Penultimate draft. Final draft forthcoming in *Biology & Philosophy*.

Moving Beyond Dichotomies: Liao, S. Matthew (ed.), Moral brains: the neuroscience of morality, Oxford University Press, 2016

Ivan Gonzalez-Cabrera[1,2]
[1]Australian National University, School of Philosophy, RSSS
[2]Konrad Lorenz Institute for Evolution and Cognition Research

Abstract: Matthew Liao's edited collection *Moral Brains: The Neuroscience of Morality* covers a wide range of issues in moral psychology. The collection should be of interest to philosophers, psychologist, and neuroscientists alike, particularly those interested in the relation between these disciplines. I give an overview of the content and major themes of the volume and draw some important lessons about the connection between moral neuroscience and normative ethics. In particular, I argue that moving beyond some of the dichotomies implicit in some of the debates advanced in the book makes the neuroscience of moral judgments much more useful in advancing normative ethics.

Keywords: Dual-process theory; moral judgments; moral neuroscience; moral psychology; moral reasoning; normative ethics

Recent theoretical and empirical research in the psychological sciences has significantly advanced our understanding of moral thinking. In this changing landscape, *Moral Brains: the neuroscience of morality* does a great job at featuring leading researchers in moral cognition from a wide range of disciplines and summarizing the last two decades or so of scientific and philosophical discussion in moral cognition. This is an especially appealing book for researchers working on moral judgments, emotions and reasoning, moral decision-making and epistemology, personality disorders associated with impaired moral judgment, and the neuromodulation of moral thinking. Researchers with a broad interest in moral psychology, philosophy of psychology, normative ethics, and metaethics will also find the book valuable. The volume supplies a collection of readings in moral psychology and neuroscience that works well as an introduction for advanced undergraduates and graduate students, as well as a stimulating reading for experienced researchers interested in the connection between moral psychology and the life sciences broadly construed.

Like most collections, *Moral Brains* explores a wide range of topics without a clear unifying theme. However, a recurrent thread running through the book concerns the relation between the neuroscience of moral judgments and normative ethics and, in particular, the implications of the former for the latter. In what follows, I will give an overview of the content and major themes of the volume and draw some important lessons from it. I will argue that moving beyond some of the implicit dichotomies that permeate some of the chapters of the book may actually help research in moral neuroscience to advance normative ethics.

1. Overview of the chapters

The book begins with a helpful overview by Matthew Liao (editor and contributor to this volume) of the main issues discussed in the fourteen articles featured in the book. The chapters are divided into four parts, which are quite variable in focus, content, and methodology. The first part focuses on the role of emotions and reasoning in moral judgments, and how these different aspects of cognition can be eventually integrated into human moral thought. The second part discusses the reliability of deontological versus consequentialist judgments, focusing on the most recent version of Joshua Greene's dual-process theory of moral judgment. The third part presents new findings and methods on the neuroscience of moral judgments, emphasizing the importance of clinical, pharmacological, and model organisms in the study of moral cognition. The fourth part deals with fundamental theoretical issues that overlap with many of the debates addressed in the book.

In the first part of the volume, on the role of emotions and reasoning in moral judgments, Jesse Prinz claims that there are philosophical reasons and behavioral evidence to support a version of moral sentimentalism according to which emotions are a constitutive part of sincere moral judgments. Since the model is considered compatible with most of the empirical evidence in moral neuroscience, Prinz argues that this version of moral sentimentalism can help us to understand how different brain structures contribute to moral cognition. Jeanette Kennett and Philip Gerrans do not necessarily disagree with Prinz's constitutive model of moral judgments but they advocate in their contribution for a much broader role of reasoning in moral deliberation. Their model contrasts with that of moral intuitionist according to which moral judgments are the result of tacit affective processes that are partially encapsulated from explicit reasoning (Haidt, 2012; Haidt & Bjorklund, 2008). In Kennett and Gerrans' model, reasoning plays a major role in moral decision-making once the diachronic aspects of human agency are taken into consideration—e.g., the fact that an agent has to resolve long-term conflicts between opposing intuitive moral responses or deal with conflicting moral responses from different agents.

Perhaps the most compelling contribution in this section is that by James Woodward who explicitly rejects a sharp distinction between human cognitive and affective pathways to moral judgment. According to Woodward, areas commonly identified as involved in emotional processing contribute causally to the construction of moral judgments in neurotypical subjects. Brain areas such as the ventromedial prefrontal cortex (vmPFC), orbitofrontal cortex (OFC), anterior cingulate cortex, insula, amygdala, and the ventral striatum are involved in emotional processing. What these brain regions do is computing values associated with (positive and negative) reinforcers and the actions undertaken to provide those reinforcers. The computation of these values is essential for all kinds of decision-making because, otherwise, agents' choices would not be motivating. Among these structures, the vmPFC and the OFC integrate reward signals from different stimuli and representations from cognitive systems (Rolls, 2005). Since empirical evidence shows that moral judgments in neurotypical subjects are often causally influenced by value signals in the vmPFC and the OFC (Greene, Nystrom, Engell, Darley, & Cohen, 2004; Shenhav & Greene, 2010), areas commonly identified as involved in emotional processing would play a central role in moral judgment in neurotypical subjects regardless of whether moral judgments and choices are in fact supported by reasoning or effortful thinking. Moreover, this would mean that we cannot make a sharp distinction between cognitive and affective pathways to moral judgments in neurotypical subjects and that the moral judgments of this population would be 'sincere' in the sense of being intrinsically motivating.

The second part is arguably the heart of the book. A great deal of it is dedicated to discussing different aspects of Greene's dual-process theory of moral judgment and the second part specifically focuses on his most recent formulation (Greene, 2014), which is reprinted in this volume. In this contribution, Greene proposes two routes through which neuroscience research could have implications for normative ethics. In the direct route, independent normative assumptions are combined with neuroscientific research about the factors that our moral judgments are sensitive to. In the indirect route, neuroscientific research identifies the conditions under which automatic and effortful moral judgments are more cognitively efficient.

Greene's central argument focuses on the indirect route. He argues that current neuroscientific research favors a certain form of consequentialism. Drawing upon an analogy with digital SLR cameras, Greene maintains that human moral cognition operates in two complementary modes: a set of automatic settings and an effortful, general-purpose reasoning mode. According to him, we should not rely on our automatic moral settings when attempting to resolve moral problems with which we have inadequate evolutionary, cultural, or personal experience (or as Greene calls them 'unfamiliar' moral problems) since it would be a cognitive miracle if we turn out to have reliable good moral instincts under these conditions (p. 131). So effortful thinking is best suited for dealing with this class of moral problems. This has important consequences for normative ethics, for Greene maintains that automatic emotional responses typically support characteristically deontological judgments, while processes of effortful thinking typically support characteristically consequentialist ones. As defined in this chapter, the former are judgments that are naturally justified by appeals to rights, duties, and so on, whereas the latter are those that are naturally justified in terms of cost-benefit reasoning (p. 122). As a result, Greene concludes that characteristically consequentialist judgments are best suited for dealing with moral problems with which we have inadequate evolutionary, cultural, or personal experience.

This part of the book includes two comments by Julia Driver and Stephen Darwall as well as a reply to these comments by Greene. Driver is a long-standing advocate of consequentialist views in moral philosophy. She argues that consequentialist and deontological moral theories are in general immune to Greene's argument because the debate about whether moral deliberation is more reliable in consequentialist than deontological terms requires assuming a background moral theory that allows us to determine whether subjects' responses are morally correct or not. Darwall, in contrast, is a well-known expert on deontological approaches to ethics. He points out that arguing that consequentialist moral theories are more reliable than deontological ones implies claiming that consequentialism is a better theory of moral right.

Darwall's argument requires some attention. He begins emphasizing that Greene aims to support a particular form of consequentialism, namely act-consequentialism. According to act-consequentialists, an action is morally right if and only if that action yield the best available consequences, regardless of whether it would be best for us to be disposed to act upon non-consequentialist moral intuitions in order to bring about those consequences. This makes act-consequentialism an 'esoteric' moral theory (Williams, 1995, p. 165). For example, people could not be reasonably held accountable for acting upon those moral intuitions which best dispose them to bring about these consequences, even in situations in which those actions actually do not meet the act-consequentialist standards of a morally right action. This makes the notions of moral right and moral accountability conceptually independent of each other. But Darwall argues that the notions of moral right and wrong are tied conceptually to the idea of moral responsibility or accountability in the sense that if an action is wrong, then it is of a kind that is blameworthy unless the agent has an excuse (p. 167). Thus, he claims that on conceptual grounds, there are superior theories of moral right, including some versions of rule-consequentialism. A better account, for instance, would be one in which an agent is obligated to perform actions of which it is true that the general acceptance of a rule requiring those actions would have better consequences than would the general acceptance of any other rule in similar circumstances. Darwall himself does not endorse this form of consequentialism. Yet, he exemplifies with this his key conceptual claim without making non-consequentialist assumptions. For unlike act-consequentialism, this form of rule-consequentialism would make the notion of moral right conceptually tied to that of accountability. Since such a version of rule-consequentialism does lead to characteristically deontological judgments, we should not conclude that characteristically consequentialist judgments are more reliable than characteristically deontological ones.

In response to Driver, Greene's reply proceeds in terms of both the direct and the indirect route. I will focus only on the latter since Driver main argument focuses on the role of background moral theorizing in Greene's argument but the direct route relies on independent normative assumptions to reach substantive moral conclusions. So Greene's argument seems to be stronger when framed in terms of the indirect route than when framed in terms of direct route. According to the former, we should not rely on our automatic moral settings when attempting to resolve moral problems with which we have inadequate evolutionary, cultural, or personal experience because this would amount to expecting a cognitive miracle. Greene argues that no additional normative premise is required to support this claim since such a claim is true regardless of the standard we apply for determining reliability. To illustrate this, he considers the case of novice drivers who lack personal experience behind the wheel since it would be a cognitive miracle if they succeed in their first attempt at driving a car. Finally, Greene clarifies that he understands consequentialism not only as a decision procedure for unfamiliar moral problems but as a higher-order 'metamoral' standard, i.e., a normative standard that adjudicates among competing tribal values and interests (p. 175). Thus, he thinks that there is a standard for everyday cases and a standard for hard cases that is the same, even though the decision procedure changes depending on the nature of the decision problem.

In response to Darwall, Greene agrees with Darwall that act-consequentialism is unfit for directly guiding everyday moral behavior, but he denies that this entails that act-consequentialism is 'interpersonally' esoteric since people do have access to the foundational moral standards upheld by act-consequentialists. Moreover, Greene argues that since consequentialism is only a good normative guide for dealing with difficult moral problems, his argument does not entail the complete rejection of characteristically deontological judgments. Therefore, it is not a problem if the dictates of rule-consequentialism are characteristically deontological since the kind of metamoral theory he defends would accommodate both characteristically consequentialist judgments and characteristically deontological judgments.

Overall, Greene's responses seem to dodge the objections raised by Driver and Darwall. Assuming that we have inadequate evolutionary, cultural, or personal experience to solve a moral problem is in itself a moral assumption, which means that his main argument proceeds through Greene's direct, rather than indirect, route, and therefore it requires some background moral theorizing. Take the case of driving a car. According to Greene, this example only works within the range of plausible conceptions of good driving: "Of course, if by driving "well" you mean crashing immediately into a tree, then all bets are off. But within the range of plausible conceptions of good driving, we can say with confidence that new drivers cannot drive well based on automatic responses (intuition) and must instead rely on explicit, controlled decision-making" (p. 173). By parity of reasoning, this would hold true in the moral case only within the range of plausible conceptions of what making good moral judgments is. The problem would then be that determining this set of plausible conceptions seems to require moral theorizing. Even assuming that automatic moral responses are unreliable in situations in which we have inadequate evolutionary, cultural, or personal experience, we still require background moral theorizing. For we should not rely on our automatic moral settings only if (or to the extent that) we have inadequate evolutionary, cultural, or personal experience to solve a moral problem. However, assuming that we have inadequate evolutionary, cultural, or personal experience to solve a moral problem is a moral assumption. Since we need background moral theorizing to determine when (or to what extent) we have inadequate evolutionary, cultural, or personal moral experience, then the argument would still require background moral theorizing to support the assumption.

Greene also seems to overlook Darwall's key conceptual claim about the relation between the notions of moral right and moral accountability. If act-consequentialism is unfit for directly guiding everyday moral behavior as he agrees, then people cannot be held accountable on an everyday basis for following a different policy or acting in ways that do not meet act-consequentialist standards. Therefore, even if the kind of metamoral theory Greene defends encompasses characteristically consequentialist judgments and characteristically deontological judgments, his argument still would not fully support consequentialism construed as a theory of moral right to the extent that it does not address Darwall's key conceptual concern about the connection between the notions of moral right and moral accountability.

The third part of the book is perhaps the most attractive for those readers engaged in methodological issues around moral neuroscience. In their contribution, James Blair, Soonjo Hwang, Stuart White, and Harma Meffert defend an integrated emotion systems model of psychopathy, which aims to understand the functional properties of the neural systems involved in psychopathic traits and the computational implications of their dysfunction (Blair, 2007). They argue that emotional systems allow norms to acquire their prohibitive power by guiding our attitudes toward these norms and their violation. In the next chapter, Ricardo de Oliveira-Souza, Roland Zahn, and Jorge Moll focus on developmental psychopathy and acquired sociopathy. Their goal is reviewing and extending previous attempts to infer the neural underpinnings of moral cognition through research on normal and abnormal moral behavior. From a methodological point of view, they integrate information from functional neuroimaging on normal subjects as well as lesion studies on psychopaths and subjects with antisocial personality and conduct disorders either in vivo or through postmortem exam.

My highlights of this part of the book are Molly Crockett's and Jana Schaich Borg's contributions. On the one hand, Crockett's chapter focuses on the influences of the neuromodulator serotonin on moral judgment and behavior. The evidence reviewed in this chapter reveals, for instance, that pharmacological enhancement of serotonin function increases people's aversion to harmful actions, and thus makes people less likely to judge harmful actions as morally permissible in hypothetical scenarios. Similarly, increased levels of serotonin have been shown to reduce people's willingness to inflict financial harm on others in retaliation for unfair treatment in ultimatum games. Since there seem to be no healthy levels of serotonin, and it is currently impossible to determine a baseline physiological state from which we can generate reliable moral judgments, Crocket argues that the influence of serotonin could have important normative implications, as moral judgments would be sensitive to non-normative factors that are significantly variable. In other words, these results warn us about potential noise introduced by serotonin function at the implementation level of moral judgment and decision-making.

On the other hand, Schaich Borg's chapter discusses the relevance of rodent models of negative intersubjectivity in the study of moral behavior and cognition. Roughly speaking, negative intersubjectivity is the process of disliking or feeling negative (for whatever reason, selfish or not) when another individual feels bad (p. 248). Schaich Borg argues that a central reason to pursue this avenue of research is that negative intersubjectivity is an important regulatory mechanism of immoral action as shown in studies on the affective components of empathy and research on callous personality traits. Another reason is that neuroscience tools available in humans such as functional magnetic resonance imaging (fMRI) have poor temporal and spatial resolution to study the type of processes we believe are responsible for moral cognition and behavior. The question is, of course, whether rodent models are actual models of moral cognition, but Schaich Borg argues that rodent models should be complemented by similar tests in humans for validation and comparison. This emphasis on the role of comparative psychology in the study of moral behavior and cognition is particularly welcome since the study of non-human animal cognition connects research in the psychological sciences to the phylogenetic history, adaptive significance, and ontogeny of behavior and cognition. By focusing on moral action, Schaich Borg's contribution also reminds us of the risk of over-intellectualized views of moral cognition that have limited practical implications. Although understanding moral judgment might be philosophically deep and genuinely important, so is understanding why bad, overly aggressive behavior happens. In this context, non-human models of empathy and aggression control can be enlightening, even if they are cognitively impoverished under some reasonable anthropocentric standard.

In the final part of the book, Guy Kahane argues that the most interesting arguments that allow drawing interesting normative conclusions are epistemic in nature, i.e., arguments in which the causal origins of our beliefs affect their justification (pp. 290-291). Since the epistemic status of moral beliefs will frequently depend on whether their distal, as opposed to proximal, causes are reliable sources of moral evaluations, findings on the neural mechanisms of moral cognition will play only a minor role in such arguments. In the following chapter, Matthew Liao argues that heuristics involve a form of reasoning, regardless of whether one understands heuristics as an attribute substitution process (Kahneman & Frederick, 2005; Sinnott-Armstrong, Young, & Cushman, 2010) or as a fast-and-frugal algorithm (Gigerenzer, 2008). Given that intuitions entail forming conclusion-judgments not based on premise-judgments, they are different from reasoning, and thus different from heuristics understood either way. So, Liao argues, normative claims, such as those of Greene, that deontological intuitions tend to be inaccurate and unreliable like the automatic settings in a digital SLR camera would be unwarranted. In the closing chapter of this volume, Walter Sinnott-Armstrong draws heavily on his previous work (Parkinson et al., 2011; Sinnott-Armstrong, 2008; Sinnott-Armstrong & Wheatley, 2012, 2013) to argue that no single set of common and distinctive features of moral judgments that enables interesting psychological generalizations can unify them. Unification here means to be able to test which judgments are moral in order to reveal what it is that makes these judgments to be moral (p. 335). However, Sinnott-Armstrong argues that there are reasons to think that moral judgments are not unified in terms of their content, neural basis, and function—although Sinnott-Armstrong (2008) and Sinnott-Armstrong and Wheatley (2012, 2013) argue against other potential ways to unify moral judgments. This raises the question about what feature (or set of features) could possibly unify moral judgments in the sense specified above. As a result, he suggests a bottom-up methodological approach aimed to investigate more carefully defined subclasses of moral judgments that might or might not lead to the desired unification.

2. Discussion

As previously mentioned, one of the central themes of the book is the implications of moral neuroscience for normative ethics. So, in this part of the review, I would like to reflect further on this issue. For one central feature of the book is that many of the contributions, especially from philosophers, often point out how little we can actually learn from this data—Prinz, Woodward, and Kahane are particularly explicit on this point. Most of the contributions indeed focus on traditional psychological methods. This is understandable since traditional psychological methods are semantic (in the sense of targeting mental states with content about the world), which seems more informative than mere data about, say, the formal computations of cognitive systems or how they are implemented in actual neural systems. One important exception is Crockett's contribution since she focuses on how moral judgments respond to neuromodulators such as serotonin that are, in principle, morally inane and not clearly linked to morally relevant distal factors (see Kahane, pp. 294-295, in this volume for discussion).

The claim I want to defend now closely follows that of Woodward in this volume, for I want to argue that moving away from certain dichotomies prevalent in, but not exclusive to, Greene's dual process theory of moral judgments makes neuroscience much more useful in advancing normative ethics. More specifically, I want to challenge the following assumptions: first, the idea that either we rely on automatic moral settings or we rely on conscious reasoning, and second, the idea that either we have adequate evolutionary, cultural, or personal experience, or we have not.

Rejecting these dichotomies makes easier to derive normative conclusions from premises about neuroscientific facts by focusing on the interaction between automatic settings and effortful thinking as well as on the coordination and integration of relevant disciplines beyond neuroscience such as evolutionary biology, cultural evolution, and developmental psychology. Even assuming that moral facts are natural facts, neuroscience alone cannot bridge the gap between premises about neuroscientific facts and the moral implications that we aim to derive from those facts. Moral judgments (understood as mental states) are just not reducible to facts about neural architecture, as Sinnott-Armstrong argues in this volume and elsewhere (Parkinson et al., 2011), and the causal connection between our neural organization and the relevant facts that our moral judgments are supposed to track (whatever they are) does not seem reconstructable by neuroscientific research alone.

Regarding the first dichotomy, it is not true that we rely on either one mode of cognition or the other since automatic settings and effortful thinking interact to influence moral judgment—Woodward makes a similar point in this volume with respect to the emotion/reason dichotomy. For example, effortful thinking can influence the prediction error upon which our automatic mode of cognition operates (Daw, Gershman, Seymour, Dayan, & Dolan, 2011). Similarly, automatic settings may provide estimates which we employ through effortful thinking when forced by computational complexity to prune its online evaluation of options (Crockett, 2013). Furthermore, even if effortful thinking requires to override our default intuition and replace it by, say, conscious reasoning, the capacity to overrule intuitive responses is also a function of factors such as the metacognitive feeling of rightness in the initial response (Thompson, 2009; Thompson, Prowse Turner, & Pennycook, 2011). Hence, it is not true that we rely on either one mode of cognition or the other since we can rely on both automatic settings and effortful thinking.

Relying on both modes of cognition can reduce computational noise. Computational noise can be defined as the chance variability of judgments due to the influence of irrelevant factors. The complex calculations associated with effortful thinking are often accurate but they are not immune to computational noise—e.g., time and stress pressure, limitations on attention, speed, the ability to multitask, and depletion of other cognitive resources. Similarly, there is also noise associated with incomplete and inefficient learning associated with our automatic settings. Information gathered through experience is always partial and learning from it requires significant time. Yet the interaction between automatic settings and effortful thinking can help to reduce the computational noise of each other. For instance, effortful thinking can train our automatic settings through offline simulation (Ji & Wilson, 2007), which reduces the exploratory risk and cost associated with prolonged reinforced learning in the latter. Since effortful thinking can influence the prediction error upon which the automatic settings are learned (Daw et al., 2011), it can also reduce computational noise by speeding up learning. Moreover, automatic settings can help to reduce computational noise associated with effortful thinking by providing estimates which are used to prune the options that the latter evaluates (Crockett, 2013).

Less computational noise increases the computational robustness of the overall decision-making system. Computational robustness is the ability of a computational system to maintain its functionality across a diverse array of operational conditions. In the context of moral decision-making, it would mean something like making good moral choices in a wide range of circumstances. Reducing the chance variability of judgments due to the influence of irrelevant factors would consequently increase the chances of making good moral choices across a number of possible scenarios. Therefore, relying on both automatic settings and effortful thinking can increase the computational robustness of the overall decision-making system.

Because relying on both systems can increase the computational robustness of our moral decision-making, we can investigate how to increase this form of robustness significantly more by looking at the interaction between both systems than by looking just at the relative robustness of each system independently. To put it another way, looking at the interaction between automatic settings and effortful thinking can help us to advance normative ethics significantly more than thinking of these systems separately, for the more we understand how to increase the computational robustness of moral decision-making, the more we can advance normative ethics. A deeper understanding of these interactions and their consequences for computational performance can help us, for instance, to find more robust moral principles and theories.

Regarding the second dichotomy, if it were true that either we have adequate evolutionary, cultural, or personal experience to solve a moral problem, or we have not, then it should not be the case that sometimes we have partially adequate experience about morally relevant facts. But it is difficult to conceive a moral problem in which all our evolutionary, cultural, or personal experience turns totally inadequate. For example, we often have evolutionary, cultural, and personal experience about intentional facts that is relevant for moral evaluation. So, it is not true that either we have adequate evolutionary, cultural, or personal experience to solve a moral problem, or we have not, and thus we frequently have both adequate and inadequate experience to solve moral problems.

This shows how we need to rely on normative ethics to tell us what facts are morally relevant and when we have gathered information about them through our evolutionary, cultural, or personal experience. The relevance of some of these facts could sometimes be controversial, but not always. Relying on uncontroversially relevant moral facts puts us on the safe side, as Greene remarks. But as Humeans repeatedly remind us, we cannot logically derive a conclusion with explicitly moral content from premises without moral content—a claim that could be true even if moral predicates were synonymous with non-moral predicates (Pigden, 2010). This means that Driver is right to emphasize the background role of moral theorizing, contrary to Greene's assumption (p. 171). Whatever the metaphysical status of those facts is, we need moral theorizing to shed light on what facts are morally relevant in a particular moral situation and whether they support our premises about the adequacy of our evolutionary, cultural, or personal experience.

Moreover, we would need facts about our evolutionary, cultural, and developmental history that connect facts about our neural wiring with morally relevant facts. That is, we would need to rely not only on facts about our internal wiring on the one hand and on normative ethics to pinpoint morally relevant facts on the other but also on connection facts that link these two. Connection facts are facts about how our internal organization registers and tracks external circumstances. Considerations about these connection facts allow us to assess the reliability of our inner neural wiring to track those morally relevant facts—this tracking cannot just be a matter of luck as if we took moral decisions by throwing a dice (see Kahane, p. 294, in this volume). The life sciences can contribute much to this project because understanding the connection between cognitive machinery and relevant moral facts requires explaining how such machinery evolved, how it develops, and how it relates to our environment under ecological conditions that we often engineered through our cultural practices and which affect cognitive performance. Since we need facts about our evolutionary, cultural, and developmental history that connect facts about our neural wiring with morally relevant facts, then we need to integrate research on the life sciences more broadly (including the study of cultural evolution) for moral neuroscience to advance normative ethics.

This point is nicely illustrated by Driver's question on whether it may also be considered a cognitive miracle that moral judgments track moral truth at all, to which Greene replies that, in his understanding of cognitive evolution, it is generally adaptive to have true beliefs. Yet this line of reply makes too many assumptions about the evolutionary link between cognition and moral truth. Perhaps having mental states that track facts about our environment is adaptive but it is less clear why these mental states have to be belief-like. Perhaps having belief-like mental states that track facts about our environment is adaptive but it is less clear why those facts have to be moral. Perhaps having belief-like mental states that track moral facts is adaptive but it is less clear why this was so in the hominin lineage—we still need an evolutionary story about how having belief-like mental states that track morally relevant facts (assuming that those facts exist) was indeed fitness-enhancing. Moreover, even if having true beliefs were always adaptive, it does not follow from that that all adaptations (cognitive or not) are traits for having true beliefs. Systems involved in moral cognition can be adaptations, although not necessarily adaptations for tracking moral facts—e.g., cognitive mechanisms for language can be adaptations for solving coordination problems between multiple agents rather than adaptations for tracking environmental facts.

To sum up, *Moral Brains* explores a wide range of issues in moral psychology, even if still too attached to traditional debates such as the role of emotions and reasoning in moral cognition or the reliability of deontological versus consequentialist moral thinking. The title of the book might be indeed somewhat deceiving since not all chapters engage with actual neuroscience and not all of them with the same breadth and depth. Yet this is a rather enjoyable feature of the book and certainly an essential part of its take-home message. For bridging the gaps between neuroscientific facts and moral philosophy is an integrative enterprise, which requires a more detailed understanding of how we relate as organisms to our environments. Moral neuroscience is not, after all, reducible to mere neuroscience.

Acknowledgements: For helpful comments and suggestions, I would like to thank Claudia Passos-Ferreira and Zoe Drayson.

References

Blair, R. J. R. (2007). The amygdala and ventromedial prefrontal cortex in morality and psychopathy. *Trends in Cognitive Science, 11*(9), 387-392.

Crockett, M. J. (2013). Models of morality. *Trends in Cognitive Sciences, 17*(8), 363-366.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-Based Influences on Humans’ Choices and Striatal Prediction Errors. *Neuron, 69*(6), 1204-1215.

Gigerenzer, G. (2008). Moral Intuition = Fast and Frugal Heuristics? In W. Sinnott-Armstrong (Ed.), *Moral Psychology: The Cognitive Science of Morality: Intuition and Diversity* (Vol. 2, pp. 1-26). Cambridge, MA: MIT Press.

Greene, J. D. (2014). Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics. *Ethics, 124*(4), 695-726.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*(2), 389-400.

Haidt, J. (2012). *The righteous mind: why good people are divided by politics and religion*. New York, NY: Pantheon Books.

Haidt, J., & Bjorklund, F. (2008). Social intuitionists answer six questions about morality. In W. Sinnott-Armstrong (Ed.), *Moral Psychology: The Cognitive Science of Morality: Intuition and Diversity* (Vol. 2, pp. 181-217). Cambridge, MA: MIT Press.

Ji, D. Y., & Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience, 10*(1), 100-107.

Kahneman, D., & Frederick, S. (2005). A Model of Heuristic Judgments. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 267-293). Cambridge: Cambridge Universty Press.

Parkinson, C., Sinnott-Armstrong, W., Koralus, P. E., Mendelovici, A., McGeer, V., & Wheatley, T. (2011). Is Morality Unified? Evidence that Distinct Neural Systems Underlie Moral Judgments of Harm, Dishonesty, and Disgust. *Journal of Cognitive Neuroscience, 23*(10), 3162-3180.

Pigden, C. R. (2010). *Hume on Is and Ought*. Basingstoke: Palgrave Macmillan.

Rolls, E. T. (2005). *Emotion explained*. Oxford: Oxford University Press.

Shenhav, A., & Greene, J. D. (2010). Moral judgments recruit domain-general valuation mechanisms to integrate representations of probability and magnitude. *Neuron, 67*(4), 667-677.

Sinnott-Armstrong, W. (2008). Is moral phenomenology unified? *Phenomenology and the Cognitive Sciences, 7*(1), 85-97.

Sinnott-Armstrong, W., & Wheatley, T. (2012). The disunity of morality and why it matters to philosophy. *The Monist, 95*(3), 355-377.

Sinnott-Armstrong, W., & Wheatley, T. (2013). Are moral judgments unified? *Philosophical Psychology, 27*(4), 451-474.

Sinnott-Armstrong, W., Young, L., & Cushman, F. (2010). Moral Intuitions. In J. Doris (Ed.), *The Moral Psychology Handbook* (pp. 246-272). Oxford: Oxford University Press.

Thompson, V. A. (2009). Dual-process theories: A metacognitive perspective. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 171-196). Oxford: Oxford University Press.

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology, 63*(3), 107-140.

Williams, B. (1995). The point of view of the universe: Sidgwick and the ambitions of ethics. In J. E. J. Altham & R. Harrison (Eds.), *Making sense of humanity and other philosophical papers*, *1982-1993* (pp. 153-171). Cambridge: Cambridge University Press.