



Neurons and normativity: A critique of Greene's notion of unfamiliarity

Michael T. Dale

To cite this article: Michael T. Dale (2020): Neurons and normativity: A critique of Greene's notion of unfamiliarity, *Philosophical Psychology*, DOI: [10.1080/09515089.2020.1787972](https://doi.org/10.1080/09515089.2020.1787972)

To link to this article: <https://doi.org/10.1080/09515089.2020.1787972>



Published online: 05 Jul 2020.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

ARTICLE



Neurons and normativity: A critique of Greene's notion of unfamiliarity

Michael T. Dale

Department of Philosophy, The University of Texas at Austin, Austin, Texas, USA

ABSTRACT

In his article "Beyond Point-and-Shoot Morality," Joshua Greene argues that the empirical findings of cognitive neuroscience have implications for ethics. Specifically, he contends that we ought to trust our manual conscious reasoning system more than our automatic emotional system when confronting unfamiliar problems; and because cognitive neuroscience has shown that consequentialist judgments are generated by the manual system and deontological judgments are generated by the automatic system, we ought to trust the former more than the latter when facing unfamiliar moral problems. In the present article, I analyze one of the premises of Greene's argument. In particular, I ask what exactly an unfamiliar problem is, and whether moral problems can be classified as unfamiliar. After exploring several different possible interpretations of familiarity and unfamiliarity, I conclude that the concepts are too problematic to be philosophically compelling and, thus, should be abandoned.

ARTICLE HISTORY

Received 10 September 2018
Accepted 2 January 2020

KEYWORDS

Greene; deontology; consequentialism; moral psychology; neuroscience; evolution; unfamiliarity; utilitarianism; familiarity

1. Introduction

Over the last two decades, the philosopher-turned-neuroscientist Joshua Greene has argued that the empirical findings of psychology and cognitive neuroscience have implications for ethics. Greene has published across a wide spectrum of journals,¹ and thus has used a variety of techniques, but the general thrust of his thesis has always remained the same: recent empirical findings in the brain sciences not only support the consequentialist outlook, they also undermine deontology. For Greene, a characteristically consequentialist judgment focuses on the overall best consequences, often ignoring any strict rules or duties.² A characteristically deontological judgment, on the other hand, is one that puts much heavier emphasis on such rules, rights, duties, and obligations. To assert that one ought not break a rule, no matter what, is to make a characteristically deontological judgment.^{3,4}

While Greene uses a few different strategies to support his thesis,⁵ this article will look closely at only one of his arguments. Specifically, I will be focusing on an argument put forward in Greene's article entitled "Beyond Point-and Shoot Morality." This argument contends that we ought to trust our manual conscious reasoning system more than our automatic emotional system when confronting unfamiliar* problems;⁶ and because cognitive neuroscience has shown that consequentialist judgments are generated by the manual system and deontological judgments are generated by the automatic system,⁷ we ought to lean toward consequentialism when facing unfamiliar moral problems (Greene, 2014). In the present article, I critically examine the second premise of this argument.⁸ In particular, I ask what exactly an unfamiliar* problem is and whether moral problems can be classified as unfamiliar*. After exploring several different possible interpretations of familiarity* and unfamiliarity*, I conclude that the concepts are too problematic to be philosophically compelling, and thus should be abandoned.

2. The unfamiliarity* argument

Greene begins by applying the dual-process model of the brain (Evans & Stanovich, 2013) to the moral domain. This dual-process theory of moral decision-making maintains that there are two overarching neural systems (which are each made up of a group of smaller neural subsystems) underlying moral judgment: one associated with emotion, and the other associated with deliberate conscious reasoning (Greene, 2014, pp. 698–708).⁹ The emotional system generates automatic responses to specific stimuli: when I see someone harm an innocent person, sometimes I have an automatic emotional response that the action is wrong without understanding exactly why I think it is wrong, that is, without the influence of my conscious reasoning.¹⁰ These responses – or emotional intuitions – are part of our automatic system, and they exist for reasons of efficiency (Greene, 2014, p. 696, 2013). If a child is drowning, it is not advantageous for a bystander to consciously consider all the reasons for and against rescuing the child – time is of the essence. Thus, it is best if the bystander has an automatic emotional reaction that compels her to jump in without thinking. Similarly, if you see a large carnivore approaching you, it would be better to have an automatic flight response than a conscious deliberation about what to do.

However, the efficiency of this emotional system has its disadvantages. In situations in which it is best to actually think through the different options, these emotional intuitions can lead us astray. Indeed, we may need a manual system to override our emotional responses, and this manual system is our conscious reasoning. With regard to the drowning child, perhaps there is absolutely no way to save the child, and if the bystander were to jump in, she

would die too. In such a situation, perhaps it is best to let the child drown, even if our emotions are telling us to jump in. Thus, sometimes it is better to engage our *flexible* manual system to countermand our *efficient* automatic system.

With this theory in the background, Greene then continues on to his main premises. The first premise (P1) maintains that we ought to use our manual system when dealing with unfamiliar*¹¹ situations and problems because of its superior reliability (Greene, 2014, pp. 714–715).¹² Specifically, because our automatic system can only adequately respond to situations that it is either conditioned to respond to or genetically programmed to respond to, there is a significant chance that it will not competently address a novel situation. Now, contrast this with our manual system, that is, our conscious reasoning. In novel circumstances, our manual system is able to use its general reasoning abilities to try to figure out what the best response is. Admittedly, it will probably take longer than the automatic system to come to a decision, but the fact that it can carefully analyze all the aspects of the situation makes it overall the best neural network to use when dealing with unfamiliarity*.

Greene then argues (P2) that many important moral problems can probably be classified as unfamiliar* if we are to define unfamiliarity* as something with which we have “inadequate evolutionary, cultural, or personal experience” (Greene, 2014, p. 714). Global poverty, climate change, terrorism, and trolley dilemmas are all problems that are too new for our genes and culture to have automatic responses to address them adequately (Greene, 2014, p. 716). Of course, as Greene admits, we can’t know with absolute certainty that such problems will qualify as unfamiliar* because we don’t (yet) know everything about our cultural and evolutionary inheritances. However, if we are to make “educated guesses” with this definition in mind, he “strongly suspects” that that these moral problems – and many others – will qualify as unfamiliar* (Greene, 2014, p. 716).

At this point, Greene’s normative case is starting to take shape. If our manual system is more reliable than our automatic system when making decisions about unfamiliar moral problems, then we have reason to trust the moral judgments that are generated by our conscious reasoning more than the judgments generated by our emotions. Now, how do we know which judgments are generated by which system? Greene turns to the field of cognitive neuroscience to answer this question, and here we see his third premise (P3) emerge. Greene and his colleagues conducted an fMRI study to investigate participants’ brain functioning while making judgments about different types of moral dilemmas, and the results were striking. First, it was revealed that deontological judgments are generated by our emotional processing system because subjects who made deontological judgments about moral situations showed heightened activity in a network of neural

regions associated with emotion (the ventromedial prefrontal cortex, the precuneus, the posterior cingulate cortex, the superior temporal sulcus, the inferior parietal lobe, and the amygdala). Second, it was shown that consequentialist judgments are generated by our conscious reasoning system because subjects who made a consequentialist judgment about a moral situation showed heightened neural activity in a collection of regions associated with conscious reasoning (the dorsolateral prefrontal cortex, the inferior parietal lobe, and the posterior cingulate cortex) (Greene, 2008; Greene et al., 2004, 2001). Now, let's formalize Greene's argument:

Background model: the dual-process theory of moral judgment. Some moral judgments are generated by an automatic, emotion-based system, while others are generated by a manual, conscious reasoning-based system.

P1: We ought to trust our manual system more than our automatic system when facing unfamiliar* problems.

P2: Many important moral problems are unfamiliar* problems.

P3: Deontological judgments are generated by an emotional neural network, while consequentialist judgments are generated by a conscious reasoning neural network.

From these three premises, he concludes that because our deontological judgments are generated by an emotional neural network, they are an output of our automatic system, and because our consequentialist judgments are generated by a conscious reasoning neural network, they are an output of our manual system (C1); and because we ought to trust our manual system more than our automatic system when making decisions about unfamiliar moral problems, we ought to trust our consequentialist judgments more than our deontological judgments when making decisions about unfamiliar moral problems (C2) (Greene, 2014, pp. 716–725).¹³ To be clear, Greene isn't asserting that we should never make deontological judgments. Instead, he is saying that we have a fairly convincing reason for why we should, in general, trust our consequentialist inclinations more when we are faced with unfamiliar moral problems.

3. A first pass

The key to Greene's second premise is the concept of unfamiliarity*. If he can come up with an understanding of unfamiliarity* such that many important moral problems qualify as unfamiliar*, then, perhaps,¹⁴ he may be off and running toward his normative conclusion (C2). However, as

Paulo (2018, p. 17) points out, Greene's conception of unfamiliarity* is ambiguous due to its lack of technical explication. The following clues are all that we are given. First, as mentioned in the last section, Greene states that unfamiliar* problems are those "with which we have inadequate evolutionary, cultural, or personal experience" (Greene, 2014, p. 714). Familiar problems*, on the other hand, are those that we can deal adequately enough with our evolutionarily, culturally, and personally instilled automatic responses. Greene (2014) writes:

Automatic settings can function well only when they have been shaped by trial-and-error experience. This may be the experience of our biological ancestors, as reflected in, for example, a genetic predisposition to fear snakes. Our automatic settings may be shaped by the experience of our cultural "ancestors," as reflected in a fear of guns, despite one's having never been harmed by one. Finally, our automatic settings are shaped by our own trial and error, as when one learns to fear hot stoves by touching them. These three mechanisms – genetic transmission, cultural transmission, and learning from personal experience – are the only mechanisms known to endow human automatic cognitive processes with the information they need to function well. (p. 714)

Therefore, when our genetic, cultural, and personal predispositions are enough to adequately address a problem, then we can classify that problem as familiar*, and when they are not enough to address the problem, we can classify it as unfamiliar*.

Problems that probably fall into the latter category, according to Greene, are those that have come about through recent cultural developments – he mentions climate change, global terrorism, global poverty, and bioethics (Greene, 2014, p. 716). In his book *Moral Tribes*, he also mentions problems related to big government versus small government, the role of religion in public life, abortion, stem cell research, the teaching of evolution in public schools, and the regulation of the financial industry (Greene, 2013, pp. 99, 348).¹⁵ All of these problems arose relatively recently in our evolutionary and cultural experience, and thus, it makes sense that we wouldn't have preprogrammed responses developed through trial and error that can adequately address them.

What about familiar* problems? Greene describes problems such as poverty, disease, war, exploitation, and personal violence as "age-old," so I assume these would be classified as familiar* problems.¹⁶ In the case of personal violence, when someone physically attacks us, we have specific responses (shaped by evolution, culture, and our personal life history) that are automatically triggered to address the situation – for example, protecting our head, running away, fighting back, and so on – and since our automatic responses are usually enough to address familiar* problems such as these, our manual system is not needed.

However, if Greene is going to use familiarity* as a technical term – and a technical term upon which his entire normative conclusion rests, no less –

then it's going to need more explanation.¹⁷ To begin to understand why, let's consider exactly how much evolutionary, cultural, and personal experience is necessary for a problem to be considered familiar*. One possibility is "even a little bit," which is to say, for a problem to be classified as unfamiliar*, there needs to be no preprogrammed genetic, cultural, or personal responses that are aimed at addressing it. However, this cannot be the case because every problem requires at least some preprogrammed responses if we are going to have any chance of addressing it. Take Greene's example of driving a car. When we first approach learning how to drive, according to Greene, it is an unfamiliar* problem because we have never driven a car before and because there is no reason for driving to be part of our genetic inheritance (Greene, 2014, p. 714). However, there is actually a significant amount of evolutionarily and culturally instilled information and processing that is needed to even approach such a problem. Coordinating hand and eye movements, object tracking, seeing, recognizing distance, reading signs; these are just some of the evolutionarily and culturally instilled abilities that are engaged in a way that enables us to solve the problem of driving. Even the ability to learn new skills is almost certainly an evolutionarily instilled capacity.¹⁸

Therefore, it cannot be that for a problem to be unfamiliar*, it needs to be completely automatic-response free. It must be that it requires some automatic responses but also some manual responses – but where exactly should Greene draw the line? Let's consider driving again. As mentioned, many of the capacities that enable us to (learn how to) drive are instilled automatic responses, but there is *something* new about driving. Our hominid ancestors didn't drive (probably), so there would be no genetic inheritance specific to driving. It's also new to our personal life history, since it is our first time driving – though one could argue that playing driving and using car toys as a child is in some way practice for adult driving. The cultural inheritance question, though, is more complicated. As Greene uses the example of driving a car as an unfamiliar* problem, we can assume he believes that there are no culturally instilled automatic responses that are used to address the problem. However, in the block quote above, he describes our fear of guns as culturally instilled. This is interesting because guns are a relatively new cultural invention, just as cars are, and if the fear we feel when seeing a gun is instilled in us by cultural inheritance (which seems correct), then what about learning to drive? The generations before us have driven, so why is it not the case that they have passed down norms to us about driving a car, such as watching out for other cars and pedestrians, making sure the gauges are all working, braking when we see traffic up ahead, and so on? Indeed, if we fear guns because adults have taught us that we should fear them, and this has led to a fear response to all stimuli that look like guns, then perhaps we can drive – at least partly – because we have ridden in cars growing up, played with toy cars, been told what to do while in cars, and so on.

It's at this point that I begin to lose sight of how driving a car can be seen as an unfamiliar* problem, and this makes it difficult to use the example to pinpoint exactly how many (or what kind of) automatic responses are needed before a problem becomes familiar*. One possible route is to maintain that the fear response we adopt about guns is simple, straightforward, and reliably manifested. Driving a car, on the other hand, is complicated, and while it may be true that some of the capacities used to learn how to drive are instilled by evolution and others are instilled by our cultural experience, the act itself – with all the different capacities and abilities organized in a particular way – is new.

However, isn't that the case for almost every decision that we make, for even if we are facing a very similar decision that we had to make a few months earlier (or our ancestors had to make a few millennia earlier), we are encountering it at a different time, and thus, with a different set of moods, emotions, cultural knowledge, recent experiences, and environmental circumstances. Going back to the "age-old" problems that Greene discusses, it's true that our ancestors have been dealing with disease for millions of years, but in almost every decision that we make about disease (even if it is itself an "age-old" disease), we still need to consider the particular circumstances (e.g., the person's history, financial burdens, the state of the art of the medical field, the overall medicinal approach, etc.). Which is all to say, in almost every decision that we make, our brains are never dealing with the same inputs nor is the context ever the same. Thus, it is almost never the case that we have made the exact same decision before.¹⁹

Therefore, we are currently at a strange spot with regard to the example of driving a car. As we've seen, it's possible to argue that many of the capacities that are required to drive a car are, in fact, evolutionarily, culturally, and personally instilled, which leaves only a little room for the manual capacities that supposedly make the problem unfamiliar*; and that would mean that only a small amount of manual responses is necessary for a problem to be classified as unfamiliar*. However, at the same time, every decision needs at least a tiny amount of manual response due to different contextual issues, which means that we are dealing with a difference between a familiar* and an unfamiliar* problem that is almost certainly too small to be useful for classification.

However, maybe we have gotten off track due to the complicated nature of driving a car. Indeed, maybe it's not a good example to use when attempting to find a technical definition of the concept of familiarity* after all, and Greene only meant it as a way to help us begin to have a general understanding of the notion. Let's now turn to some of the more normatively loaded – and thus ultimately more topical – examples. Perhaps through exploring them we can find a more compelling way to understand the notions of familiarity* and unfamiliarity*.

4. Moral problems

Greene (2014, p. 716) “strongly suspect[s]” that the bizarre and unrealistic trolley problem qualifies as unfamiliar*.²⁰ Again, there does seem to be something that is intuitively true about this. If we are considering the dilemma for the first time, it is new to us and we need to take a moment to figure out what we think the best action is. Yet, it remains difficult to use this example to help develop a technical understanding of familiarity* and unfamiliarity*. For one, it’s again true that many aspects of the problem are familiar* to us. We know what all the objects in the scenario are (e.g., trolleys) due to our cultural heritage; we also know what it means to harm someone, to kill someone, and to save someone, probably as a result of some combination of our genes and personal life history. We are even able to imagine what it would (roughly) feel like to carry out a certain decision (e.g., push the large man onto the tracks).

At the same time, though, there is something different about approaching a moral dilemma than attempting to drive a car. We saw in the last section that the driving example became difficult to classify once we considered all of the culturally instilled information we have about driving before we ever get behind the wheel. In the trolley problem, however, the situation seems to be a bit clearer cut. While we are familiar* with many aspects of the dilemma (otherwise, we would probably not have any way to even understand it or begin to think about it), neither we nor our genes have needed to face this specific dilemma before; which is all to say, the pieces of the dilemma are familiar* but the final deliberation about what to do in the situation is not evolutionarily, culturally, or personally instilled, and is thus novel. By way of analogy, consider the construction of a new car.²¹ While all of the materials and the principles of engineering are familiar to the builders, the final product is something new, distinctive, and (at least the first time they build one) unfamiliar*.

With this new way of considering the issue, we may be able to clear up some of the confusions we were having before about learning to drive. While it is still the case that we are (genetically and culturally) familiar with many aspects of driving before we ever try it (indeed, even more so than the trolley dilemma because of all of the times we are exposed to driving while growing up), the act of driving itself is something novel to us the first time we attempt to do it – and that is why it takes a significant period of time before we can master the skill. Again, the components are familiar* but the activity itself is a combination of all of the familiar* components into a final product that is novel and unfamiliar*. At this point, we may have the beginning of our sought-out definition. Specifically, for a problem to be unfamiliar*, we can be familiar* with many of the capacities and skills necessary to solve the problem, but the

specific way in which the capacities and skills are applied must be novel in such a way that we need to use effortful manual processing to solve it.

However, how much effortful manual processing are we talking about here? As mentioned in the previous section, all decisions require some amount of effortful processing due to the different context one faces from one moment to the next. Could Greene maintain that there is some minimal amount of manual processing that is acceptable for familiar* problems, but after this minimal threshold has been surpassed, the problem then becomes unfamiliar*? Perhaps, but it's difficult to see how such a threshold could be convincingly established. To see why, let's consider the most likely way Greene would go about establishing it. As I explained in the second section, Greene contends that there are two general systems inside the brain: one associated with emotion (made up of the ventromedial prefrontal cortex, the precuneus, the posterior cingulate cortex, the superior temporal sulcus, the inferior parietal lobe, and the amygdala) and the other with conscious reasoning (made up of the dorsolateral prefrontal cortex, the inferior parietal lobe, and the posterior cingulate cortex). Could Greene use this neurological portrait to come up with his much-needed threshold? That is, could he claim that a certain amount of neural processing in the area of the brain associated with conscious reasoning is the necessary amount to establish unfamiliarity*?

Again, we have the hint of possibility here, but once the neurological evidence is considered (as it has been by Bluhm, 2014; Berker, 2009; Christensen & Gomila, 2012; Dean, 2010; Klein, 2011; Moll & De Oliveira-Souza, 2007; Prinz, 2016), it becomes difficult to get on board with such a possibility. Consider first the makeup of Greene's conscious reasoning system. As Greene himself acknowledges (Greene et al., 2004, p. 397), the posterior cingulate cortex (PCC) is primarily associated with emotion (Kober et al., 2008; Maddock, 1999; Maddock et al., 2003). Moreover, the inferior parietal lobe (IPL) is linked to both conscious reasoning and emotion. Interestingly, this is again not contradicted by Greene's own research, as he includes the IPL in both his conscious reasoning system and his emotional system (Greene et al., 2004, pp. 391–392). Furthermore, Borg et al. (2008) found that the IPL showed heightened activity when subjects made moral judgments about heavily emotional situations, such as incest. On top of this, there is even evidence that the dorsolateral prefrontal cortex (DLPFC) helps adjudicate between conflicting inputs, some of which have been shown to be primarily emotion based (Klein, 2011). Thus, it cannot be claimed that these regions make up a conscious reasoning-based neural network.

There are similar problems with Greene's emotional neural system. We already know that the PCC is associated with emotion, so there is no controversy there. However, both the precuneus (PC) and the superior

temporal sulcus (STS) are linked to many nonemotional tasks that play a part in conscious reasoning. Specifically, the PC plays a role in processing visual images, memory, awareness, consciousness, and taking a first-person perspective (Cavanna & Trimble, 2006; Vogeley & Fink, 2003; Vogeley et al., 2004; Vogt & Laureys, 2005). All this has led to a meta-analysis of emotional processing to not even include the PC in any of the major emotional neural networks (Kober et al., 2008, p. 14). This is not to say that the PC is not at all associated with emotion. Greene is right that it is to a certain extent, but the empirical evidence is still too indefinite, and therefore, it's difficult to confidently claim that it is part of an emotional system.

Similarly, the STS has been shown to be important not only in emotion, but also in audiovisual tasks, theory of mind, and face recognition (Hein & Knight, 2008). Could it be the case that the STS shows activity during deontological judgments (which, according to Greene, are emotionally generated) because subjects are thinking about the large man's perspective? We simply do not know, and therefore, the evidence is unclear either way. Furthermore, patients who have suffered lesioning in the VMPFC lose significant ability in decision making, planning, and understanding and following social rules, which are all linked to conscious reasoning (Damasio, 1994, pp. 54–79). Finally, even though the amygdala is popularly known as an emotional center, recent evidence has linked it to cognitive processes such as attention and associative learning (Holland & Gallagher, 1999; Pessoa, 2008). Thus, if we also take into account the previous analysis on the IPL, only one of the six regions of Greene's emotional neural network (i.e., the PCC) is solely associated with emotion.

Importantly, Greene and his colleagues acknowledge some of these inconsistencies and attempt to explain them. In particular, they argue that the PCC might be activated during consequentialist judgments because it is playing a role in motivation (Greene et al., 2004). That is, the DLPFC and the IPL are the systems doing the actual cognitive processing; the PCC is just there to motivate the subject to put forward the moral judgment. However, as they acknowledge, there is no evidence to back up this claim. No study has ever found that the PCC is concerned primarily with motivation, and thus, all they are doing is putting forward one possible interpretation of the evidence (which they admit to). However, if that's the case, then an alternative interpretation is just as plausible. Instead, it could be that the PCC is the primary neural area generating the consequentialist response, and the DLPFC and the IPL are only playing subsidiary roles, perhaps only helping the PCC to construct a moral judgment (from a primarily emotional response) and put it forward in a coherent way. Of course, I am not suggesting that this is the truth of the matter. I am only pointing out that, as far as the neuroscientific evidence suggests, pretty much any interpretation of the data is acceptable, including interpretations that fully support and fully undermine Greene's dual-process theory.

More recently, due to the seemingly conflicting research on VMPFC functioning (Bechara et al., 1994; Damasio, 1994; Grabenhorst & Rolls, 2011; Koenigs & Tranel, 2007; Koenigs et al., 2007; Moll & De Oliveira-Souza, 2007; Rangel & Hare, 2010), Greene advocates a slightly different understanding of the VMPFC than he does in most of his other work. Based on neurological evidence that the VMPFC shows a significant amount of activity during purely emotional assessments and even more activity during judgments that integrate both emotional and deliberately cognitive assessments, Shenhav and Greene (2014) argue that the VMPFC plays the role of integrator when multiple competing inputs are triggered. This would mean that when the VMPFC shows activity during a deontological judgment, it is not predominately generating an emotional response; it is instead mediating the inputs from the other, more primary emotion-based system.

While there is indeed some – albeit tentative – evidence of the VMPFC’s integrative role (Grabenhorst & Rolls, 2011; Rangel & Hare, 2010), this does not take away from the fact that lesioning of the VMPFC impairs certain (very specific) conscious reasoning-based abilities (e.g., decision making, planning, and understanding social rules), and there is no *prima facie* reason to believe that damage to an integrating system would have such an effect on only those specific abilities and not others. Therefore, yes, the VMPFC could be playing an integrative role when subjects make deontological judgments (which in and of itself is interesting: why would integration need to occur if a deontological judgment is primarily emotion based?), but it is also still a very live possibility that the VMPFC is associated with (at least some) conscious reasoning, and perhaps, that reasoning is having an effect on generating deontological judgments.

In sum, the neurological evidence is just too nebulous. Thus, while there is certainly the *possibility* that, sometime in the future, we will have a sufficiently accurate understanding of the brain to maintain that certain neural processing really is only (or at least primarily) conscious reasoning based, we are not yet there, and this makes it very difficult for Greene to claim that a certain amount of neural processing in specific areas is enough to establish the threshold of unfamiliarity*.^{22,23}

However, there is another, potentially more troubling problem lurking here. Even if we did have an accurate and nuanced understanding of human neural processing, would it be possible to come up with any sort of standard? While evolutionary biology is currently far from the stage at which it could come up with a list of uncontroversial genetic endowments for the human species such that problems being addressed by those endowments could be classified as genetically unfamiliar*, it’s certainly a theoretical possibility.²⁴ However, coming up with a universal standard for what is culturally and personally familiar* seems straightforwardly impossible, as every person is going to have a different cultural inheritance due to their

specific personal life history. Consider some of the unfamiliar* moral problems that Greene discusses. Could we agree that climate change is unfamiliar*? It's certainly true that this specific moral problem is new in the history of our species.²⁵ Thus, we may all have a similar lack of genetic inheritance with regard to climate change. However, how we actually approach the problem will differ significantly from person to person. One person – the daughter of climate scientists – might be raised to always think about the importance of the long-term survival of the earth, so for her – due to her cultural inheritance and personal life history – the problem is not unfamiliar at all. Her friend, however, is a budding philosopher, and while he may have thought a bit about what our obligations are to the next generations, he hasn't thought much about climate change specifically. Thus, for him, the problem is mildly unfamiliar*. Consider, also, a devoutly religious person, a fabulously wealthy person who was raised to not think much about future generations, a poor person who never learns to read, a person who believes in pagan deities, even a person who believes that the earth is on the back of a turtle! Due to the amazing diversity of human understanding and belief, this list could go on for hundreds of pages, and what that means is that a person will only very rarely approach the problem of climate change in the exact same way as another person, and many will approach it in significantly different ways. If this is the case, though, then it is very difficult to see how there could be a standard of unfamiliarity* (or familiarity*) for the problem of climate change.

Some might argue that I am overstating the importance of diversity. Sure, there are a certain number of experts that specialize in these problems (e.g., bioethicists, religious leaders, climate scientists), but such experts add up to a small minority of the population. Thus, for the vast majority of people, problems such as climate change will remain unfamiliar* enough to call for the use of (a certain amount of) deliberate conscious reasoning. This would mean that there would be some exceptions of familiarity* to some of these moral problems, but that wouldn't cause any significant problems for Greene's account.

However, I think that the problem of diversity is much more significant. Admittedly, if the moral problems under consideration were very abstract and difficult to understand, such as those discussed by metaethicists, then there would be a fairly clear line between familiarity* and unfamiliarity*. Those that have had sufficient training and experience in metaethics could be classified as familiar* with such problems, while those without any philosophical training would be at a loss about how to even begin to approach such discussions. However, the moral problems under consideration for Greene's account are not abstract and difficult to understand; they are problems that people think about and discuss on a fairly common basis – climate change, global terrorism, poverty, bioethics, the role of religion in

public life, abortion, stem cell research, the teaching of evolution in public schools, the regulation of the financial industry. Almost everyone has spent at least some time considering these issues,²⁶ and that means that they will have at least some familiarity* with them. Again, though, the specific type of familiarity* will differ depending on the person. True, many of them will not be considered experts on these issues, but this isn't important. All that matters is that there is such a varying degree of familiarity* both within and between populations that it becomes unfeasible to choose a particular place to draw a line.

Someone might respond by pointing out that some (perhaps many) of the people who have thought about these issues haven't thought about them in a compelling and coherent way (like the experts have), and thus, we can classify them as being as unfamiliar* with the issues as someone who really knew nothing about them. However, this would be straightforwardly untrue, as a person who has considered one of these moral issues is familiar with the issue, even if that familiarity hasn't come about through a more formal or academic setting. Indeed, consider the difference between a professional bioethicist specializing in end-of-life treatment and a woman who is curious about her options once she gets much older and reads a handful of books on the topic. Certainly, there is a difference of familiarity* in degree, but it can't be said that there is a difference in kind. Moreover, this way of conceptualizing unfamiliarly* seems destined to cause more problems than it's worth, as Greene would then have to come up with some sort of conceptualization of the "correct" type of familiarity, and down that road, the threat of arbitrary bias looms large. For instance, what makes the expert equipped with particular knowledge and training more familiar* with an issue than someone who has personally dealt with the issue firsthand but does not have an academic degree in it?

In response to these worries, Greene might claim that he doesn't need to come up with such a standard. Instead, the theory could be implemented on a more individual level: if a particular person is not familiar* with a moral problem, then she should trust her deliberate conscious reasoning, but if she is familiar* with it – due, perhaps, to her cultural upbringing – then she can rely more on her automatic settings.

Admittedly, this idea does sound promising at the theoretical level, but take a moment to consider how feasible it would be to implement. It would mean that people would have to somehow accurately understand and measure how familiar* they are with particular moral problems as they face them – but how could people come to know such information, especially in light of the well-known fact that humans are not very good at understanding their own mental states and unconscious attitudes?²⁷ Would they need to find out from a neuroscientist how much deliberate conscious processing their brain is generating when they approach the problem?

I suppose that's a theoretical possibility, but if Greene wants his theory to apply to more than just a select few individuals, then he should probably avoid this approach.

Thus, it does seem to be important for there to be some sort of standard of unfamiliarity*. It would not necessarily need to be universal, as there is variability with regard to moral problems across cultures, but Greene would need some way to classify certain moral problems as unfamiliar* that at least a significant portion of some specific populations could get behind. Otherwise, Greene would be stuck with telling people to figure out for themselves whether they are familiar* enough with a problem to rely on their automatic settings. Not only is this extra step cumbersome and unappealing, it also seems straightforwardly unrealistic in light of already discussed fact that every moral problem requires both a significant amount of automatic processing and at least a small amount of manual processing. Indeed, he would be expecting people to search for and find quite a small target.

5. Me versus us, us versus them

In the previous sections, I attempted to put forward the most compelling understandings of familiarity* and unfamiliarity* that could come out of Greene's discussion in "Beyond Point-and-Shoot Morality." (Admittedly, he does mention one other way we could understand these concepts. In particular, he claims that any time there is significant disagreement about a moral issue, that indicates that people are having conflicting intuitions about the issue, which means that there isn't enough cultural or genetic knowledge to understand the issue clearly, and therefore, the problem should probably be classified as unfamiliar* (Greene, 2014, p. 716). However, I decided to not address this interpretation because it is so problematic that it only weakens Greene's case. Indeed, not only does the objection of cultural and personal diversity apply to it, it is also just straightforwardly true that people disagree about all type of problems, including familiar* ones.)

However, if we are to turn to some of Greene's other work – specifically, his book *Moral Tribes* – there is one final way he could potentially define the terms in a more technically robust manner.²⁸ In particular, Greene could equate familiar* problems to what he calls "Me versus Us" problems, which are those concerned with strife between members of the same group (Greene, 2013, pp. 98–99, 348, 2017, p. 73). As such conflict was common throughout our evolutionary history, it is likely that natural selection instilled specific mechanisms to help address it; that is, mechanisms that suppress selfish desires and encourage cooperation. One example could be the propensity for a group to pool resources to help a sick member. This

ability to value the “Us” over the “Me” was certainly likely to enhance the survival of the group, and thus, probably can be traced back to evolutionarily ancient origins.

On the other hand, “Us versus Them” problems, which concern conflicts between groups, are much more recent, according to Greene. Of course, group conflict itself is age old, but because most of our evolutionarily instilled mechanisms are tribalistic insofar as they value in-group over out-group members, the attempt to value the “Them” over (or at least equal to) the “Us” is relatively new in the history of our species. Which is all to say, intergroup cooperation is not something that evolution would have selected for, and because of this, it’s unlikely that there have been many mechanisms instilled to address it.

Could Greene use these two concepts to define familiarity* and unfamiliarity*? That is, due to the likelihood that evolution equipped us with mechanisms to deal with intragroup conflict but not intergroup conflict, could he classify problems relating to the former as familiar* and the latter as unfamiliar*? If he did take this route, it does seem like it would fit well with (at least some of) the examples that he brings up. Greene wants to classify climate change, global terrorism, and global poverty as unfamiliar*, and while it could be argued that smaller-scale versions of these problems are age old (e.g., poverty and the use of resources within a group), the newer aspect of needing to address these problems on a global scale through intergroup cooperation could lead to them being convincingly classified as unfamiliar*. On the other hand, intragroup poverty, disease, exploitation, and personal violence only require cooperation within a group and are therefore likely to be adequately addressed by instilled mechanisms.

Overall, this seems to be a promising possibility. However, it starts running into problems once we consider again the complicated nature of cultural and personal familiarity*. As described, if we develop a culturally instilled response to a problem that is enough to address the problem, then it is no longer unfamiliar*. However, if this is the case, then it becomes much more difficult to describe intragroup group problems as familiar* and intergroup problems as unfamiliar*. Recall the diversity of cultures, beliefs, religions, and viewpoints discussed in [Section 4](#). As each of these belief systems has its own view of intergroup conflict and how to solve it, every culture will be coming to the intergroup cooperation table with different amounts of familiarity*. Admittedly, this would have been less of an issue before a few thousand years ago because intergroup cooperation was probably not a priority then,²⁹ but, now that globalization is part of everyday life, it is simply a matter of fact that each belief system has its own way of thinking about global problems.

For example, while it might be true that a small minority of isolated foraging cultures have never considered the issues of climate change, global

terrorism, and global poverty, the vast majority of cultures existing today are aware of these issues. Thus, there is already a certain amount of familiarity* that is culturally instilled in most populations with regard to these moral problems. Of course, different cultures feel differently about how to address these problems (e.g., some think we ought to help the poor in other countries while others do not), but that doesn't take away from the fact that they have considered (and thus, have become at least somewhat familiar* with) such problems. Due to this, it would be difficult to claim that intergroup problems are unfamiliar* in the way Greene's theory needs them to be (i.e., unfamiliar* to the point at which automatic responses would not be adequate to address them).

Furthermore, if it is true that cultural influences expose most populations to intergroup problems, then that means that people will have thought about these issues at least somewhat, and this brings us back to the problem of individual-level diversity. Indeed, some people – due to their familial upbringing and personal life history – might, for example, be especially familiar* with global terrorism but only vaguely familiar* with disease, and this would mean that their deliberate conscious reasoning should be used with regard to an intragroup problem, as opposed to an intergroup problem. Some might want to again argue that this would only result in a few exceptions, but this seems very unlikely due to the already mentioned influence of globalization. Most people discuss and consider all of the moral problems Greene brings up (whether they be intragroup problems or intergroup problems), and thus, there is no telling who will end up being relatively unfamiliar* with exploitation, who will end up being fairly familiar* with climate change, who will end up being extremely familiar* with global poverty, and so on.

Unfortunately for Greene, this problem crops up with every attempt to define familiarity*, and it reveals itself to be quite devastating. Is there a way to get around it? For example, could Greene just drop the idea of familiarity* due to culture and personal life history, and just stick with evolutionarily instilled familiarity*? If he did this, then this way of defining familiar* problems as “me versus us” and unfamiliar* problems as “us versus them” could work. However, it wouldn't stand up under scrutiny because, at the end of the day, if Greene wants to maintain that certain (moral) problems are unfamiliar* while others are familiar*, what could plausibly be the difference between culturally instilled familiarity* and evolutionarily instilled familiarity*? They are both equipping us with the knowledge and capability to address a problem, so it is difficult to see how only the latter could be classified as familiar*. Indeed, an automatic response is an automatic response. What difference does it make if one comes about through the trial-and-error processes of evolution and another through the trial-and

-error processes of cultural learning? None at all, it seems, and thus, the problem of diversity beleaguers Greene's second premise until the end.

6. Broader implications

In this article, I outlined what seemed to be the most promising elucidations of the concepts of unfamiliarity* and familiarity* in the hope of finding a more robust and, ultimately, compelling understanding of the second premise of Greene's unfamiliarity* argument. In so doing, I explored some more technical neurological approaches, as well as other more theoretical approaches. In the end, however, the problems (especially the problem of cultural and personal diversity) proved too significant.

To see how this affects Greene's unfamiliarity* argument, recall the formalized version of the argument:

Background model: the dual-process theory of moral judgment, where some moral judgments are generated by an automatic, emotion-based system, while others are generated by a manual, conscious reasoning-based system.

P1: We ought to trust our manual system more than our automatic system when facing unfamiliar* problems.

P2: Many important moral problems are unfamiliar* problems.

P3: Deontological judgments are generated by an emotional neural network, while consequentialist judgments are generated by a conscious reasoning neural network.

C1: Deontological judgments are an output of our automatic system, while consequentialist judgments are an output of our manual system.

C2: We ought to trust our consequentialist judgments more than our deontological judgments when making decisions about unfamiliar moral problems.

If the concept of unfamiliarity* is incoherent, P2 quickly collapses because there is no way to classify a certain group of moral problems as unfamiliar* problems. Importantly, though, notice that P1 is also no longer tenable. P1 contends that our manual system is more trustworthy than our automatic system in certain circumstances, but if Greene is unable to compellingly classify or define such circumstances, then there is no longer any reason to accept this premise – and if Greene loses P1, he loses his “ought,” which, of course, is the key to C2. He can still associate deontological judgments with emotion and consequentialist judgments with conscious reasoning,³⁰ but he can no longer claim that we ought to trust one of

those neural networks (and the judgments generated from it) in a particular set of circumstances.

If Greene's unfamiliarity* argument is no longer tenable, this has important implications for his overall project. In "Beyond Point-and-Shoot Morality," Greene puts forward two arguments for why the findings of cognitive neuroscience and psychology have implications for ethics. One argument is the unfamiliarity* argument (he calls it the "indirect route"), which is what I discussed in this article. The other argument (which he calls the "direct route," and can be found in Greene, 2014, pp. 711–713) focuses on why the majority of participants choose to sacrifice the one person in the 'switch' dilemma but also choose not to push the large man onto the tracks in the 'footbridge' dilemma (Greene et al., 2004, 2001).³¹ According to Greene, these two dilemmas are asking the same question (i.e., would you sacrifice one to save five?), so the puzzle is why are participants answering them in different ways? His explanation is that participants have an (evolutionarily instilled) emotional reaction to the idea of physically pushing the man onto the tracks, and this prevents them from advocating the consequentialist judgment even though they normally would choose to sacrifice one life to save five.³² Thus, the following normative argument can be put forward (cf. Greene, 2010, p. 16; Paulo, 2018, p. 7):

P1: People's deontological judgments in response to 'trolley' problems are strongly influenced by the presence of personal force.

P2: The presence of personal force is morally irrelevant to the moral acceptability of actions such as these.

C: People's deontological judgments in response to 'trolley' problems are strongly influenced by at least one morally irrelevant factor, personal force, and are therefore at least somewhat unreliable.

However, as Kumar and Campbell (2012) point out, this argument is unwarranted. Most of us can agree on the idea that the presence or absence of personal force is a morally irrelevant factor, so we know that something is going wrong somewhere with people's moral judgments about 'trolley' dilemmas. However, we cannot know exactly where because it could equally well be the case that the consequentialist judgments in the 'switch' case should be discounted because they are overly sensitive to the impersonal nature of the harm. Importantly, Greene (2014, p. 713) himself acknowledges this.³³ Thus, the best Greene can do with his "direct route" is this:

P1: People's judgments in response to 'trolley' problems are strongly influenced by the presence or absence of personal force.

P2: The presence or absence of personal force is morally irrelevant to the moral acceptability of actions such as these.

C: People's judgments in response to 'trolley' problems are strongly influenced by at least one morally irrelevant factor, personal force, and are therefore at least somewhat unreliable.

While intriguing, this argument doesn't have any specific normative clout. Greene knows this, though, and it seems to be one of the main reasons he feels the need to put forward a more general theory that has a further reaching normative conclusion (Greene, 2014, p. 713). Enter the unfamiliarity* argument, which makes claims about which types of (moral) judgments we should trust more broadly. However, if my arguments in this article are correct, and the concept of unfamiliarity* is not philosophically tenable, then Greene must retreat back to his significantly more limited "direct route" that, while interesting, won't do much to tilt the standoff between deontology and consequentialism.

Notes

1. For example, see Greene et al. (2001); Greene (2003), 2008, 2009, Greene (2013), 2014).
2. The term 'characteristically' is important for Greene. To say that a judgment is characteristically consequentialist is to say that the judgment is most naturally justified by the consequentialist outlook, regardless of the actual reasoning that led to the judgment – the same applies for deontological judgments. Therefore, importantly, it is possible for someone to use deontological reasoning to come to a characteristically consequentialist conclusion (Greene, 2014). Of course, this behavioral understanding of consequentialist and deontological judgments is not in line with how the concepts are commonly used in the philosophical literature (e.g., see note 3), but Greene (2014, p. 699) is careful to acknowledge this. Now, there may be significant problems with defining the concepts in this way (such as those pointed out in Kahane, 2012; Paulo, 2018), but to discuss them would be outside the scope of this article.
3. Of course, this brief explanation doesn't do justice to the nuances of the deontological enterprise. To take just one example, deontology often allows for exceptions, and thus, the absolutist tone of this definition is in some ways misleading. However, because this is how Greene defines the notion, and it is a *good enough* definition insofar as it emphasizes the primary focus of the deontologist's enterprise (i.e., the act itself, as opposed to its consequences), I will stick with it and refrain from delving into the (albeit important) details.
4. A helpful way to see the difference between these two views is through the now canonized trolley problem. See note 20 for an explanation.
5. See note 1.
6. I use an asterisk to indicate I am referring to Greene's technical notion of familiarity and unfamiliarity.

7. From now on, I will drop the term ‘characteristically,’ but it should be understood that whenever I refer to consequentialist or deontological judgments, I have Greene’s more behavioral understanding of these concepts (discussed in note 2) in mind.
8. The premises of the argument will be laid out in the next section.
9. Importantly, these two systems usually conflict when a person is making a difficult moral decision, and it is the “winner” of this conflict that dictates the final decision.
10. Jonathan Haidt (2012) is famous for arguing that almost all moral judgments are generated by a system like this.
11. Greene (2014) defines “unfamiliar problems” as ones “with which we have inadequate evolutionary, cultural, or personal experience” (p. 714). See the following paragraph.
12. Unfortunately, Greene does not go into any specific detail about what it means for a system to be reliable, but we can extrapolate from his argument that reliability probably has to do with the ability of a system to generate true beliefs about the world. This, however, brings up the (extremely important) question of whether such true beliefs accurately track morally true beliefs; that is, whether the practical reliability of a neural system has any significant connection to the moral reliability of the system. This potential problem hangs over Greene’s entire argument, but as my critique lies elsewhere, I will not be discussing it in this article. See Königs (2018) and Paulo (2018) for further discussion.
13. For the fully formalized version of the argument, see [Section 6](#).
14. I say ‘perhaps’ because there may well be significant problems with the other premises.
15. In his book, he isn’t discussing the notion of (un)familiarity*, but he does describe these problems as newer problems that are unique to our modern way of life, so I am assuming that Greene would probably classify them as unfamiliar*.
16. Again, this comes from a discussion unrelated to the concept of familiarity*, but these seem to fit the bill. They are also all that we are given, as he does not put forward any specific examples of familiar* problems in “Beyond Point-and-Shoot Morality.”
17. See the final section for a discussion about how exactly the loss of P2 affects his argument.
18. Indeed, one can well imagine that it would be costly and unnecessary for a species that relies only on strict instinctual responses to evolve such an ability.
19. I put the qualifier ‘almost’ here because it may be the case that in some rare circumstances, the decision is exactly the same, but, in all honesty, I have a difficult time imagining such a scenario.
20. The standard ‘switch’ version of the dilemma is where you see that an out-of-control trolley is heading toward five people standing on the track. If you flip the lever in front of you, the trolley will switch onto a different track with only one person standing on it. Do you let the five people die (which would be to make a characteristically deontological judgment), or do you pull the lever and cause the trolley to run over the one person (which would be to make a characteristically consequentialist judgment)? The ‘footbridge’ version of the trolley dilemma requires pushing a large man onto the tracks to stop the trolley from hitting five people.
21. To stick with the driving theme.
22. Of course, this is only a practical defeater, and thus, it still leaves open the possibility that we could match familiarity* and unfamiliarity* with specific neurological functioning. However, as this particular attempt at interpreting Greene’s concepts rests fundamentally on the possibility of neurological measurement, the impracticality of it is important.

23. Some readers may notice that this discussion of the neurological evidence has broader implications. Indeed, if it is true that the research is nebulous in the way that I suggest, it casts doubt on not only P3 of Greene's argument but also the entire dual-process theory of moral judgment. Importantly, I do think that such doubt should be cast, but as this particular article is focused on P2, it would be outside its scope to get into a detailed discussion of its implications for P3 and the background model.
24. However, again, this level understanding seems fairly far off, which adds on to the methodological worries brought up in the previous paragraphs.
25. Sacrificing our short-term interests for a long-term goal, though, is probably not so new. Thus, there is a fundamental aspect of this problem that could be seen as quite old. More on this in the next section.
26. Even people in nonindustrialized societies deal with many of these issues (or, at least, very similar ones), albeit on a more local scale.
27. For evidence of this well-supported finding in cognitive science, see Gawronski and Payne (2010), Mai et al. (2011), and Perkins and Forehand (2012).
28. To be clear, Greene never explicitly maps out this possibility. It is merely connecting the dots in an attempt to give the most charitable reading to Greene's notion of unfamiliarity*. Königs (2018) was the first to make this connection.
29. Of course, hunter-gatherer groups often engage in alliances with nearby groups, but I imagine that, for the most part, such alliances were always in the interest of the group itself and would take a backseat if they came into conflict with the priorities of the group.
30. Some of the evidence in this article, though, did indirectly call into question this association as well (see note 23).
31. For a recap of these dilemmas, see note 20.
32. 'Normally' should be understood as something along the lines of "from an abstract, removed perspective, unpolluted by emotion." Of course, this seems to be already assuming the consequentialist framework, but as I am only discussing broader implications in this section, I'm going to pass over such (admittedly important) details.
33. In his own words, "are we oversensitive to personal force in response to *footbridge*, or under-sensitive in response to *switch*?" (Greene, 2014, p. 713).

Acknowledgments

For all of their helpful feedback and support, I would like to thank Jerry Samet, Julia Driver, Marion Smiley, Bertram Gawronski, Sinan Dogramaci, Jamie Kelly, Kate Moran, David Sosa, Joshua Greene, Selim Berker, William Edgecomb, Elliot Goodine, Sarah Carr, Skylar Brannon, Sanaz Talaifar, Annie Jung, and four anonymous reviewers from *Philosophical Psychology*. I'd also like to thank the members of the Brandeis University Speaker Series and the Boston Area Moral Cognition Group.

Disclosure statement

No potential conflict of interest was reported by the author.

References

- Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 50(1–3), 7–15. [https://doi.org/10.1016/0010-0277\(94\)90018-3](https://doi.org/10.1016/0010-0277(94)90018-3)
- Berker, S. (2009). The normative insignificance of neuroscience. *Philosophy & Public Affairs*, 37(4), 292–329. <https://doi.org/10.1111/j.1088-4963.2009.01164.x>
- Bluhm, R. (2014). No need for alarm: A critical analysis of Greene’s dual-process theory of moral decision-making. *Neuroethics*, 7(3), 299–316. <https://doi.org/10.1007/s12152-014-9209-0>
- Borg, J., Lieberman, D., & Kiehl, K. (2008). Infection, incest, and iniquity: Investigating the neural correlates of disgust and morality. *Journal of Cognitive Neuroscience*, 20(9), 1529–1546. <https://doi.org/10.1162/jocn.2008.20109>
- Cavanna, A., & Trimble, M. (2006). The precuneus: A review of its functional anatomy and behavioral correlates. *Brain*, 129(3), 564. <https://doi.org/10.1093/brain/awl004>
- Christensen, J. F., & Gomila, A. (2012). Moral dilemmas in cognitive neuroscience of moral decision-making: A principled review. *Neuroscience and Biobehavioral Reviews*, 36(4), 1249–1264. <https://doi.org/10.1016/j.neubiorev.2012.02.008>
- Damasio, A. (1994). *Descartes error: Emotion, reason, and the human brain*. Penguin.
- Dean, R. (2010). Does neuroscience undermine deontological theory? *Neuroethics*, 3(1), 43–60. <https://doi.org/10.1007/s12152-009-9052-x>
- Evans, J., & Stanovich, K. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8(3), 223–241. <https://doi.org/10.1177/1745691612460685>
- Gawronski, B., & Payne, B. (Eds.). (2010). *Handbook of implicit social cognition: Measurement, theory, and applications*. Guilford Press.
- Grabenhorst, F., & Rolls, E. T. (2011). Value, pleasure, and choice in the ventral prefrontal cortex. *Trends in Cognitive Sciences*, 15(2), 56–67. <https://doi.org/10.1016/j.tics.2010.12.004>
- Greene, J. (2003). From neural ‘is’ to moral ‘ought’: What are the moral implications of neuroscientific moral psychology? *Nature Reviews Neuroscience*, 4(10), 846–849. <https://doi.org/10.1038/nrn1224>
- Greene, J. (2008). The secret joke of Kant’s soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology: The neuroscience of morality* (pp. 35–79). MIT Press.
- Greene, J. (2009). The cognitive neuroscience of moral judgment. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences IV* (pp. 987–999). MIT Press.
- Greene, J. (2010). Notes on ‘the normative insignificance of neuroscience’ by Selim Berker. Unpublished manuscript.
- Greene, J. (2013). *Moral tribes: Emotion, reason, and the gap between Us and Them*. Penguin Press.
- Greene, J. (2014). Beyond point-and-shoot morality: Why cognitive (neuro)science matters for ethics. *Ethics*, 124(4), 695–726. <https://doi.org/10.1086/675875>
- Greene, J. (2017). The rat-a-gorical imperative: Moral intuition and the limits of affective learning. *Cognition*, 167(1), 66–77. <https://doi.org/10.1016/j.cognition.2017.03.004>
- Greene, J., Nystrom, L., Engell, A., Darley, J., & Cohen, J. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44(2), 389–400. <https://doi.org/10.1016/j.neuron.2004.09.027>
- Greene, J., Sommerville, R., Nystrom, L., Darley, J., & Cohen, J. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537), 2105–2108. <https://doi.org/10.1126/science.1062872>

- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Pantheon Books.
- Hein, G., & Knight, R. (2008). Superior temporal sulcus—it's my area: Or is it? *Journal of Cognitive Neuroscience*, 20(12), 2125–2136. <https://doi.org/10.1162/jocn.2008.20148>
- Holland, P. C., & Gallagher, M. (1999). Amygdala circuitry in attentional and representational processes. *Trends in Cognitive Science*, 3(2), 65–73. [https://doi.org/10.1016/S1364-6613\(98\)01271-6](https://doi.org/10.1016/S1364-6613(98)01271-6)
- Kahane, G. (2012). On the wrong track: Process and content in moral psychology. *Mind & Language*, 27(5), 519–545. <https://doi.org/10.1111/mila.12001>
- Klein, C. (2011). The dual track theory of moral decision-making: A critique of the neuroimaging evidence. *Neuroethics*, 4(2), 143–162. <https://doi.org/10.1007/s12152-010-9077-1>
- Kober, H., Feldman Barrett, L., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T. (2008). Functional grouping and cortical-subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *Neuroimage*, 42(2), 998–1031. <https://doi.org/10.1016/j.neuroimage.2008.03.059>
- Koenigs, M., & Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: Evidence from the ultimatum game. *The Journal Of Neuroscience*, 27(4), 951–956. <https://doi.org/10.1523/JNEUROSCI.4606-06.2007>
- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to prefrontal cortex increases utilitarian moral judgments. *Nature*, 446(7138), 908–911. <https://doi.org/10.1038/nature05631>
- Königs, P. (2018). On the normative insignificance of neuroscience and dual-process theory. *Neuroethics*, 11(2), 195–209. <https://doi.org/10.1007/s12152-018-9362-y>
- Kumar, V., & Campbell, R. (2012). On the normative significance of experimental moral psychology. *Philosophical Psychology*, 25(3), 311–330. <https://doi.org/10.1080/09515089.2012.660140>
- Maddock, R. (1999). The retrosplenial cortex and emotion: New insights from functional neuroimaging of the human brain. *Trends in Neuroscience*, 22(7), 310–316. [https://doi.org/10.1016/S0166-2236\(98\)01374-5](https://doi.org/10.1016/S0166-2236(98)01374-5)
- Maddock, R., Garrett, A., & Buonocore, M. (2003). Posterior cingulate cortex activation by emotional words: fMRI evidence from a valence decision task. *Human Brain Mapping*, 18(1), 30–41. <https://doi.org/10.1002/hbm.10075>
- Mai, R., Hoffmann, S., Helmert, J., Velichkovsky, B., Zahn, S., Jaros, D., EH Schwarz, P., & Rohm, H. (2011). Implicit food associations as obstacles to healthy nutrition: The need for further research. *The British Journal of Diabetes & Vascular Disease*, 11(4), 182–186. <https://doi.org/10.1177/1474651411410725>
- Moll, J., & De Oliveira-Souza, R. (2007). Moral judgments, emotions and the utilitarian brain. *Trends in Cognitive Sciences*, 11(8), 319–321. <https://doi.org/10.1016/j.tics.2007.06.001>
- Paulo, N. (2018). In search of Greene's argument. *Utilitas*, 31(1), 38–58. <https://doi.org/10.1017/S0953820818000171>
- Perkins, A., & Forehand, M. (2012). Implicit self-referencing: The effect of nonvolitional self-association on brand and product attitude. *Journal of Consumer Research*, 39(1), 142–156. <https://doi.org/10.1086/662069>
- Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews Neuroscience*, 9(2), 148–158. <https://doi.org/10.1038/nrn2317>
- Prinz, J. (2016). Sentimentalism and the moral brain. In S. M. Liao (Ed.), *Moral brains: The neuroscience of morality* (pp. 45–73). Oxford University Press.

- Rangal, A., & Hare, T. (2010). Neural computations associated with goal-directed choice. *Current Opinion in Neurobiology*, 20(2), 262–270. <https://doi.org/10.1016/j.conb.2010.03.001>
- Shenhav, A., & Greene, J. (2014). Integrative moral judgment: Disassociating the roles of the amygdala and ventromedial prefrontal cortex. *The Journal of Neuroscience*, 34(13), 4741–4749. <https://doi.org/10.1523/JNEUROSCI.3390-13.2014>
- Vogeley, K., & Fink, G. (2003). Neural correlates of the first-person-perspective. *Trends in Cognitive Sciences*, 7(1), 38–42. [https://doi.org/10.1016/S1364-6613\(02\)00003-7](https://doi.org/10.1016/S1364-6613(02)00003-7)
- Vogeley, K., May, M., Ritzl, A., Falkai, P., Zilles, K., & Fink, G. (2004). Neural correlates of first-person perspective as one constituent of human self-consciousness. *Journal of Cognitive Neuroscience*, 16(5), 817–827. <https://doi.org/10.1162/089892904970799>
- Vogt, B., & Laureys, S. (2005). Posterior cingulate, precuneal & retrosplenial cortices: Cytology & components of the neural network correlates of consciousness. *Progress in Brain Research*, 150(1), 205–217. [https://doi.org/10.1016/S0079-6123\(05\)50015-3](https://doi.org/10.1016/S0079-6123(05)50015-3)