useful at engaging in rewarding forms of cooperation and that, in contrast, bad moral beliefs are those that do not have such features. For a neuromoral theorist of this kind, improving our ability to make good moral judgments consists in using the science of MOJUM to improve our ability to form beliefs with the right kind of cooperation-enhancing features. *Mutatis mutandis* for error theorists who distinguish between good and bad moral judgments by appealing to things other than the cooperation-enhancing features of moral beliefs. Accounts of this sort fit well with non-moral forms of normative grounding and in particular with the collective usefulness view: the idea would be that moral beliefs are useful fictions when they have some features—e.g. they help us engage in useful forms of cooperation—that further widely shared goals.

In what follows, we remain neutral on the questions that have to do with meta-ethical disputes, though many of the issues we will discuss intersect in very complex ways with meta-ethical issues. However, the existence of the options we have just mentioned is useful to show that conflicting attempts at deriving normative conclusions from empirical evidence about MOJUM face a common problem. When a neuromoral theorist claims that an improved understanding of MOJUM enables humans to make *better* moral judgments, what kind of normative standard is this "better" referring to? Neuromoral theories in general have to answer this question. We claim that this question gives rise to a dilemma, which we will call the 'normative grounding dilemma'. As already noted, there are either moral or non-moral interpretations of this grounding. The moral interpretation coincides with the 'meta-moral view'. 'Collective usefulness' is the most plausible way of fleshing out the non-moral option, as we will see from the authors discussed below. According to the normative grounding dilemma, as we conceive of it, neuromoral theories incur problems either on the moral or on the non-moral reading of the normative grounding. On the one hand, as we will see, adopting the meta-moral view poses problems of circularity and, more generally, of justificatory grounds for neuromoral accounts. On the other hand, adopting the 'collective usefulness' view requires that some specific and, in some cases, hard-to-find empirical evidence be collected. We will argue that neuromoral theories need to tackle these difficulties if they are to become persuasive philosophical theories.

We will show how the 'normative grounding dilemma' affects three neuromoral theories: one suggested by Michael Gazzaniga, one put forward by Gerd Gigerenzer, and one developed by Joshua Greene. We do not aim at being exhaustive, but we intend these illustrations to show that an explicit treatment of this issue is necessary in order to make progress in this area more generally. We will also discuss these three proposals in relation to (ii)—how an understanding of MOJUM can help humans form good moral judgments. We believe these three proposals are interesting partly because of what they say about the role that conscious reflection ought or ought not to play in attempts to enhance our moral-belief-formation processes. As we will see, Greene and Gazzaniga think, though for very different reasons, that conscious reflection is a praiseworthy way to tackle moral problems, while Gigerenzer thinks that conscious reflection ought not to be used when trying to solve difficult moral

problems. So, these three proposals attribute very different roles to conscious thought and reflection and thereby represent a diverse sample of neuromoral theories.

MAKING IMPLICIT MORALITY EXPLICIT

Gazzaniga's View: Descriptive and Normative Claims

One kind of neuromoral theory is suggested by Michael Gazzaniga in the following passage:

> I would like to support the idea that there could be a universal set of biological responses to moral dilemmas, a sort of ethics, built into our brains. My hope is that we soon may be able to uncover those ethics, identify them, and to begin to live more fully by them. I believe we live by them largely unconsciously now, but that a lot of suffering, war, and conflict could be eliminated if we could agree to live by them more consciously. [6, p. xix]

According to this proposal, there is a universal hardwired moral competence, similar to the Chomskian linguistic competence (Universal Grammar, UG). Let us call this moral competence Universal Morality (UM). Once the implicit principles that constitute UM are identified and explicitly described through psychological investigation, people will be able to deploy such principles in conscious deliberation whenever they need to deal with a moral problem, thereby enhancing their ability to token good moral judgments.

On this view, the implicit principles of UM are in general reliable and trustworthy. Nevertheless, people sometimes end up forming defective moral judgments and behaving against these implicit principles. UM does not always manage to affect our decision-making processes and our behavior in the right way. If and when explicit knowledge of the principles of UM is acquired, human beings will be able to start following these principles *deliberately* and, in this way, they will be able to ensure—or at least to increase the probability—that the moral judgments they form are good ones. It will not be necessary to perform this conscious application of the implicit principles all the time, but only in some difficult cases, those in which there are good reasons for believing that unconscious moral thinking is unlikely to be the expression of the correct operation of UM.[2] Hence, the explicit application of UM would allow us to make *better* moral judgments, more reliable and trustworthy than they would otherwise be. Since conscious reflection is likely to make the explicit application of UM easier, conscious reflection is then a good way to perfect our moral life.

Problems

On this account, humans are able to form moral judgments in at least two ways: via the unconscious operation of UM and via a conscious and deliberate application of principles. One issue that a neuromoral theorist of this kind needs to address is that of explaining what he means when he says that people *ought to* consciously deploy the principles of UM. As said above, we label this the 'normative grounding problem'.

We distinguished 'meta-moral' and 'collective usefulness' responses to the normative grounding problem. We can illustrate both here. If the meta-moral approach is taken, the claim "we ought to consciously apply the principles of UM" turns out to express a moral judgment and the "ought" is a moral one. This response constitutes the first horn of the 'normative grounding dilemma' and it seems problematic, since it is unclear whether this moral judgment is actually true and what strategy a neuromoral theorist of this sort can use to justify it. Suppose this theorist suggests that our grasp of the normative moral standards derives from UM itself. This would mean that the only way we have to morally justify the moral claim "we ought to consciously apply UM's principles" is by appeal to the principles of UM. On this view, it is by appealing to the principles of UM that we can justify the claim that, in order to make reliable first-order moral judgments, we should make explicit use of the

[2] A possible view, for example, is that in some contexts people make many *performance errors* when thinking about moral matters, and that an understanding of UM will help us to avoid such errors. Notice that the proponent of this view would have to find a principled way to make a competence/performance distinction in the context of morality, which may turn out to be a difficult task.

principles of UM, at least in some cases. A strategy like this could make the justification of Gazzaniga's normative claim problematically circular. How can we, from such a principle, derive a justification of the claim that the principles ought to be deliberately deployed? What if, for example, the principles of UM were unreliable in the first place? The neuromoral theorist would have to find a way to show that the circularity is virtuous rather than vicious, and it is not clear, again, what strategies she could deploy to achieve this aim. We think that the upholder of the meta-moral view owes an answer to the circularity problem.

Another possible response to the 'normative grounding problem' is that the grounding comes from non-moral facts, and in particular from facts about collective usefulness. This constitutes the second horn of the 'normative grounding dilemma' described above. The claim would be that the deliberate deployment of the principles of UM is the best way people have—either as individuals or as collectives—to secure some commonly desired outcomes, such as the reduction of suffering and the avoidance of conflict in the world. But if this is the right way of interpreting the 'ought', one could ask whether the non-moral ought statement is actually true. Let us grant, for the sake of argument, that there is such a thing as a UM. Is it true, as Gazzaniga suggests, that the conscious deployment of UM is the best way available to humans to avoid conflict, the suffering that conflict generates, and to pursue similar goals? What are the reasons for thinking that UM—which presumably evolved in response to environmental pressures that are very different from current human environments—does not contain judgment-forming principles that in contemporary societies are likely to lead, at least in some circumstances, to conflict, suffering and other unpleasant things that people would prefer to avoid? The hypothesis that UM is *not* the best way available to humans to increase 'collective usefulness' cannot be excluded *a priori*. In order to exclude this hypothesis, Gazzaniga would have to marshal empirical evidence about how human societies work in different historical and cultural contexts and about the effects that the conscious and unconscious deployment of the principles of UM have in such contexts. However, this kind of empirical evidence is not easy to gather. In order for the 'collective usefulness' view to be a convincing solution for the 'normative grounding problem' this evidence needs to be provided.

To conclude, whether Gazzaniga's normative claim is interpreted in moral terms or in non-moral terms, its justification seems to be problematic. Other scholars have developed and explored the hypothesis that there is a universal moral competence more in depth than has Gazzaniga [7–9]. There is no room for analyzing such views here (cf. Sterelny [10] for a discussion). Nevertheless, the passage quoted above points towards an interesting option available to those interested in the view that the scientific understanding of MOJUM can have normative implications. Our claim is simply that any neuromoral theorist who wants to appeal to a universal moral competence in order to make normative claims similar to those made by Gazzaniga would have to face the 'normative grounding dilemma'.

Simple Heuristics that Make Us Moral

Gigerenzer's View: Descriptive and Normative Claims

According to a different view, the science of MOJUM can have a positive impact on people's ability to form moral judgments not because it will allow them to consciously apply some principles that have so far been sub-conscious, but rather because it will make people understand that they ought *not* to rely on conscious deliberation when they make moral decisions. One author who adopts a theory of this kind is Gerd Gigerenzer [11, 12].

Gigerenzer's focus is on fast and frugal heuristics. These are unconscious strategies for fixing belief and for making decisions, designed to solve problems in a rapid way and with the use of only small amounts of incomplete information. According to Gigerenzer, the heuristics people deploy to solve moral problems and to form moral judgments are not specific for thinking only about moral issues. Rather, they are general-purpose tools whose deployment comes natural when one thinks about moral as well as non-moral problems. Gigerenzer, developing some of Herbert Simon's [13,14] ideas on 'bounded rationality' and on 'satisficing' strategies (as opposed to maximizing ones), has elaborated an account of the role that heuristics play in making human cognition successful. On this view, simple heuristics—rather than the informationally rich processes of conscious deliberation and maximization—are the things "that make us smart" [15–17].

Gigerenzer [11, p. 3] claims that the very same heuristics can in different circumstances result in morally good or morally bad decisions. For example, a simple rule such as "if there is a default option, do nothing about it (i.e. tacitly accept the default)" may result in morally good or morally bad decisions, depending on what the default option actually is. In one of Gigerenzer's own examples, an opt-in default for organ donation will result in most people's tacit acceptance of the default option, that is, a (perhaps unconscious) decision in favor of organ donation [cf. 18]. The same applies to heuristics such as "do what the majority of the members of your group do." In a society in which most people behave in racist ways, this heuristic will result in racist behavior.

One way in which, on this account, the science of MOJUM can help people make good moral judgments is the following. A proper understanding of the mind and of its heuristics 'toolbox' will help one see that many of the current theories of decision making—including many theories of *moral* decision making—are wrong. These theories, especially those that appeal to maximizing strategies, are wrong both in that they do not accurately represent the way people make decisions and, as a consequence, in that they give bad advice about the proper way of making decisions. An understanding of the heuristics toolbox will help one see that one *ought not* to follow such advice when trying to solve decision-making problems, including *moral*decision-making problems. One ought to rely on the heuristics toolbox instead. Trying, via conscious deliberation, to counteract or bypass the outputs generated by the heuristics toolbox is in general a bad idea.

Gigerenzer claims that the goodness of various types of consequences (health, income, freedom, etc.) is incommensurable, and that in any case working out immensely complex causal pathways is computationally intractable. Moreover, citing Lipsey and Lancaster [19], he dismisses a strategy of 'approximating' maximization: in general, trying to approximate what might work in ideal conditions—including ideal computational conditions—will not lead to the best outcomes in actual conditions. Therefore, insofar as conscious reflection and its informationally rich processes are an attempt to approximate maximization, they are going to be counterproductive. In the real world simple heuristics outperform maximizing strategies and, more generally, informationally rich strategies. This also applies to moral judgments: those produced by the heuristics toolbox are more trustworthy than those produced by informationally rich processes. Hence, Gigerenzer argues that doing the kinds of maximizing computations that some utilitarians recommend is bad advice. One way in which a proper scientific understanding of MOJUM can help us make good moral judgments is by helping us avoid being led astray by bad theories.

Some of Gigerenzer's writings suggest also another way, which will be relevant in our discussion below, in which an understanding of the heuristics toolbox can have a positive impact on our moral-judgment-forming ability. As mentioned above, often the quality (good or bad) of the output of the heuristics toolbox depends on the context in which the toolbox is used. Because of this, knowledge of the toolbox and of its interactions with the range of decision-making environments could be used to increase the chances that people will find themselves in contexts in which the heuristics they are going to use will result in good judgments [12, 20]. Knowledge of the heuristics toolbox can be used in this way both by individual agents and by policy makers who have some control on the contexts in which certain populations make choices (cf. the notion of 'choice architect' in [21]). Notice that the focus here is on shaping the decision-making *context* rather than on changing the heuristics themselves, whose deployment is (supposedly) triggered automatically and may be difficult to avoid.

Problems

Let us consider Gigerenzer's critique of maximizing utilitarianism first. In what sense, on his view, is the advice given by the maximizing utilitarian *not* to be followed? What is the normative standard by which one should evaluate the maximizing utilitarian's recommendations? Are those recommendations bad from a moral viewpoint or from a non-moral viewpoint?

One thing to notice is that it is open to the maximizing utilitarian to recommend making use of Gigerenzer's heuristics. According to maximizing act utilitarianism, an action is morally right if it maximizes aggregate utility. Since trying to make evaluations and comparisons that are too difficult for creatures like us is certainly *not* conducive to the maximization of utility, a maximizing utilitarian of this sort can use Gigerenzer's own arguments and recommend not doing all the calculations that

Gigerenzer also recommends not doing.[3] In particular cases, this utilitarian can also recommend using the heuristics that Gigerenzer describes, at least if some degree of control on the heuristics one deploys is possible—whether it is direct or indirect control, e.g. via environmental manipulations. Furthermore, on a standard reading of maximizing rule utilitarianism, an action is morally right if it is produced by decision-making procedures whose deployment—in organisms like us, in a world like ours—tends, at least in the long term, to maximize aggregate utility. Again, this kind of utilitarianism is compatible with Gigerenzer's focus on heuristics. If the heuristics turn out to be the decision-making procedures whose deployment in general tends to maximize aggregate utility, then a utilitarian of this sort can recommend using them, or can recommend not trying to sabotage them.

Gigerenzer makes the following comment on the view that maximization can be seen as a normative standard (as opposed to a decision procedure):

> I must admit that I fail to understand the logic. […] Even if someone were to stumble over the best action by accident, we would not recognize it as such and be able to prove that it is indeed the best. How can maximization serve as a norm for rightness if we can neither determine nor, after the fact, recognize the best action?[4] [25, p. 44]

There are many complex issues here that we do not want to enter into. But if Gigerenzer is right about maximization being computationally out-of-reach for creatures like us and in a world like ours, then perhaps a maximizing utilitarian conception of the normative standard can be accepted only by accepting the claim that humans can never have knowledge of moral facts, because—due to their computational limitations—they can never have knowledge of what action or what decision-making procedure is the one that maximizes utility. This seems to be a hard bullet for the maximizing utilitarian to bite, and so such a utilitarian would presumably want to argue that maximization is *not* in the relevant sense computationally out-of reach for organisms like us, at least not in general and not systematically. Let us concede, at least for the sake of argument, that Gigerenzer's claims, when properly reconstructed, *can* be seen as a critique of the maximizing utilitarian's account (or accounts) of the normative standard for actions and decision-making procedures. Even so, an alternative account of the nature of such a normative standard needs to be provided. But this leads Gigerenzer into the 'normative grounding dilemma'. As in the preceding discussion of Gazzaniga, we can consider two solutions: one moral (i.e. the meta-moral view), one non-moral (i.e. the collectively usefulness view).

One option is that when Gigerenzer makes claims of the sort "you ought to rely on the heuristics toolbox rather than on informationally rich conscious deliberation", the 'oughts' are moral ones. Gigerenzer's normative claims would then express moral judgments about moral-judgment-formation processes, but these second-order moral judgments need to be justified. What resources could one use to justify them? Appealing to the heuristics toolbox in this context would seem to be circular. We have noted above that heuristics do not necessarily lead to results that are morally commendable, since they do so just in the presence of the right kind of environment. Therefore, heuristics seem to possess little or no justificatory power on their own. Hence, the meta-moral option leaves Gigerenzer's normative claim ungrounded. These issues are key to the success of Gigerenzer's neuromoral theory, but they are not issues that Gigerenzer explores.

Another option, which represents the second horn of the 'normative grounding dilemma', is that the 'oughts' in Gigerenzer's normative claims refer to some non-moral standard such as 'collective usefulness'. The view being proposed then would be that relying on the heuristics toolbox is the best way people have—either as individuals or collectives—to secure certain important desired outcomes, such as peaceful relations, stable and rewarding cooperation, etc. However, this option is also

---

[3] Mill [22, p. 69] himself addressed this classic objection: "[…] [D]efenders of utility often find themselves called upon to reply to such objections as this—that there is not time, previous to action, for calculating and weighing the effects of any line of conduct on the general happiness. This is exactly as if any one were to say that it is impossible to guide our conduct by Christianity, because there is not time, on every occasion on which anything has to be done, to read through the Old and New Testaments."

[4] This is an example of what Shelly Kagan [23, p. 64] calls "the most common objection to consequentialism"; see Lenman [24] for an influential recent development. However, according to Kagan, the objection generalizes to all moral theories, since any plausible moral theory must allow some role for consequences. And the problem is not peculiar to maximizing theories, since any consequences will causally echo into the unknowable future [23, pp. 64–69].

problematic and the second horn of the 'normative grounding dilemma' does not fare better than the first, as we will now show.

Gigerenzer refers to what he calls "ecological rationality." Heuristics are ecologically rational in the sense that they are the best or the only way to achieve "success." Success is measured in terms of "accuracy, frugality and speed of decisions" and involves "means suited to certain goals" [26, p. 13]. But what are these goals? Gigerenzer suggests that some important goals for judging the ecological rationality of heuristics are desire satisfaction and the fixation of true beliefs [26, *ibid*.]. He seems to be claiming that we ought to rely on the heuristics toolbox in the sense that doing so is the best way we have to satisfy our individual or collective desires, including those that concern or involve attaining true beliefs. In an article on 'moral satisficing', he writes:

> My aim is not to provide a normative theory that tells us how we ought to behave, but a descriptive theory with prescriptive consequences, such as how to design environments that help people to reach their own goals. [12, p. 530]

This statement is somewhat puzzling, in that a prescriptive theory is normally understood as one that tells people what they ought to do. We believe that Gigerenzer is saying the following. People can select certain goals, acquire certain desires. Some of these goals are labeled "moral goals" and involve certain kinds of behaviors and outcomes. The labeling in Gigerenzer's view is not very important: it is just a way of marking certain desires as particularly significant, and the label can be applied differently in different cultures [12, p. 543]. Once a moral goal has been selected, one has to identify the best means of realizing it. The distinction Gigerenzer makes between normative and prescriptive is the distinction between a theory that tells you how to select your moral goals and a theory that tells you what to do in order to realize a moral goal you have selected.

It is certainly possible to question Gigerenzer's distinction between normative and prescriptive and the way he deals—or rather avoids dealing—with what he classifies as normative issues in the narrow sense. We leave this for another occasion. We want to notice though that once Gigerenzer's 'oughts' are interpreted as non-moral 'oughts' of collective usefulness, it becomes clear that heuristics are not necessarily the best way to increase 'collective usefulness' itself. Despite what Gigerenzer says about the general reliability of the heuristics toolbox and the general trustworthiness of its products, he does not provide specific empirical evidence to the effect that using the toolbox is the best way we have, across contexts, to avoid conflicts, ensure fruitful cooperation, etc. As we have already noted in Gazzaniga's case, this evidence would not be easy to marshal. By contrast, certain goals—especially collective goals that are determined through complex negotiation and collective deliberation—seem to be such that we are more likely to realize them through slow and informationally rich conscious reflection [27]. For example, in order to decide which organ donation policy may engender the maximal benefit to ailing patients in need of transplantation, it might be necessary to carefully study people's decision-making processes so as to understand under which conditions they are more likely to donate their organs. This study seems to involve effortful, step-wise thinking. After all, Gigerenzer concedes that the success of heuristics is determined by context and surely informationally rich processes are needed in order to determine which heuristics are going to be successful in which contexts. More generally, informationally rich conscious reflection seems to be the best tool social engineers and choice architects have for affecting environments in ways that will lead people to decide and act—perhaps via the deployment of simple heuristics—in ways that are conducive to certain desired outcomes. That is, even if we were to grant that Gigerenzer is right at the level of everyday behavior and choice, this would not mean that he is also right at the level relevant to 'choice architects' [21], that is those who have to make decisions that will determine tendencies in people's choices and behaviors. This is the level at which policy making normally takes place. In order to evaluate and compare heuristics and the outcomes they generate in different possible choice environments, policy makers have to perform informationally rich analyses, some of which may well have to appeal to Gigerenzer's own research. So, Gigerenzer's attack on conscious moral reflection seems to have at best limited applicability. His normative claim ('we ought to rely on the heuristics toolbox instead of informationally rich strategies') seems to be problematic even when interpreted in the non-moral sense.

Automatic vs. Manual Morality

Greene's View: Descriptive Claims

Gazzaniga believes that we ought to use conscious reflection to identify the principles of universal morality and to use them explicitly. Gigerenzer in contrast believes that we ought to avoid using conscious and informationally rich processes to sabotage the workings of the heuristics toolbox. Joshua Greene agrees with Gazzaniga that conscious reflection is important for the trustworthiness of moral judgments, but his reasons for this view are very different from Gazzaniga's. Greene argues in favor of a dual-process theory of moral thinking [28–34]. On this view, there are two distinct psychological systems responsible for the formation of moral judgments. By analogy with cameras, Greene calls the two systems the "automatic mode" and the "manual mode."

The *automatic mode system* (or, in brief, automatic system) generates moral judgments on the basis of emotional reactions triggered by (the perception or the thought of performing) particular actions. Such emotional reactions occur automatically, quickly, by means of unconscious processing, even though the reactions themselves may be conscious. If the reaction is a negative one, the resulting moral judgment will also in general be negative: it will be a judgment that the action ought *not* to be performed. "Moral intuitions" is the name Greene gives to the emotion-based moral judgments generated by this system [cf. also 35].

By contrast, the *manual mode system* (or, in brief, manual system) generates moral judgments through slow, effortful, and conscious reflection. One way in which this can happen is the following. Different possible actions are considered and the consequences that such actions are likely to produce are identified and compared. If this comparison arrives at the conclusion that a specific action is likely to bring about bad consequences—or consequences that are worse than those brought about by the alternatives—then the person makes the judgment that the action ought *not* to be carried out. It is not just the outcome of the process that is conscious in this case, but also the process itself.

Some of Greene's research that is relevant here focuses on a distinction between actions that involve personal force and actions that do not:

> An agent applies personal force to another when the force that directly impacts the other is generated by the agent's muscles, as when one pushes another with one's hands or with a rigid object. Thus, applications of personal force, so defined, cannot be mediated by mechanisms that respond to the agent's muscular force by releasing or generating a different kind of force and applying it to the other person. [31, p. 365]

Consider actions that involve bringing about physical harm to another human being in an intentional way and by means of personal force, in the sense just defined. According to Greene's hypothesis, through the operation of the automatic system, such actions generate negative emotional reactions. As a result, a moral judgment against the permissibility of these actions is formed. In contrast, if one has to make a decision about a similar action that does *not* involve personal force, the automatic system is (at least in general and *ceteris paribus*) less likely to produce strong negative reactions. In such cases, the manual system may be able to exert its control over the tokening of the moral judgment and a judgment in favor of the permissibility of the action may be formed. There are various factors that, Greene argues, "push the buttons" of the automatic system. Personal force is only one of them. On this account, the automatic system is always "on" and cannot be deactivated. The manual system, in contrast, can only be activated through conscious effort.

Much of the experimental evidence that Greene brings in support of his views was obtained by Greene himself and his collaborators. These experiments focus on moral dilemmas in which subjects have to evaluate the moral permissibility or impermissibility of specific actions. Two of the dilemmas that play a central role are *switch* and *footbridge*. In *switch*, deciding to hit a switch will turn a runaway trolley away from a track where five people are located and onto a track where there is another individual. As a result of this, the individual located on this other track will be killed but the five people—who would otherwise have died—will be saved. In *footbridge*, one can push a person, a bystander, off a footbridge and into the path of a runaway trolley. This will result in the death of the bystander, but the lives of five individuals further down the track—who would otherwise have died—will be saved. When told about the *switch* scenario, most subjects say that, in a situation like that, hitting the switch would be the right thing to do, or at least that it would be permissible. This is, as Greene puts it, a 'characteristically consequentialist judgment', since it is easy to justify it in consequentialist terms and difficult to justify it in deontological terms. In contrast, when confronted with the *footbridge* scenario, most subjects say

that it would be wrong to push the man off the bridge. This is a 'characteristically deontological judgment', as it is easy to justify it in deontological terms and difficult to justify it in consequentialist terms. Why is there this difference in responses? The evaluated action in *footbridge* involves personal force and so, according to Greene's hypothesis, it activates the automatic system and generates strong negative reactions, which—in most subjects—lead to the formation of a judgment against the permissibility of the action, a characteristically deontological judgment. In contrast, the action evaluated in *switch* does not involve personal force and does not activate the automatic system. Hence, when reflectively considering the action, the subject can evaluate its consequences and conclude that, given the circumstances, it is permissible to kill one to save five.[5] That is, in the absence of strong emotional interference, the subject can form a characteristically consequentialist judgment.[6]

Brain areas independently associated with emotional responses—the middle frontal gyrus (MFG), the superior temporal sulcus (STS), and the angular gyrus—are more activated when the subject is asked to solve *footbridge*-like dilemmas, i.e. in contexts in which subjects tend to give deontological responses. In contrast, areas independently associated with higher cognition—the dorsolateral prefrontal cortex (DLPFC) and the anterior cingulate cortex (ACC)—are more activated when the subjects are asked to solve *switch*-like dilemmas, i.e. in contexts in which subjects tend to give consequentialist responses [28, 29]. According to Greene, the best explanation for the functional Magnetic Resonance Imaging (fMRI) results and the pattern of responses to moral dilemmas is the dual-process hypothesis summarized above.

Greene's psychological account is controversial and has been much discussed in the literature.[7] We will concentrate instead on Greene's claims about the normative implications of his dual-process account.

Greene's View: Normative Claims

According to Greene, the newly discovered neuro-psychological facts have normative implications and can be used to help people make better moral judgments. Greene wants to claim that the deliverances of the manual system are more trustworthy and ought to be preferred to the deliverances of the automatic system when, as in *footbridge* scenarios for example, there is a conflict between the two. These normative implications, he says, cannot be derived from the scientific claims *alone*. In order to arrive at them, one needs to combine the scientific claims with independent normative judgments. In spite of this, he says, the normative implications can be surprising and important [32–34].

A useful example that Greene discusses, though it is not related to his own empirical research, is the following. It has been discovered that decisions of capital punishment juries are sometimes affected by (often unconscious) racist biases. Consider the normative claim that capital juries *ought not* to be affected by such biases. The empirically supported claim *when combined with the normative claim just mentioned* implies that capital juries sometimes take decisions that are different from those they ought to take. It follows that such juries ought to try, or to try harder, to avoid being influenced by racist biases.

Similarly, because of the automatic system, the way people form moral judgments is affected by the presence or absence of personal force in the evaluated action. But this factor, says Greene, is morally

[5] The presence/absence of personal force explains *some* of the difference between *footbridge* and *switch*. The results of other moral dilemmas suggest that other factors (e.g. whether the harm is intentionally and actively produced) are relevant too and that there are interaction effects between these other factors and personal force [31].
[6] The 'undisturbed' operation of the manual system will not result in such moral judgment in all cases. For example, if subjects are asked to solve *footbridge* first and *switch* immediately afterwards, they will often give a deontological answer to both *footbridge* and *switch* [36]. This could be due to the fact that, having just given a deontological answer to *footbridge*, and being unsure about the existence of a moral difference between *footbridge* and *switch*, subjects—in order to maintain coherence—end up giving a deontological answer to *switch* too, despite the fact that the harm hypothesized in *switch* does not elicit a negative emotional response via the automatic system.
[7] Greene claims that there is much independent evidence in support of the dual-process theory of moral judgment, coming from the studies of psychopaths, individual differences, the way particular tasks interfere or enhance emotional reactions or reasoning abilities, etc. Cf. [34]. Greene's account has generated much controversy [37–45] and various alternative proposals have been made [7, 45–47].

irrelevant. That is, it is wrong—and thereby one ought not—to base one's decision about the moral permissibility of an action on this factor. This claim is, in Greene's views, uncontroversial:

> Were a friend to call you from a footbridge seeking moral advice, would you say, 'Well, that depends… Will you be pushing or using a switch?' If questions such as this […] are not on your list of relevant moral questions, then you, too, should say "no" [when asked whether moral judgments ought to be sensitive to factors like personal force]. [34, § IV]

The neuropsychological discovery that personal force "pushes the buttons" of the automatic system and the normative claim that personal force is morally irrelevant, when taken together, imply that people sometimes reach the wrong moral conclusions. That is, people make moral decisions that are different from those they ought to make. They ought to try, or to try harder, to avoid being influenced by factors like the presence/absence of personal force in the action they are evaluating.[8]

Given that personal force exerts its influence on moral decision-making via the operation of the automatic system, on this view, people ought in some circumstances to avoid relying on the automatic system in the formation of moral judgments. Since the automatic system cannot be deactivated, in order to avoid relying on it one ought to stop oneself from endorsing the emotion-based moral intuitions that such system generates. Insofar as the workings of the automatic system are driven by morally irrelevant factors, we ought not to trust such system when we make moral decisions, we ought to discard or ignore the intuitions it generates, and we ought instead to rely on the manual system.

There is a further normative claim that Greene wants to make. This is the claim that his discoveries about the dual-process workings of the moral mind favor consequentialist as opposed to deontological approaches to morality: one ought to be a consequentialist and not rely on deontological moral theories. What is the argument for this claim?

As already noted, characteristically deontological judgments are normally produced by the automatic system, while characteristically consequentialist judgments are normally produced by the manual system. Conscious reflection can generate characteristically consequentialist judgments by considering various actions and comparing their likely consequences. But conscious reflection can also be involved, in some circumstances, in the formation or at least the post-hoc justification of characteristically deontological judgments. This happens when conscious reflection is used to identify and formulate principles that systematize (or explain, or rationalize) the various—and, in some cases, *prima facie* conflicting—moral intuitions with which one finds oneself. The starting point of this process is the set of moral judgments generated by the automatic system, most of which are characteristically deontological. When one tries—through conscious reflection—to identify some principles that systematize such judgments, the principles will also in general be deontological. According to Greene, once the psychological details are uncovered, one realizes that what deontological philosophers normally do is nothing but systematizing moral intuitions in the way just described. These philosophers are in the "intuition chasing" business: they "out-folk the folk" and "rearrange our prejudices" [32, 34, cf. also 35].

Hence, Greene is *not* claiming that *all* kinds of conscious reflection are trustworthy. There are some forms of conscious deliberation that are actually subservient to the automatic system. These forms of conscious reflection ought not to be trusted. In other words, because of the way deontological principles are arrived at, and because of the role played by automatic-mode-generated intuitions, deontological principles should be mistrusted. One ought not to believe deontological theories because their construction and development is affected by morally irrelevant factors.

In contrast, consequentialist theories do not in any way rely on automatic-mode-generated moral intuitions. Greene, following Singer [49], suggests that *at most* consequentialist theories can be said to rely on "rational intuitions." Again following Singer, as an example of what such rational intuitions might be, Greene refers to Sidgwick's ethical axioms, which are intuitive in the sense that they are

[8] Consider another example. According to [34], some recent research suggests that moral permissibility judgments are affected—through the emotions generated by the automatic system—by how close one is to the person whose death is needed in order to save a greater number of lives. The closer one is, the less likely it is that one will judge the action permissible. Since, says Greene, in accordance for example with Peter Singer, distance is morally irrelevant, one ought to try to avoid being affected by distance when forming moral judgments. But see Kamm [48] for the view that distance is, at least in some circumstances, morally relevant.

supposed to be self-evident truths that can be accessed through philosophical reflection.[9] Thus, "consequentialism should get points for not chasing intuitions and […] some of its competitors should lose points for doing so" [34, cf. also 32].

Problems

Greene claims that people *ought not* to trust the automatic-mode system and the emotion-based deontological intuitions it produces, and that they *ought* to trust the manual-mode system and its consequentialist judgments instead. But both on a meta-moral interpretation of this claim and on a non-moral, 'collective usefulness' interpretation, serious problems arise, as we will show below. Thus, Greene's neuromoral theory faces the 'normative grounding dilemma' too.

To understand which of the two options Greene favors, consider this quote:

> [The psychological] experiments identify factors to which our judgments are sensitive. This information may be combined with independent normative assumptions concerning the kinds of things to which our judgments ought to be sensitive. This combination can lead us to new, substantive moral conclusions. In other words, scientific information can allow us to trade in difficult "ought" questions for easier "ought" questions, and thus advance ethics. [34, §IV]

On Greene's picture, it is wrong to let one's moral decisions be affected by personal force and similar factors. So, if we discover that these factors affect our moral judgments via the automatic system, we have to conclude that it is wrong to rely on such a system and the intuitions it generates. We also have to conclude that it is wrong to give characteristically deontological answers to moral dilemmas like *footbridge*.

Is this wrongness *moral* wrongness? Greene is not entirely explicit on this, but the analogy with capital juries suggests that he intends the 'oughts' to be understood in moral terms. After all, when one says that capital judgments ought not to be sensitive to race, the 'ought' is most plausibly interpreted as a moral one. It is *morally* wrong for a jury to make decisions about capital punishment on the basis of racist prejudices; and it is *morally* wrong for jurors to rely on their intuitions if such intuitions are (known to be) affected by unconscious racist biases. Similarly, it could be argued, it is *morally* wrong for a person to rely on their emotion-based moral intuitions if these intuitions are (known to be) affected by morally irrelevant factors. Saying that these factors are morally irrelevant seems equivalent to saying that the decisions they affect are likely to be *morally wrong* decisions. And, surely, it is morally wrong to (knowingly) make morally wrong decisions. After all, morally wrong decisions are likely to lead to morally wrong actions.[10]

Hence, Greene's claim is more naturally interpreted in a meta-moral way. If so, Greene's strategy seems to consist in using the empirical discoveries in order to discard some of our moral judgments while holding fixed—and with the help of—some other moral judgments. The moral judgments that get discarded are the deontological answers to *footbridge*-like dilemmas. The moral judgment that is held fixed is the claim that we ought not to base our judgments on factors such as the presence or absence of personal force. If this is indeed Greene's strategy, the moral judgments that are held fixed are second-order moral judgments, i.e. meta-moral judgments, moral judgments about moral judgments. More specifically, they are moral judgments about moral-judgment-formation processes. However, it is not clear why we should save these second-order moral judgments rather than the first-order moral judgments (such as the spontaneous deontological response to *footbridge*) with which they conflict. It does not seem that second-order judgments have any kind of general priority relative to first-order moral judgments.

---

[9] Here is the third of these axioms: "The good of any one individual is of no more importance, from the point of view (if I may say so) of the Universe, than the good of any other" [50, p. 382].

[10] It could be argued that to be wrong about moral facts is morally wrong, and that this moral wrongness is supplementary relative to the wrongness of a moral action that originates from such erroneous moral knowledge. Alternatively, it could be argued that errors in moral knowledge are morally wrong just because they lead to morally wrong actions. We do not want to enter this interesting and important issue here, since what concerns us at the moment is to show that to follow a morally irrelevant factor can be plausibly seen as an instance of moral wrongness (as opposed to other kinds of wrongness).

According to Greene, everyone agrees that one ought not to decide on the permissibility of actions on the basis of personal force, i.e. everyone agrees that personal force is a morally irrelevant factor. However, we doubt that everyone does agree. Empirical evidence should be provided to buttress this claim. When our moral judgments are discovered to be in conflict with one another, perhaps via the acquisition of some empirical information, we need decide which ones to hold fixed and which ones to revise. In the case under consideration, one possible decision would be this: save the deontological answers to *footbridge* and the like and give up the second-order moral judgment that personal force is morally irrelevant. Greene may be right that many would find this option unpalatable, but many people find the discarding of deontological answers unpalatable too. The claim that most people would prefer to revise their answers to dilemmas like *footbridge* than to revise their (alleged) views on the moral irrelevance of personal force needs to be empirically supported.

However, let us suppose that Greene is right on the existence of this preference. The fact that most people have this preference would not show that the preference is justified, that it accurately tracks anything normatively significant. It may be the case that the agreement is due to widely shared but misleading intuitions rather than anything more substantive. Greene could try to show that such principles are self-justifying, or they are ultimately derivable from self-justifying principles, or that they are justified because they are supported by first order intuitions, that is, first-order intuitions other than those generated in footbridge-like dilemmas. Whichever strategy Greene chooses, a lot of work needs to be done to show that the strategy has some chance of success.

Given his reference to Sidgwick, Greene would perhaps like to argue that it is possible to explain and vindicate the intuition that it is better to keep the belief that personal force is morally irrelevant rather than the deontological answers to *footbridge* in terms of rational intuition. On this view, the preference in question would be the result of a process of reliable rational intuition rather than the result of unreliable emotion-based intuition. But unless we are told more about the nature of rational intuition and the way it is supposed to work, it is hard to see whether this line of argument has any chances of success.

Hence, it is not evident that Greene's theory possesses the resources to defend a meta-moral interpretation of its normative claims, since it is not clear why we should save the moral irrelevance of personal force and similar factors rather than widespread first-order moral judgments about cases such as *footbridge*. However, although it is more natural to interpret Greene's claim along meta-moral lines, it is possible to abandon the meta-moral interpretation and adopt the non-moral 'collective usefulness' view instead. But the truth of the claim that one ought, from the collective usefulness standpoint, to avoid using the automatic system and to adopt a consequentialist moral theory is an empirical issue, and a difficult one to investigate.[11] Even if this interpretation avoids many of the pitfalls to which the meta-moral interpretation can lead, it can be supported only through a very ambitious empirical research program showing that, across different cultures and epochs, relying on the manual mode system is more helpful for securing stable, fruitful cooperation and avoiding conflict than following the suggestions of the automatic mode system. Without this empirical backing, a 'collective usefulness' reading of Greene's normative claim remains unconvincing.

CONCLUSIONS

We have considered three different neuromoral theories, that is, three different theories according to which an accurate scientific account of the mental formation and processing of moral judgments has normative implications and can be used to increase the likelihood that good moral judgments are made. These theories provide three different accounts of the cognitive architecture of moral judgment and of the role that conscious thinking has or can have in the production of moral judgments. They also differ in the way they classify the various moral-judgment-formation processes as more or less trustworthy, in some cases as a result of the role that conscious thinking plays or does not play in such processes. Although these theories differ in their normative assessments of these processes and of moral judgments considered as the outcome of these processes, we have identified a recurring problem, the normative grounding problem. All of these theories need to specify according to which normative standard improved cognitive science may enable us to make *better* moral judgments. We have argued

[11] Some of Gigerenzer's worries about the drawbacks of informationally rich thinking, at least in the context of everyday cognition, may be relevant here.

that this normative standard can be specified in two ways: the meta-moral way and the non-moral way, which can be most plausibly understood in terms of collective usefulness. We have also argued that both solutions are deeply problematic. The meta-moral interpretation seems to make the neuromoral normative claims conceptually difficult to justify, but it also opens important empirical issues, such as in Greene's case. The 'collective usefulness' interpretation is conceptually simpler and in some ways more plausible, but requires difficult empirical work concerning the effects of the various moral-judgment-formation processes on 'collective usefulness' across different contexts. No matter which interpretation of normative grounding is chosen, the normative claims put forth by neuromoral theories incur significant problems.

Acknowledgments

The authors would like to thank Joshua Greene for sharing unpublished materials and two anonymous referees for their helpful comments. Research was supported by the European Institute of Oncology (TB), the Umberto Veronesi Foundation (TB), and the VolkswagenStiftung (RR).

References

1. Kant, I. 1996/1788. Critique of practical reason. In *Practical philosophy*, ed. M. Gregor. Cambridge: Cambridge University Press. First published 1788.

2. Boyd, R. 1988. How to be a moral realist. In *Essays on moral realism*, ed. G. Sayre-McCord, 181–228. Ithaca: Cornell University Press.

3. Copp, D. 2008. Darwinian skepticism about moral realism. *Philosophical Issues*18(1): 186–206.

4. Joyce, R. 2002. *The myth of morality*. Cambridge: Cambridge University Press.

5. Joyce, R. 2006. *The evolution of morality*. Cambridge: MIT Press.

6. Gazzaniga, M. 2005. *The ethical brain*. New York: Dana Press.

7. Mikhail, J. 2007. Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences* 11(4): 143–152.

8. Hauser, M., L. Young, and F. Cushman. 2008. Reviving Rawls's linguistic analogy: Operative principles and the causal structure of moral actions. In *Moral psychology vol. 2: The cognitive science of morality*, ed. W. Sinnott-Armstrong, 107–143. Cambridge: The MIT Press.

9. Dwyer, S. 2009. Moral dumbfounding and the linguistic analogy. Methodological implications for the study of moral judgment. *Mind and Language* 24(3): 274–296.

10. Sterelny, K. 2012. *The evolved apprentice. How evolution made humans unique*. Cambridge: MIT Press.

11. Gigerenzer, G. 2008. Moral intuition=fast and frugal heuristics? In *Moral psychology vol. 2. The cognitive science of morality*, ed. W. Sinnott-Armstrong, 1–26. Cambridge: MIT Press.

12. Gigerenzer, G. 2010. Moral satisficing: Rethinking moral behavior as bounded rationality. *Topics in Cognitive Science* 2: 528–554.

13. Simon, H. 1955. A behavioral model of rational choice. *Quarterly Journal of Economics* 69: 99–118.

14. Simon, H. 1956. Rational choice and the structure of the environment.*Psychological Review* 63: 129–138.

15. Gigerenzer, G., P.M. Todd, and ABC Research Group. 1999. *Simple heuristics that make us smart*. New York: Oxford University Press.

16. Gigerenzer, G. 2000. *Adaptive thinking*. New York: Oxford University Press.

17. Gigerenzer, G., and R. Selten (eds.). 2001. *Bounded rationality*. Cambridge: MIT Press.

18. Johnson, E., and D. Goldstein. 2003. Do defaults save lives? *Science* 302: 1338–1339.

19. Lipsey, R.G., and K. Lancaster. 1956. The general theory of the second best.*Review of Economic Studies* 24: 11–32.

20. Todd, P.M., and G. Gigerenzer. 2007. Mechanisms of ecological rationality: Heuristics and environments that make us smart. In *Oxford handbook of evolutionary psychology*, ed. R. Dunbar and L. Barrett. Oxford: Oxford University Press.

21. Sunstein, C.R., and R.H. Thaler. 2003. Libertarian paternalism is not an oxymoron.*The University of Chicago Law Review* 70(4): 1159–1202.

22. Mill, J.S. 1998/1861. In *Utilitarianism*, ed. R. Crisp. Oxford: Oxford University Press.

23. Kagan, S. 1998. *Normative ethics*. Boulder: Westview Press.

24. Lenman, J. 2000. Consequentialism and cluelessness. *Philosophy and Public Affairs* 29(4): 342–370.

25. Gigerenzer, G. 2008. Reply to comments. In *Moral psychology vol. 2. The cognitive science of morality*, ed. W. Sinnott-Armstrong, 41–46. Cambridge: MIT Press.

26. Gigerenzer, G., and T. Sturm. 2012. How (far) can epistemology be naturalized?*Synthese* 187(1): 243–268.

27. Sunstein, C.R. 2008. Fast, frugal and (sometimes) wrong. In *Moral psychology, vol. 2. The cognitive science of morality*, ed. W. Sinnott-Armstrong, 27–31. Cambridge: MIT Press.

28. Greene, J.D., et al. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293: 2105–2108.

29. Greene, J.D., et al. 2004. The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44: 389–400.

30. Greene, J.D., et al. 2008. Cognitive load selectively interferes with utilitarian moral judgment. *Cognition* 107: 1144–1154.

31. Greene, J.D., et al. 2009. Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition* 111: 364–371.

32. Greene, J.D. 2008. The secret joke of Kant's soul. In *Moral psychology vol. 3. The neuroscience of morality*, ed. W. Sinnott-Armstrong, 35–79. Cambridge: MIT Press.

33. Greene, J. D. 2010. *Notes on* "*The Normative Insignificance of Neuroscience*" *by Selim Berker*. Available on http://www.wjh.harvard.edu/~jgreene/GreeneWJH/Greene-Notes-on-Berker-Nov10.pdf (accessed May 2nd, 2012).

34. Greene, J. D. (unpublished). Beyond point-and-shoot morality: Why cognitive (neuro)science matters for ethics. Forthcoming on *Ethics*.

35. Haidt, J. 2001. The emotional dog and his rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 108(4): 814–834.

36. Schwitzgebel, E., and F. Cushman. 2012. Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind and Language* 27(2): 135–153.

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

38. Haidt, J., and S. Kesebir. 2010. Morality. In *Handbook of social psychology*, 5th ed, ed. S. Fiske, D. Gilbert, and G. Lindzey, 797–832. Hoboken: Wiley.

39. Kahane, G. 2013. The armchair and the trolley: An argument for experimental ethics. *Philosophical Studies* 162(2): 421–445.

40. Kahane, G., et al. 2012. The neural basis of intuitive and counterintuitive moral judgment. *Social Cognitive Affective Neuroscience* 7(4): 393–402.

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

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

43. McGuire, J., et al. 2009. A reanalysis of the personal/impersonal distinction in moral psychology research. *Journal of Experimental Social Psychology* 45(3): 577–580.

44. Mikhail, J. 2008. Moral cognition and computational theory. In *Moral psychology vol. 3. The neuroscience of morality*, ed. W. Sinnott-Armstrong, 81–91. Cambridge: MIT Press.

45. Moll, J., et al. 2005. The neural basis of human moral cognition. *Nature Reviews Neuroscience* 6: 799–809.

46. Haidt, J. 2007. The new synthesis in moral psychology. *Science* 316(5827): 998–1002.

47. Haidt, J. 2012. *The righteous mind. Why we are divided by politics and religion*. New York: Pantheon Books.

48. Kamm, F. 2007. *Intricate ethics*. New York: Oxford University Press.

49. Singer, P. 2005. Ethics and intuitions. *The Journal of Ethics* 9: 331–352.

50. Sidgwick, H. 1907. *The methods of ethics*, 7th ed. London: Macmillan.