4

THE NEUROSCIENCE OF MORAL JUDGMENT

Joanna Demaree-Cotton and Guy Kahane

1. Introduction

We routinely make moral judgments about the rightness of acts, the badness of outcomes, or people's characters. When we form such judgments, our attention is usually fixed on the relevant situation, actual or hypothetical, not on our own minds. But our moral judgments are obviously the result of mental processes, and we often enough turn our attention to aspects of these processes—to the role, for example, of our intuitions or emotions in shaping our moral views or to the consistency of a judgment about a case with more general moral beliefs.

Philosophers have long reflected on the way our minds engage with moral questions—on the conceptual and epistemic links that hold between our moral intuitions, judgments, emotions, and motivations. This form of armchair moral psychology is still alive and well, but it's increasingly hard to pursue it in complete isolation from the growing body of research in the cognitive science of morality (CSM). This research is not only uncovering the psychological structures that underlie moral judgment but, increasingly, also their neural underpinning—utilizing, in this connection, advances in functional neuroimaging, brain lesion studies, psychopharmacology, and even direct stimulation of the brain. Evidence from such research has been used not only to develop grand theories about moral psychology but also to support ambitious normative arguments.

Needless to say, these normative arguments are contentious, as is, more generally, the relation between the CSM and traditional philosophical accounts of moral judgment. Where some assert that empirical evidence could resolve longstanding ethical debates (see, e.g., Churchland, 2011; Greene, 2008, 2016), others argue that neuroscience has no normative significance whatsoever (Berker, 2009).

The aim of the present chapter is to bring a measure of increased clarity to this debate. We will proceed as follows. We shall begin with general reflections about the potential bearing of neuroscience on moral epistemology. Focusing on the issue of the reliability of our moral judgments, we shall suggest that neuroscientific findings have limited epistemic significance considered on their own; they are likely to make an epistemic difference only when "translated" into higher-level psychological claims.[1] But neuroscientific findings are

anyway best understood as merely one stream of evidence feeding into the CSM, a broader scientific enterprise whose focus is primarily at a higher level of description. Questions about the normative significance of neuroscience are therefore unhelpful unless "neuroscience" is understood in this broader sense.

These general reflections will guide the rest of the chapter. We will briefly introduce some key theories and findings that have dominated the "first wave" of recent cognitive science of morality (circa 2000–2010) that much of the philosophical debate has focused on: the "moral grammar" theory defended by Mikhail and others, Haidt's social intuitionist model, and Greene's dual process model.[2] We then consider more closely several key themes and debates that have shaped this research, highlighting both their potential normative significance and important recent developments in empirical research that further complicate the scene, including the rejection of a sharp dichotomy between cognition and emotion,[3] and a departure from strong nativist assumptions to interest in "moral learning".[4]

2. Is Neuroscience Relevant to Moral Epistemology?

The epistemic status of our moral intuitions and judgments often depends on their immediate causal history. Most obviously, it is widely thought that moral judgments resulting from an *unreliable process* are not epistemically justified.[5] Think, for example, of forming moral judgments about others' actions purely on the basis of their race, or whether you are fond of them, rather than on the basis of the circumstances, nature, and consequences of their action. Few, if any, hold that these are reliable ways of forming moral judgments.[6]

Reliability may not be the only feature of judgment-forming processes that affects the normative status of moral judgments. Many moral epistemologists hold that epistemic justification requires not only reliability but also forming one's judgments in response to appropriate, morally relevant reasons. This requirement may also be morally important. For example, a correct, reliably produced moral judgment might nevertheless fail to express a virtuous character unless it was formed in response to the reasons that make that moral judgment correct.[7]

This brief sketch of the epistemic significance of the processes that generate our moral judgments both offers a straightforward argument supporting an epistemic role for the CSM and sets important constraints on that role.

First the argument. What kinds of processes our moral judgments depend on influences the normative status of those judgments in various ways. But what kinds of processes produce our moral judgments is an empirical matter. So insofar as the CSM sheds light on the processes that produce our moral judgments, it seems straightforwardly relevant to moral epistemology.

But while the CSM obviously has much to say about these processes, to impact moral epistemology it must shed light on the *epistemically relevant* aspects of these processes. And here things get more complicated.

Think of the kinds of things the CSM can tell us about the sources of our moral judgments. It may clarify the *personal-level* processes that underlie them: what features of the case are consciously registered, whether or not these judgments result from explicit deliberation, whether they are based in felt emotions, etc. The CSM can also uncover the *sub-personal information processing* that underlies our judgments.[8] Our judgments may

be the result of complex unconscious computations, shaped by implicitly held principles and responsive to features of a case in ways that escape our conscious awareness. Finally, the CSM can offer an account of these processes at the *neural* level: by identifying brain circuitry and patterns of neural activation involved in generating these judgments.

Now when a judgment-forming process is specified in personal-level terms—and in many cases, also in information processing terms—we have philosophical and empirical resources available to us to make reasonable judgments as to whether that type of process is likely to produce reliable, reason-responsive moral judgments or not. Our earlier examples of unreliable processes took this form: racial prejudice or biased motives are paradigmatic examples of processes that undermine the epistemic standing of a judgment.[9] Similarly, knowing which features of a moral case are being registered (and which ignored), what implicit or explicit principles are being applied, what kind of deliberative process is being followed—these can also be directly evaluated for epistemic respectability.

So, insofar as we can infer what psychological process is being implemented by a given neural process, neuroscience can indirectly inform us about the reliability of moral judgments. But unless we can make such inferences, how can we determine whether a neural process is likely to issue in accurate moral judgments?

It seems that we could only evaluate the reliability of a neural process by observing what moral judgments result from it and then using armchair methods to directly evaluate the judgments themselves. This is likely to require controversial commitments to substantive normative claims. It risks circularity, if we end up justifying a set of moral judgments by appeal to the reliability of a neural process that has been certified as reliable precisely *because* it has produced the relevant judgments. One could perhaps try to avoid such circularity by arguing that certain controversial moral judgments were produced by a reliable neural process because that process also reliably produces judgments that we all agree to be correct—though it is not obvious why the reliability in the latter context must carry over to the former (Kahane, 2016). But in any event, the work neuroscience does on this approach is pretty minimal: identifying types of processes that we then try to correlate with patterns of moral judgments. What exactly is involved in these processes, at the neural level, is irrelevant.[10]

These methodological considerations lead to a more substantive claim. The claim is that a moral judgment has the epistemic properties it does in virtue of higher-level psychological rather than neural properties of the judgment-forming process.

On most accounts of moral justification, it is the psychological properties of judgment-forming processes that are epistemically relevant. It seems clear why this would be so for internalist accounts of justification, which stress the importance of forming beliefs on the basis of good reasons. After all, the kinds of states and processes that can constitute good or bad reasons are best described in personal-level terms, e.g.: processes of deliberation, the experience of certain emotions, weighing up of evidence, or the perception of certain non-moral features. However, reliabilist accounts of justification also emphasize the importance of psychological processes. This is because reliabilism has had to face the challenge (the so-called "generality problem") of telling us *which* way of specifying "the" process leading up to a judgment is the relevant one for evaluating the judgment's justification. Attempted solutions have tended to stress the relevance of the reliability of the *psychological* process on which the judgment is founded.[11] This is partly because it seems important to ensure

that beliefs are "well-founded" or "properly based"—i.e. produced by a process involving evidentially relevant mental states. By contrast, how a psychological process is physically implemented seems intuitively irrelevant to epistemology; if two people arrive at the same moral judgment by weighing up the same evidence in the same fashion, it seems irrelevant to the epistemic status of that judgment whether their neural hardware differed in any interesting way.[12]

So a purely neural description of the processes leading to a moral judgment will on its own tell us little about its epistemic status. We first need to map these neural properties onto higher-level psychological ones. Still, you might think that if there were a straightforward, one-to-one mapping between psychological processes and neural ones, we could—especially as neuroscience continues to progress—simply translate epistemically relevant psychological processes into neural terms. If so, questions about the epistemic status of the process leading up to a moral judgment could in principle be answered in exclusively neural terms. However—whilst we are increasingly able to make reasonable inferences about psychology from neuroscientific data—there are good reasons for thinking that the relation between psychological and neural types isn't going to be straightforward in this way.

First, many (if not all) types of psychological processes seem to be *multiply realized*—they may be realized by distinct neural arrangements in different individuals, or even within the same individual over time. For example, emotional processing that is normally supported by paralimbic brain areas in nonclinical populations might be supported by the lateral frontal cortex in psychopaths (Kiehl, 2008). This represents one way in which simple one-to-one mapping can fail.[13]

Secondly, any given brain area or network is likely to be involved in many distinct psychological processes in different contexts (see e.g. Pessoa, 2013), and it may make little sense to ask whether neural activation in such a network counts as a reliable process overall. For example, activity in the amygdala may contribute to a reliable psychological process-type in certain cases (e.g., representing abhorrent behavior as negative) and an unreliable psychological process-type in others (e.g., hostility to outgroup members).

Finally, many epistemically important distinctions that are salient when judgment-forming processes are described in psychological terms are unlikely to carve distinctions that are significant at the neural level. Of course, there must be some neural difference between reliable and unreliable psychological processes in a given context.[14] But this difference needn't be one that is useful for describing the functional organization of the brain. For example, while there are competing accounts of the considerations that are relevant to the moral evaluation of some person, action, or institution, it seems unlikely that the consideration of these "morally relevant factors" on the one hand and patently morally irrelevant factors will map onto two interestingly distinct neural kinds on any such account. Indeed, the very same neural network may be involved in both, just with subtly different patterns of activation. Conversely, unless researchers embrace an exceedingly simplistic moral scheme on which (e.g.) nothing but physical pain and pleasure are relevant to the morality of an action, person, or institution, the patterns of neural activity involved in perceiving or cognizing morally relevant considerations may be quite heterogeneous—at least as heterogeneous as the considerations themselves. So it may be only when we describe what these patterns of activity represent at a higher level that it will become salient that they fall into epistemically significant groups.

Consequently, psychological descriptions are likely to retain their primacy over neural ones in moral epistemology. There is thus little point in considering the normative significance of neuroscience independently of the specific role of neurobiological claims within the larger body of evidence and theorizing in the CSM, much of which is anyway, at least at this stage, largely focused on higher-level cognitive processes.[15] Neurobiological evidence from, say, neuroimaging or psychopharmacology can support or challenge theories of moral psychology, but these theories ultimately stand or fall in light of the whole body of relevant empirical evidence, which will frequently involve traditional questionnaires, evidence about response times, introspective reports, and the like.

3. Three Approaches to the Cognitive Science of Morality

The turn of the century has witnessed a dramatic surge in scientific interest in moral judgment. After decades of seemingly fruitless attempts to solve the so-called problem of consciousness, cognitive science has turned its attention to morality. Spurred by the development of functional neuroimaging and general trends in cognitive science, cognitive scientists rejected the stale rationalist developmental theories of Piaget and Kohlberg and instead emphasized the role of innate, automatic and—somewhat more controversially—affective processes in driving moral judgment.[16]

The CSM has so far been dominated by three main approaches. The **universal moral grammar** approach—championed by Harman, Mikhail, Hauser, and Dwyer—builds on Rawls's early work to draw a direct analogy between moral psychology and Chomskian linguistics.[17] On this view, our core moral judgments reflect the working of a "moral organ": unconscious computations map input about causation, intention, and action onto innately represented moral principles to produce universally shared intuitions about moral permissibility and wrongness. Jonathan Haidt's **social intuitionist** approach also emphasizes the centrality of automatic intuitions in shaping moral judgment. On Haidt's view, however, our moral intuitions result from rapid emotional reactions—a claim supported by studies purporting to show that we can manipulate moral judgments by triggering emotions such as disgust.[18] The universal moral grammar and social intuitionist approaches sharply disagree about the role of emotion in moral judgment but agree that moral judgment is almost exclusively the product of automatic intuitions, not of conscious reasoning. For the universal moral grammar theorists, explicit reasoning plays a minimal role in our core moral "competence"—just as the conscious application of rules play a minimal role in our grammatical competence. For Haidt, such reasoning is largely used to rationalize intuitive views we would hold anyway and to pressure others into sharing our views. The third approach, Joshua Greene's **dual process model**, shares much with Haidt's but adds an important twist.[19] Like Haidt, Greene holds that a great deal of moral judgment is shaped by immediate "alarm-bell"-like emotional reactions (a "system-1" type process) and that much of the justification offered in support of our moral views—including much of the theorizing of moral philosophers—is merely ex-post rationalization. But Greene also claims to find evidence for an important exception, arguing that utilitarian judgments are uniquely based in explicit reasoning (a "system-2" type process). And Greene has famously argued that this difference—which he traces to distinct neural structures—supports a normative argument favoring utilitarianism.

These are but brief sketches of these main three approaches.[20] In what follows we will consider some of their claims more closely. The rest of the chapter is organized thematically. We will consider the epistemic import of several key debates in recent moral psychology: debates about the *domain-specificity* of moral cognition, about the respective roles of *emotion* and *reason*, and about *nativist* and *learning* approaches. We will review some of the core evidence for the three approaches but also ways in which more recent research has cast doubt on some of their key claims. In line with the discussion of the previous section, we will highlight key findings at the neural level but always as they bear on psychological theories that support claims about personal-level processes of potential epistemic importance.

4. Domain-Specificity and General Capacities

Psychologists distinguish between domain-specific and domain-general capacities. Domain-specific capacities are dedicated to a target domain (possible examples: grammatical competence, face recognition); domain-general capacities apply across domains (example: general intelligence). When we form moral intuitions and judgments, are we utilizing a capacity that is specific to morality—or even a "moral module"—or are we merely drawing on general psychological capacities? Relatedly, are there specific brain areas dedicated to moral cognition? If there is a moral module, it will be natural to expect it to be realized in distinctive neural circuitry (Suhler & Churchland, 2011), though in principle a module could be realized in a more distributed network.

The universal moral grammar (UMG) approach is most clearly committed to the existence of such a moral module, but the other approaches have also been friendly to domain-specificity: Haidt's moral foundations theory, which seeks to develop the social intuitionist approach by describing in more detail the mechanisms that produce innate, affective intuitions, claims that they are produced by domain-specific, functional modules (Graham et al., 2012). Even Greene sometimes suggests that emotional "system 1" judgments are the product of evolved, domain-specific capacities.[21]

The evidence so far has not been very kind to the idea of a dedicated moral module, at least not one with unique neural correlates. Early moral neuroimaging studies used fMRI to investigate which brain areas were associated with specifically moral content by comparing conditions where participants assessed stimuli with moral content to conditions where they assessed nonmoral but otherwise similar content.[22] These studies found that moral cognition employs many neural networks involving areas distributed around the brain, including the ventromedial and dorsomedial prefrontal cortex (vmPFC and dmPFC), the temporoparietal junction (TPJ), the precuneus, the posterior cingulate cortex (PCC), the amygdala, and the temporal pole. Moreover, these brain areas are not used exclusively for moral cognition. Rather, the brain networks involved in morality overlap extensively with those involved in theory of mind, emotion, and a host of other functions, including imagination, memory, and causal reasoning—all of which are capacities that we also use for nonmoral cognition.[23]

This has not, however, sounded the death toll of domain-specificity. Haidt and Joseph (2011) have defended moral modules against neuroscientific objections of this kind by highlighting that psychological modules are not the same thing as neurobiological modules.[24] As noted earlier, psychological mechanisms might be domain-specific even if not based in

a specific area of the brain; activity in distributed, overlapping neural circuits can produce specialized mechanisms at the psychological level. For example, some UMG theorists proposed that morality relies on a unique *interaction* between theory of mind and emotions to produce distinctively moral cognition (Hauser, 2006, 219; Hauser & Young, 2008). And current research has not clearly distinguished the question of the existence of a mechanism dedicated to producing moral outputs from the question of the existence of a mechanism dedicated to producing evaluative or normative outputs more generally; the latter might exist even if the former doesn't.[25] Nevertheless, the view that moral cognition is based in more general psychological capacities currently seems more plausible.

The empirical question of domain-specificity ties in with one common worry about moral intuitions, namely that it's hard to see how we could calibrate them in a noncircular fashion. If the processes in question produced *nothing* but moral outputs, this worry would be reinforced. By contrast, if moral intuitions result from domain-general capacities that also produce nonmoral outputs, then we have an independent way of assessing the reliability of the processes that generate our moral intuitions. This would also address the worry that our moral intuitions are produced by a mysterious faculty that is "utterly different from our ordinary ways of knowing everything else" (Mackie, 1977).

The empirical question of domain-specificity also relates in a fairly direct way to the question of epistemic reliability: to answer questions about reliability, we first need to identify the processes whose reliability we are trying to assess. For example, Nado (2014) has argued that evidence of domain-specificity of different categories of intuitions, including moral intuitions, means that the reliability of each category of intuitions must be assessed separately. Others have argued that moral intuitions inherit the reliability of the domain-general processes that produce them.

One such strategy is an influential line of response to Sharon Street's Darwinian challenge to moral realism.[26] In brief, Street (2006) argues that our moral beliefs are extensively shaped by evaluative dispositions that have an evolutionary origin. Since it's highly unlikely that such evolutionary pressures track an objective realm of moral facts, we have no reason to think that our moral judgments are reliable if we endorse moral realism. In reply, Parfit has argued that when we assess the intrinsic plausibility of our moral intuitions we are employing the same rational capacity we use when we assess epistemic, logical, and other a priori claims.[27] Since there *was* evolutionary advantage in having a general rational capacity that tracks epistemic and logical facts, we have good reason to think that this capacity is reliable—a conclusion we can then carry over to the moral domain. One may wonder, however, whether reliability in one domain must automatically carry over to a completely different one. But more importantly, for our purposes, this strategy relies on falsifiable empirical claims: it assumes that at least reflective moral intuitions are the product of a domain-general capacity.

Now, as we saw earlier, the current balance of the evidence doesn't give much support to the idea of a domain-specific moral module. Unfortunately, however, it also doesn't lend much support to Parfit's aforementioned strategy. As we shall see, evidence from psychopathy and lesion studies suggests that the *content* of our moral judgments is strongly dependent not on whether we possess general rational capacities but on whether we have certain emotional sensibilities. It is very unlikely that we form moral judgments simply by exercising general rational capacities.

These considerations take us directly to another debate at the center of recent empirical moral psychology—that about the respective roles of emotion and reason in moral judgment.

5. Emotion and Deliberation

A long tradition in moral philosophy, stretching back to Plato and Kant, emphasized a sharp distinction between reasoning and emotions, with cool reasoning the source of practical rationality and moral knowledge and emotions an irrational, distorting influence on moral judgment.[28] This division between irrational emotion and rational reasoning dominated early work in the CSM. It's clearly in the background of Haidt's social intuitionist model and Greene's dual process model, with the main difference being that, according to the social intuitionist model, reasoning hardly ever influences our moral judgments. In contrast, according to Greene's model, genuine moral reasoning produces one category of moral judgment, namely "utilitarian" judgments—such as the judgment that it's permissible to sacrifice one person as a means to saving others—while "deontological" judgments result from emotions that function in an irrational, "alarm-like" way that often overrides rational processing. By contrast, UMG theorists, who are more favorable toward such "deontological" judgments, insist that the immediate intuitions on which they are based result from "cool" unconscious computations.

However, this sharp division between emotion and reason is out of touch with important recent work in moral epistemology. It has been argued, for example, that emotions are often needed to bring morally relevant features to our attention and that they may even be necessary for grasping their moral importance. Emotions, then, can play an epistemically beneficial, perhaps even essential, role in the pursuit of moral knowledge (e.g., Arpaly, 2003; Jaggar, 1989; Little, 1995; see also De Sousa, 2014).[29] Whilst few deny that certain emotional experiences have an epistemically negative effect—for example, intense, overwhelming emotions that distort our perception of the situation and prevent us from considering relevant evidence—philosophers in this camp would argue that "cold" reasoning can also be subject to bias and blind spots, especially when entirely severed from emotional input.

Recent developments in neuroscience strongly support this alternative way of thinking about emotion and reason (Moll et al., 2008; Nichols, 2002; see also May and Kumar, Chapter 7 of this volume; Railton, 2017, especially 5.2). In the mid-twentieth century, it was thought that emotional and cognitive processes were supported by distinct, dedicated brain regions (the "limbic system" and the neocortex, respectively). However, there is growing consensus that there is no such sharp anatomical divide in the highly interconnected human brain (e.g. Barrett & Satpute, 2013; LeDoux, 2012; Lindquist & Barrett, 2012; Okon-Singer et al., 2015; Pessoa, 2008). Brain areas involved in emotion play a crucial role in cognitive functions such as learning, attention, and decision making, and emotions depend on brain regions involved in various cognitive operations. Consequently, cognitive scientists increasingly reject the characterization of psychological processes involved in moral judgment as either "emotional" or "cognitive" (e.g. Cushman, 2013; Huebner, 2015; Moll, De Oliveira-Souza & Zahn, 2008). If emotions are not easily separable from the "cognitive" means by which we represent, process, and evaluate the world, then this opens up the possibility

that emotional processes—not just "cold" reasoning—can make rational, evidence-sensitive contributions to moral judgment.

This suggestion is supported by recent research on the network of brain areas that Greene associated with "emotional" moral judgments—in particular, the right temporo-parietal junction (rTPJ),[30] the amygdala, and the ventromedial prefrontal cortex (vmPFC) (Greene et al., 2001; Greene et al., 2007).

The "emotional" process Greene identified appears to be based on an initial unconscious analysis of what seem like morally relevant factors concerning the intentions of agents and their causal relationship to harm. A number of recent EEG and imaging studies suggest that moral violations are initially identified in the rTPJ (Harenski, Antonenko et al., 2010), prior to experienced emotion.[31] The rTPJ is crucial for making moral distinctions on the basis of whether harm to persons is intended, unintended, or merely accidental (Schaich Borg et al., 2006; Chakroff & Young, 2015). So, if you artificially disrupt rTPJ activity, participants fail to take intentions into account (Young, Camprodon et al., 2010), judging, say, that deliberate attempts to harm that fail are morally acceptable, or that harming someone by accident is wrong.

The EEG studies suggest that processing continues next in the amygdala. While the amygdala is involved in affect, it is a mistake to think of it as simply generating gut feelings. The amygdala is involved in associative learning and is a crucial node in the "salience" network, which, given that we have limited time and computational resources, uses affective cues to help us pay greater attention to information that is likely to be contextually important—for example, because it has previously been associated with threats, norm violations, or benefits (Barrett & Satpute, 2013; Pessoa & Adolphs, 2010). This allows the amygdala to rapidly identify potentially relevant features and, through extensive connections to the visual and prefrontal cortex, to direct attention and processing resources to those features.

Both the amygdala and the rTPJ feed information about potentially morally relevant features to frontal areas including the vmPFC.[32] Whilst the vmPFC allows emotionally processed information to influence moral judgment, recent fMRI studies suggest it doesn't do this by generating "alarm-like" reactions designed to dominate judgment, as Greene's dual process model claimed. Rather, it facilitates a productive interaction between emotions and reasoning.

Firstly, the vmPFC is part of the so-called "default mode" brain network that is responsible for imagination, visualization, and empathic understanding.[33] One way the affectively charged salience network directs attention to morally relevant features is by recruiting the default mode network to imaginatively simulate the perspectives of harmed parties (Chiong et al., 2013).

Secondly, the vmPFC allows us to represent different pieces of information generated by different brain areas as evaluatively relevant for the moral question at hand, converting information into "common currency" (Huebner, 2015) and allowing us to feel its relevance (Young, Bechara et al., 2010). This fits with imaging studies showing that its activity isn't associated directly with specific inputs to moral judgment—such as emotional responses or the calculation of costs and benefits (narrowly construed)—but specifically with attempts to make an overall moral judgment (Hutcherson et al., 2015; Shenhav & Greene, 2014); nor is it associated with a specific *type* of conclusion, such as deontological or utilitarian judgments (Kahane et al., 2012) or the judgment that some controversial issue is wrong or not

wrong (Schaich Borg et al., 2011). Finally, there don't appear to be inhibitory relationships between brain areas associated with emotions and those associated with cost-benefit analysis (Hutcherson et al., 2015); this again supports the hypothesis that the vmPFC does not facilitate the influence of emotions at the expense of other information but rather allows us to weigh information together to produce "all-things-considered" judgments—a view that is more consonant with common forms of moral deliberation (see Kahane, 2014).

So it seems as if "emotional" processes, mediated by areas such as the rTPJ, amygdala, and vmPFC, allow us to process morally relevant information and thus can contribute to good moral reasoning. This is supported by studies of clinical populations with neural abnormalities that disrupt these circuits, such as psychopaths, patients with vmPFC damage, and patients with behavioral-variant fronto-temporal dementia (FTD). While much of their ability to reason remains intact, these clinical populations suffer from emotional deficits, including diminished empathy, that restrict their "cognitive" abilities to correctly perceive, attend to, and take into account morally relevant properties.[34] For example, patients with vmPFC lesions and FTD patients have an impaired ability to infer which emotional states others are experiencing (Shamay-Tsoory & Aharon-Peretz, 2007); vmPFC patients struggle with decision making because they lack appropriate emotional reactions (Damasio, 1994) and they fail to respond to harmful intentions (e.g., judging that attempted murder is permissible;Young, Bechara et al., 2010), and FTD patients display sociopathic tendencies. Psychopaths display abnormal functioning in the amygdala, vmPFC, and TPJ when presented with moral transgressions and consequently fail to attend to and correctly process morally salient properties such as harm and mental states (Decety et al., 2015; Harenski, Harenski et al., 2010; Hoppenbrouwers et al., 2016).

These clinical populations also give abnormally high rates of the so-called "utilitarian" judgments—judging that it is morally permissible to sacrifice one person as a means to saving others. Rather than being the product of good moral reasoning unfettered by irrational emotion, it seems these judgments are associated with these patients' failure to register complex facets of moral value (Gleichgerrcht et al., 2011).

Indeed, in nonclinical populations, the so-called utilitarian judgments that Greene associated with the dorsolateral prefrontal cortex (dlPFC) are not associated with genuinely utilitarian, impartial concern for others but rather with rational egoism, endorsement of clear ethical transgressions, and lower levels of altruism and identification with humanity (Kahane et al., 2015). Furthermore, in a study by FeldmanHall and colleagues (2012) in which participants decided whether to give up money to stop someone receiving painful electric shocks, activity in the dlPFC was associated with self-interested decisions and decreased empathic concern, while the vmPFC was associated with prosocial decisions.[35] Moreover, the "emotional" circuits seem to facilitate truly impartial, altruistic behavior. For example, Marsh and colleagues (2014) found that extraordinary altruists have relatively enlarged right amygdala that are more active in response to other people's emotions—which they are *better* at identifying.

This is not to say that reasoning supported by the dlPFC cannot contribute positively to moral judgment. The dlPFC, part of the frontoparietal control network, is considered especially important for holding goals and norms in working memory and thus is important for overriding intuitive responses on reflective grounds. For example, Cushman and colleagues (2012) found that activation of the dlPFC was associated with the condemnation

of harmful omissions—perhaps due to an attempt to treat acts and omissions equally on the basis of consistency (see Campbell & Kumar, 2012). So while reasoning sometimes suffers from moral blind spots or is used for egoistic goals, it can make an epistemically positive contribution to moral judgment, especially when conjoined with emotional processes (see also May & Kumar, Chapter 7 of this volume).

The neuroscientific work reviewed here is not conclusive. Recording brain activity through fMRI or EEG cannot directly establish the causal impact of emotions or reasoning on judgment, and we must be cautious in drawing conclusions about normal moral judgment from those suffering from neuropsychological impairments (e.g. Bartels & Pizarro, 2011; Huebner, 2015; Kiehl, 2008).

Nevertheless, on balance, current evidence suggests that both emotions and reasoning contribute to moral judgment and that moral judgment may operate at its best when reasoning and emotion interact. Indeed, the influences of "emotion" and "reasoning" are not always cleanly separable, as when we use empathy to understand the effects of an action on others or when we weigh up moral reasons whose moral import was drawn to our attention using affective cues.[36]

These findings fit nicely with a variety of positions in moral epistemology according to which moral reasons, whether represented by intuitions or deliberated about in reasoning, should be weighed together to produce all-things-considered judgments, and that emotions and reasoning can work together to achieve this (Kahane, 2014). These findings also pose a serious challenge to attempts to debunk moral judgments on the ground that they are influenced by emotion or to regard them as epistemically sound merely because they are based on reasoning. Rather, attempts to debunk moral judgments will have to depend on more fine-grained descriptions of the psychological process in question.

We have said a fair amount to show that it is both philosophically simplistic and empirically problematic to sharply contrast emotion with reason. This contrast is motivated not only by an unjustifiably dismissive view of emotion but also by a narrow understanding of reason. Longer response times are the central form of evidence for effortful explicit deliberation in moral judgment, but the mere fact that someone takes longer to reach a moral conclusion hardly shows that this process is more reliable. That surely also depends on *what* exactly the deliberation involves. Taking longer to form a moral judgment can be a bad sign—think of someone who takes a while to decide whether pushing a man off a footbridge *just for fun* is permissible. Researchers in CSM rarely give a detailed account of what moral reasoning is supposed to involve, but it is usually assumed to involve putting together an explicit moral argument.[37] This would not help establish a sharp epistemic contrast between deliberation and emotion and intuition given that arguments require premises and, one might argue, these would ultimately need to be based on intuitions. But in any event, intuitions and emotions aren't just necessary inputs to deliberation: they are often directly involved in it. Deliberation can just involve forming more reflective intuitions about a given case, and a great deal of moral deliberation consists of weighing opposing reasons and considerations, a process that in large part involves higher-order intuitions to the effect that one set of reasons outweighs another (Kahane, 2014); we've discussed evidence about the neural basis of this process. Moreover, feelings such as certainty, confidence, and doubt play a key role in shaping such deliberation. Future work in the CSM will need to operate with richer conceptions of both emotion and reason.

6. Moral Nativism and Moral Learning

We turn, finally, to the question of innateness and learning. Biological traits are the product of a complex interaction between innate structure and environment. What is at issue is thus the *extent* to which aspects of our moral psychology are relatively determined in advance of experience. According to moral nativists, a full account of human moral psychology will need to make significant reference to features of it that are organized in advance of experience (see e.g. Haidt & Joseph, 2011; Graham et al., 2012). According to non-nativists, by contrast, moral judgments are best explained as the product of learning processes interacting with the environment, and little reference to specialized innate structure is needed.[38]

Much work in the CSM has been dominated by strong nativist assumptions. UMG theorists claim that moral judgment is produced by an innate moral module, while both the social intuitionist and dual process models explain patterns of moral intuitions by reference to emotional responses selected by evolution.[39] However, evidence for these claims is fairly limited. Nativists frequently appeal to evidence that certain patterns in moral judgments are found universally across human cultures (e.g. Dwyer et al., 2010; Haidt & Joseph, 2004). But universal patterns of moral judgment could also be explained by learning mechanisms if we assume that relevant kinds of environmental input are universal. Stronger evidence for nativism comes from the application of Chomskian poverty-of-the-stimulus arguments to morality. For example, UMG theorists (and Haidt, 2001, 826–827) appeal to developmental evidence regarding the speed with which young children develop moral judgments that conform to sophisticated, abstract moral rules—such as a rule against intentionally harming others as a means. They argue that children could not develop this sort of moral psychology without innate constraints, especially given the minimal sorts of explicit feedback children get about morality, which usually concerns highly specific actions (e.g. "You should not have hit your brother") rather than general principles (Mikhail, 2007; Dwyer et al., 2010).

However, Nichols et al. (2016; see also Nichols, Chapter 6 of this volume) have recently shown that simple Bayesian assumptions make it possible to quickly infer rules prohibiting acts (but not omissions) and intended harm (but not foreseen harm) from very few occasions of highly specific, minimal feedback. This suggests that we may acquire sophisticated, abstract moral rules from the application of domain-general learning mechanisms to minimal cultural input about what sort of behavior is prohibited.

Other recent work in the CSM similarly emphasizes the role of experience and learning in morality (e.g. Allman & Woodward, 2008; Crockett, 2013; Cushman, 2013; Railton, 2017; see also Campbell & Kumar, 2012), pushing against the earlier nativist status quo. This trend draws on a large body of recent work in neuroscience and computational modeling that suggests that our brains make wide use of prediction-error signals to facilitate powerful forms of learning. Evidence suggests that such learning underlies the detection of a variety of morally relevant features, including intentions and other mental states, causation, risk, reward, and expected value. Moreover, the neural mechanisms that encode reward and value (in the neuroscientist's sense) in reinforcement learning are responsive not only to personal material reward but also to abstract social values relevant to morality, such as character assessments. Although direct evidence for neural coding of specifically "moral value" and prediction errors is lacking as of yet, it seems likely that such learning also shapes moral judgment itself.

Neuropsychological evidence from clinical populations is consonant with the hypothesis that domain-general reinforcement learning mechanisms play an important role in moral judgment. Given their emotional deficits, psychopaths are generally impaired in their ability to learn from negative affective reactions to predict future harms or modify their own behavior—a deficit also reflected in abnormal patterns of moral judgment and behavior. Early damage to the vmPFC similarly impairs moral judgment.

The literature on reinforcement learning typically distinguishes between two sorts of algorithms—model-free and model-based—that are thought to characterize learning and decision making in the brain in distinct but overlapping circuits (Crockett, 2013; Cushman, 2013; Huebner, 2015; Railton, 2017; Nichols, Chapter 6 of this volume). *Model-based* algorithms compare the value of candidate actions based on a detailed *model* of all of the expected outcomes associated with them. This is computationally costly (it involves going over a lot of information), but it's also far-sighted and very flexible. By contrast, *model-free* algorithms assign value to actions in specific contexts ("states") simply based on reinforcement history—i.e. on whether the action-state pair has previously been associated with good or bad outcomes. This reinforcement learning can be achieved through experience, observation, or possibly through simulating the consequences of actions (Miller & Cushman, 2013). Model-free algorithms are relatively inflexible, with the value assigned to action representations only changeable over time.

How the model-free and model-based distinction exactly contributes to moral judgment needs further research. Although we have been writing in terms of action-selection, model-free and model-based systems can be defined over all sorts of representations (consequences, situations, etc.), so caution is advised against assuming that they cleanly underwrite the distinction between "deontological" and "consequentialist" judgments assumed by Greene's model (Ayars, 2016; Cushman, 2013), and model-free learning may not be able to explain the *persistence* of certain deontological intuitions (Railton, 2017). Nevertheless, the distinction between model-free and model-based learning may play a role in the distinction between automatic action-based and controlled outcome-based moral assessment. In particular, model-free systems may in part be responsible for the greater moral condemnation of "personal" harm, of acts over omissions, and intended harms over unintended side-effects (Crockett, 2013; Cushman, 2013).

Debates about the evolutionary (or other) sources of moral judgment are obviously of great interest, but their epistemic significance isn't straightforward. UMG theorists occasionally write as if the aim of moral philosophy is to uncover the innate "moral code" posited by UMG theory. But this is an odd idea. If this innate moral code is the product of natural selection, why should we let it guide our actions? After all, evolution "aims" at reproductive fitness, not at moral truth. Why think that dispositions that were reproductively advantageous to our ancestors in the savannah track any kind of moral truth? These kinds of considerations lead Greene (2008) to a contrary conclusion: if certain moral judgments have their source in our evolutionary history, then they should be treated with suspicion. Instead, we should use our general capacity for reason to arrive at independent, consequentialist conclusions.

Evolutionary debunking arguments of this kind have received a great deal of attention in recent years (see e.g. Kahane, 2011; Vavova, 2015).[40] One worry is that if they work at all, they will support general moral skepticism (Kahane, 2011; Ruse, 1988). We should also

distinguish such debunking arguments from the different argument that evolution-selected dispositions *were* truth-tracking in our ancestral environment but lead us astray in the very different modern context (Singer, 2005, seems to conflate these two forms of argument).

Those wishing to resist evolutionary debunking arguments often seek to deny the nativist assumption that such arguments require. It might be thought that the recent shift to moral learning offers hope for such a strategy. If our moral judgments actually have their source in moral learning then it seems that evolutionary debunking arguments cannot get off the ground. However, things aren't so simple. To begin with, evolutionary pressures may still affect the direction of moral learning, especially if the learning operates on core environmental features that have remained constant. If so, then our current moral intuitions would still have an evolutionary source in the sense the debunkers assume. Relatedly, the contribution of some innate structure has not been ruled out. In particular, moral learning must operate on a set of goals or "values", and these are almost certain to have an evolutionary source. [41]

Even if evolutionary forces did not shape our moral judgments, this hardly shows they are truth-tracking. This depends on how moral learning operates and what it operates on. Advocates of moral learning often emphasize the way such learning involves "rational" processes since they are sensitive to evidence and feedback over time (e.g. Railton, 2017). So perhaps we needn't worry, as Greene (2016) does, about hardwired intuitions that fail to adapt to modern moral problems. But there is a great gap between tracking the moral truth and being "rational" in the sense of effectively identifying general patterns in one's environment and relating them to pre-set goals. If our deontological intuitions are, for example, merely the side-effect of a Bayesian learning heuristic interpreting the target of others' condemnation (as Nichols et al. suggest), then this may be as debunking as an evolutionary explanation. On the other hand, if deontological intuitions arise through learning what behaviors are associated with callous, anti-social, and otherwise immoral character traits (as Railton, 2017, suggests—a hypothesis supported by Everett et al., 2016), then they may have a basis in morally relevant considerations. The bottom line is we need more empirical research on the nature of moral learning before debunking worries can be dropped.

Whether or not this epistemic worry can be addressed, moral learning accounts seem rather far from the idea that our moral judgments have their source in the exercise of a general rational capacity to reflect on a priori matters (Parfit, 2011). On such accounts, general capacities are indeed involved, but these are capacities to detect robust regularities in our environment. It is hard to see how such learning processes could detect the intrinsic wrongness of certain acts—they would at best support a broadly consequentialist reading of deontological intuitions (see e.g. Railton, 2017). However, experience can play a role in a priori reflection—we may need relevant experience (and hence, learning) to properly *comprehend* the content of fundamental moral principles, principles that can nevertheless be known without reliance on evidence from experience. Whether emerging accounts of moral learning are compatible with this picture remains to be seen.

7. Concluding Remarks

While scientific theorizing about morality has a long history, the CSM is a fairly new field. The approaches that have dominated it in the first decade of this century already seem out

of date or at least in need of major revision while exploration of alternative directions (e.g., relating to moral learning) have only started. We have tried to give a reasonably up-to-date survey of the key theories and findings in the area, though, inevitably, there is also a lot of interesting work we had to leave out. What does seem clear, however, is that we are seeing a rapid advance in the scientific understanding of moral psychology. It is unlikely that this growing understanding will leave moral epistemology unchanged, and we have tried to trace some of the key connections. There are no simple knock-down arguments from findings in psychology and neuroscience to exciting moral conclusions. An argument from such findings to any kind of interesting moral conclusion will need some philosophical premises, and these will often be controversial. But this doesn't show that such findings are irrelevant to moral epistemology. Arguments deploying such premises will be controversial and open to question—which is just to say that they will be no different than most arguments in moral epistemology.

Notes

1. One of us argues elsewhere (Demaree-Cotton, 2016) that support for one popular kind of skeptical argument for the general unreliability of moral judgment on the basis of results from cognitive science—namely, skeptical arguments appealing to findings that moral judgments are influenced by morally irrelevant ways of presenting information—has been overstated.
2. These models are also discussed in Chapters 1, 2, 5, 6, 7, 8, 9 and 16 of this volume.
3. The interaction of emotion and reasoning is also discussed at length in Chapter 7 of this volume.
4. Moral learning is also discussed at length in Chapter 6 of this volume.
5. For further discussion see Chapters 13, 17, 18 and 19 of this volume, where a variety of concepts of epistemic justification are analyzed along with the relation of reliability to them.
6. On externalism about moral justification (Shafer-Landau, 2003), unreliability may directly entail lack of justification. On internalism, unreliable moral judgments may be justified if we aren't aware of this unreliability.
7. These views can take very different forms e.g. Arpaly, 2003; Hills, 2010. See too Chapter 25 of this volume, where Hills argues that moral worth depends on understanding the reasons why what one is doing are good, moral, or just.
8. See Dennett, 2006, on the personal/subpersonal distinction.
9. See Chapters 13 and 14 of this volume for skeptical philosophical perspectives on such demonstrations of unreliability/reliability.
10. E.g., Berker, 2009; Kamm, 2009.
11. E.g., Alston, 1995; Comesaña, 2006. See Beebe, 2004, for an explicit argument for the relevance of psychological processes over physical realizers.
12. Cf. Davis, 2009, 35.
13. Beebe (2004) and Davis (2009) appeal to multiple realizability to argue that belief-forming processes should be specified psychologically, not physically.
14. We qualify this with "in a given context" because it is possible for a given psychological process that is reliable in one context to be unreliable in another context; in such a case you might have the same neural process supporting a psychological process that is reliable in one context but unreliable in another.
15. Greene (2016, 132, and fn.9) criticizes Berker (2009) for assuming Greene attempts to draw normative conclusions from neuroscience directly.
16. For further discussion of why neuroscientific findings are of limited normative significance see Kahane, 2016. The work of Kohlberg and his students and colleagues is discussed in Chapters 1, 2, 5 and 6 of this volume.

17. Dwyer, 1999; Harman, 2008; Hauser et al., 2008; Mikhail, 2007, 2011. See too Chapter 2 of this volume.
18. Haidt, 2001, 2012. See too Chapters 1, 2, 7, 8, 9 and 16 of this volume.
19. Greene, 2008, 2016; Greene et al., 2001.
20. Other approaches defending the centrality of immediate emotional responses are Nichols, 2002 and Prinz, 2006.
21. E.g., Greene et al., 2004, 389; Greene & Haidt, 2002, Box 1. For an argument for many such modules, see Chapter 9 of this volume.
22. Bzdok et al., 2015; Greene, 2015; Greene & Haidt, 2002; Schaich Borg et al., 2006; Young & Dungan, 2012.
23. See previous note. Also, Greene, 2015, 198; Pascual et al., 2013.
24. See especially Haidt & Joseph, 2011, 2118–2119.
25. A similar hypothesis is defended in Chapter 2 of this volume.
26. Street's argument is discussed at length in Chapter 12 of this volume.
27. Parfit, 2011, 492–497. See also de Lazari-Radek & Singer, 2012.
28. See Chapters 10 and 11 of this volume for the relevant history.
29. See Chapter 17 of this volume on moral perception and Chapter 18 on intuitions; both chapters analyze the role played by emotion in moral judgment. See too Chapter 7 of this volume on emotion and reasoning more generally.
30. The TPJ is sometimes referred to as the pSTS.
31. Decety & Cacioppo, 2012; also, Gui et al., 2016; Yoder & Decety, 2014.
32. See previous note.
33. See Li, Mai & Liu, 2014.
34. See Elliott et al., 2011, for a review.
35. Similarly, see Rand et al., 2014.
36. See Railton, 2017, especially 5.2.
37. See Saunders, 2015, for a critique of accounts of moral reasoning in the CSM.
38. The definition of "innateness" is a vexed issue in cognitive science. See Griffiths, 2009.
39. Evidence for normative understandings among chimps and other primates (Chapter 3 of this volume) opens up the possibility that aspects of moral cognition are both evolved and learned (i.e. naturally selected and culturally transmitted). For a general overview of the evolution of human moral psychology, see Chapter 9 of this volume.
40. See Chapters 12 and 13 of this volume.
41. See Chapter 9 of this volume for evolutionary explanations of moral intuitions about family obligation, incest, and a suite of phenomena related to cooperation.

References

Allman, J. and Woodward, J. (2008). "What Are Intuitions and Why Should We Care About Them? A Neurobiological Perspective," *Philosophical Issues*, 18, 164–185.

Alston, W. P. (1995). "How to Think About Reliability," *Philosophical Topics*, 23, 1–29.

Arpaly, N. (2003). *Unprincipled Virtue: An Inquiry Into Moral Agency*. New York: Oxford University Press.

Ayars, A. (2016). "Can Model-Free Reinforcement Learning Explain Deontological Moral Judgments?" *Cognition*, 150, 232–242.

Barrett, L. F. and Satpute, A. B. (2013). "Large-Scale Brain Networks in Affective and Social Neuroscience: Towards an Integrative Functional Architecture of the Human Brain," *Current Opinion in Neurobiology*, 23, 361–372.

Bartels, D. M. and Pizarro, D. A. (2011). "The Mismeasure of Morals: Antisocial Personality Traits Predict Utilitarian Responses to Moral Dilemmas," *Cognition*, 121, 154–161.

Beebe, J. R. (2004). "The Generality Problem, Statistical Relevance and the Tri-Level Hypothesis," *Noûs*, 38, 177–195.

Berker, S. (2009). "The Normative Insignificance of Neuroscience," *Philosophy & Public Affairs*, 37, 293–329.

Bzdok, D., Groß, D. and Eickhoff, S. B. (2015). "The Neurobiology of Moral Cognition: Relation to Theory of Mind, Empathy, and Mind-Wandering," in J. Clausen and N. Levy (eds.), *Handbook of Neuroethics*. Dordrecht: Springer, 127–148.

Campbell, R. and Kumar, V. (2012). "Moral Reasoning on the Ground," *Ethics*, 122, 273–312.

Chakroff, A. and Young, L. (2015). "How the Mind Matters for Morality," *AJOB Neuroscience*, 6, 41–46.

Chiong, W., Wilson, S. M., D'Esposito, M., Kayser, A. S., Grossman, S. N., Poorzand, P., Seeley, W. W., Miller, B. L. and Rankin, K. P. (2013). "The Salience Network Causally Influences Default Mode Network Activity During Moral Reasoning," *Brain*, 136, 1929–1941.

Churchland, P. S. (2011). *Braintrust: What Neuroscience Tells Us About Morality*. Princeton: Princeton University Press.

Comesaña, J. (2006). "A Well-Founded Solution to the Generality Problem," *Philosophical Studies*, 129, 27–47.

Crockett, M. (2013). "Models of Morality," *Trends in Cognitive Sciences*, 17, 363–366.

Cushman, F. (2013). "Action, Outcome, and Value: A Dual-System Framework for Morality," *Personality and Social Psychology Review*, 17, 273–292.

Cushman, F., Murray, D., Gordon-McKeon, S., Wharton, S., and Greene, J. K. (2011). "Judgment before principle: engagement of the frontoparietal control network in condemning harms of omission," *Social Cognitive and Affective Neuroscience*, 7, 888–895.

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: G.P. Putnam.

Davis, J. K. (2009). "Subjectivity, Judgment, and the Basing Relationship," *Pacific Philosophical Quarterly*, 90, 21–40.

de Lazari-Radek, K. and Singer, P. (2012). "The Objectivity of Ethics and the Unity of Practical Reason," *Ethics*, 123, 9–31.

de Sousa, R. (2014). "Emotion," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition). https://plato.stanford.edu/archives/spr2014/entries/emotion/

Decety, J. and Cacioppo, S. (2012). "The Speed of Morality: A High-Density Electrical Neuroimaging Study," *Journal of Neurophysiology*, 108, 3068–3072.

Decety, J., Chen, C., Harenski, C. L. and Kiehl, K. A. (2015). "Socioemotional Processing of Morally-Laden Behavior and Their Consequences on Others in Forensic Psychopaths," *Human Brain Mapping*, 36, 2015–2026.

Demaree-Cotton, J. (2016). "Do Framing Effects Make Moral Intuitions Unreliable?" *Philosophical Psychology*, 29, 1–22.

Dennett, D. C. (2006). "Personal and Sub-Personal Levels of Explanation," in J. L. Bermúdez (ed.), *Philosophy of Psychology: Contemporary Readings*. London: Routledge, 17–21.

Dwyer, S. (1999). "Moral Competence," in K. Murasugi and R. Stainton (eds.), *Philosophy and Linguistics*. Boulder, CO: Westview Press, 169–190.

Dwyer, S., Huebner, B. and Hauser, M. D. (2010). "The Linguistic Analogy: Motivations, Results, and Speculations," *Topics in Cognitive Science*, 2, 486–510.

Everett, J. A. C., Pizarro, D. A. and Crockett, M. J. (2016). "Inference of Trustworthiness from Intuitive Moral Judgments," *Journal of Experimental Psychology: General*, 145, 772–787.

Elliott, R., Zahn, R., Deakin, W. J. and Anderson, I. M. (2011). "Affective Cognition and its Disruption in Mood Disorders," *Neuropsychopharmacology*, 36, 153–182.

FeldmanHall, O., Dalgleish, T., Thompson, R., Evans, D., Schweizer, S. and Mobbs, D. (2012). "Differential Neural Circuitry and Self-Interest in Real vs Hypothetical Moral Decisions," *Social Cognitive and Affective Neuroscience*, 7, 743–751.

Gleichgerrcht, E., Torralva, T., Roca, M., Pose, M. and Manes, F. (2011). "The Role of Social Cognition in Moral Judgment in Frontotemporal Dementia," *Social Neuroscience*, 2, 113–122.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. and Ditto, P. H. (2012). "Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism," *Advances in Experimental Social Psychology*, 47, 55–130.

Greene, J. D. (2007). "Why Are VMPFC Patients More Utilitarian? A Dual-Process Theory of Moral Judgment Explains," *Trends in Cognitive Sciences*, 11, 322–323.

———. (2008). "The Secret Joke of Kant's Soul," in W. Sinnott-Armstrong (ed.), *Moral Psychology, Vol. 3: The Neuroscience of Morality: Emotion, Disease, and Development*. Cambridge, MA: MIT Press, 35–79.

———. (2015). "The Cognitive Neuroscience of Moral Judgment and Decision Making," in J. Decety and T. Wheatley (eds.), *The Moral Brain*. Cambridge, MA: MIT Press, 197–220.

———. (2016). "Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics," in S. M. Liao (ed.), *Moral Brains: The Neuroscience of Morality*. New York: Oxford University Press, 119–149.

Greene, J. D. and Haidt, J. (2002). "How (and Where) Does Moral Judgment Work?" *Trends in Cognitive Sciences*, 6, 517–523.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M. and Cohen, J. D. (2004). "The Neural Bases of Cognitive Conflict and Control in Moral Judgment," *Neuron*, 44, 389–400.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M. and Cohen, J. D. (2001). "An fMRI Investigation of Emotional Engagement in Moral Judgment," *Science*, 293, 2105–2108.

Griffiths, P. (2009). "The Distinction Between Innate and Acquired Characteristics," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2009 Edition). https://plato.stanford.edu/archives/fall2009/entries/innate-acquired/

Gui, D. Y., Gan, T. and Liu, C. (2016). "Neural Evidence for Moral Intuition and the Temporal Dynamics of Interactions Between Emotional Processes and Moral Cognition," *Social Neuroscience*, 11, 380–394.

Haidt, J. (2001). "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review*, 108, 814–834.

———. (2012). *The Righteous Mind: Why Good People are Divided by Politics and Religion*. New York: Pantheon Books.

Haidt, J. and Joseph, C. (2004). "Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues," *Daedalus*, 133, 55–66.

———. (2011). "How Moral Foundations Theory Succeeded in Building on Sand: A Response to Suhler and Churchland," *Journal of Cognitive Neuroscience*, 23, 2117–2122.

Harenski, C. L., Antonenko, O., Shane, M. S. and Kiehl, K. A. (2010). "A Functional Imaging Investigation of Moral Deliberation and Moral Intuition," *NeuroImage*, 49, 2707–2716.

Harenski, C. L., Harenski, K. A., Shane, M. S. and Kiehl, K. A. (2010). "Aberrant Neural Processing of Moral Violations in Criminal Psychopaths," *Journal of Abnormal Psychology*, 119, 863–874.

Harman, G. (2008). "Using a Linguistic Analogy to Study Morality," in W. Sinnott-Armstrong (ed.), *Moral Psychology, Volume 1: Evolution of Morality—Adaptation and Innateness*. Cambridge, MA: MIT Press, 345–352.

Hauser, M. D. (2006). "The Liver and the Moral Organ," *SCAN*, 1, 214–220.

Hauser, M. D. and Young, L. (2008). "Modules, Minds and Morality," in D. W. Pfaff, C. Kordon, P. Chanson and Y. Christen (eds.), *Hormones and Social Behavior*. Berlin, Heidelberg: Springer, 1–11.

Hauser, M. D., Young, L. and Cushman, F. (2008). "Reviving Rawls's Linguistic Analogy: Operative Principles and the Causal Structure of Moral Actions," in W. Sinnott-Armstrong (ed.), *Moral Psychology, Vol. 2: The Cognitive Science of Morality: Intuition and Diversity*. Cambridge, MA: MIT Press, 107–143.

Hills, A. (2010). *The Beloved Self: Morality and the Challenge from Egoism*. Oxford: Oxford University Press.

Hoppenbrouwers, S. S., Bulten, B. H. and Brazil, I. A. (2016). "Parsing Fear: A Reassessment of the Evidence for Fear Deficits in Psychopathy," *Psychological Bulletin*, 142, 573–600.

Huebner, B. (2015). "Do Emotions Play a Constitutive Role in Moral Cognition?" *Topoi*, 34, 427–440.

Hutcherson, C. A., Montaser-Kouhsari, L., Woodward, J. and Rangel, A. (2015). "Emotional and Utilitarian Appraisals of Moral Dilemmas Are Encoded in Separate Areas and Integrated in Ventromedial Prefrontal Cortex," *The Journal of Neuroscience*, 35, 12593–12605.

Jaggar, A. M. (1989). "Love and Knowledge: Emotion in Feminist Epistemology," *Inquiry*, 32, 151–176.

Kahane, G. (2011). "Evolutionary Debunking Arguments," *NOUS*, 45, 103–125.
———. (2014). "Intuitive and Counterintuitive Morality," in J. D'Arms and D. Jacobson (eds.), *Moral Psychology and Human Agency: Philosophical Essays on the Science of Ethics*. Oxford: Oxford University Press, 9–39.
———. (2016). "Is, Ought, and the Brain," in S. M. Liao (ed.), *Moral Brains: The Neuroscience of Morality*. Oxford: Oxford University Press, 281–311.
Kahane, G., Everett, J. A. C., Earp, B. D., Farias, M. and Savulescu, J. (2015) "'Utilitarian' Judgment in Sacrifical Moral Dilemmas Do Not Reflect Impartial Concern for the Greater Good," *Cognition*, 134, 193–209.
Kahane, G., Katja, W., Shackel, N., Farias, M., Savulescu, J. and Tracey, I. (2012). "The Neural Basis of Intuitive and Counterintuitive Moral Judgment," *SCAN*, 7, 393–402.
Kamm, F. (2009). "Neuroscience and Moral Reasoning; A Note on Recent Research," *Philosophy & Public Affairs*, 37, 330–345.
Kiehl, K. A. (2008). "A Reply to de Oliveira-Souza, Ignácio, and Moll and Schaich Borg," in W. Sinnott-Armstrong (ed.), *Moral Psychology, Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*. Cambridge, MA: MIT Press, 165–171.
LeDoux, J. E. (2012). "A Neuroscientist's Perspective on Debates About the Nature of Emotion," *Emotion Review*, 4, 375–379.
Li, W., Mai, X. and Liu, C. (2014). "The Default Mode Network and Social Understanding of Others: What Do Brain Connectivity Studies Tell Us," *Frontiers in Human Neuroscience*, 8. doi:10.3389/fnhum.2014.00074.
Lindquist, K. A. and Barrett, L. F. (2012). "A Functional Architecture of the Human Brain: Emerging Insights from the Science of Emotion," *Trends in Cognitive Sciences*, 16, 533–540.
Little, M. O. (1995). "Seeing and Caring: The Role of Affect in Feminist Moral Epistemology," *Hypatia*, 10, 117–137.
Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin Classics.
Marsh, A. A., Stoycos, S. A., Brethel-Haurwitz, K. M., Robinson, P., VanMeter, J. W. and Cardinale, E. M. (2014). "Neural and Cognitive Characteristics of Extraordinary Altruists," *PNAS*, 111, 15036–15041.
Mikhail, J. (2007). "Universal Moral Grammar: Theory, Evidence and the Future," *Trends in Cognitive Sciences*, 11, 143–152.
———. (2011). *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. New York: Cambridge University Press.
Miller, R. and Cushman, F. (2013). "Aversive for Me, Wrong for You: First-Person Behavioral Aversions Underlie the Moral Condemnation of Harm," *Social and Personality Psychology Compass*, 7, 707–718.
Moll, J., De Oliveira-Souza, R. and Zahn, R. (2008). "The Neural Basis of Moral Cognition: Sentiments, Concepts, and Values," *Annals of the New York Academy of Sciences*, 1124, 161–180.
Nado, J. (2014). "Why Intuition?" *Philosophy and Phenomenological Research*, 89, 15–41.
Nichols, S. (2002). "Norms with Feeling: Towards a Psychological Account of Moral Judgment," *Cognition*, 84, 221–236.
Nichols, S., Kumar, S., Lopez, T., Ayars, A. and Chan, H-Y. (2016). "Rational Learners and Moral Rules," *Mind & Language*, 31, 530–554.
Okon-Singer, H., Hendler, T., Pessoa, L. and Shackman, A. J. (2015). "The Neurobiology of Emotion-Cognition Interactions: Fundamental Questions and Strategies for Future Research," *Frontiers in Human Neuroscience*, 9, 1–14. doi:10.3389/fnhum.2015.00058.
Parfit, D. (2011). *On What Matters: Volume Two*. New York: Oxford University Press.
Pascual, L., Gallardo-Pujol, D., and Rodrigues, P. (2013). "How does morality work in the brain? A functional and structural perspective of moral behavior." *Frontiers in Integrative Neuroscience,* 7, 1–8.
Pessoa, L. (2008). "On the Relationship Between Emotion and Cognition," *Nature Reviews Neuroscience*, 9, 148–158.
———. (2013). *The Cognitive-Emotional Brain: From Interactions to Integration*. Cambridge, MA: MIT Press.

Pessoa, L. and Adolphs, R. (2010). "Emotion Processing and the Amygdala: From a 'Low Road' to 'Many Roads' of Evaluating Biological Significance," *Nature Reviews Neuroscience*, 11, 773–782.
Prinz, J. (2006). "The Emotional Basis of Moral Judgments," *Philosophical Explorations*, 9, 29–43.
Railton, P. (2017). "Moral Learning: Why Learning? Why moral? And Why Now?" *Cognition*. 167: 172–190.
Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A. and Greene, J. D. (2014). "Social Heuristics Shape Intuitive Cooperation," *Nature Communications*, 5, Article 3677.
Ruse, M. (1988). "Evolutionary Ethics: Healthy Prospect or Last Infirmity," *Canadian Journal of Philosophy*, 14 (Supp), 27–73.
Saunders, L. F. (2015). "What Is Moral Reasoning?" *Philosophical Psychology*, 28, 1–20.
Schaich Borg, J., Hynes, C., Van Horn, J., Grafton, S. and Sinnott-Armstrong, W. (2006). "Consequences, Action, and Intention as Factors in Moral Judgments: An fMRI Investigation," *Journal of Cognitive Neuroscience*, 18, 803–817.
Schaich Borg, J., Sinnott-Armstrong, W., Calhoun, V. D. and Kiehl, K. A. (2011). "Neural Basis of Moral Verdict and Moral Deliberation," *Social Neuroscience*, 6, 398–413.
Shafer-Landau, R. (2003). *Moral Realism: A Defence*. Oxford: Oxford University Press.
Shamay-Tsoory, S. G. and Aharon-Peretz, J. (2007). "Dissociable Prefrontal Networks for Cognitive and Affective Theory of Mind: A Lesion Study," *Neuropsychologia*, 45, 3054–3067.
Shenhav, A. and Greene, J. D. (2014). "Integrative Moral Judgment: Dissociating the Roles of the Amygdala and Ventromedial Prefrontal Cortex," *The Journal of Neuroscience*, 34, 4741–4749.
Singer, P. (2005). "Ethics and Intuitions," *The Journal of Ethics*, 9, 331–352.
Street, S. (2006). "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies*, 127, 109–166.
Suhler, C. L. and Churchland, P. (2011). "Can Innate, Modular 'Foundations' Explain Morality? Challenges for Haidt's Moral Foundations Theory," *Journal of Cognitive Neuroscience*, 23, 2103–2116.
Vavova, K. (2015). "Evolutionary Debunking of Moral Realism," *Philosophy Compass*, 10, 104–116.
Yoder, K. J. and Decety, J. (2014). "Spatiotemporal Neural Dynamics of Moral Judgment: A High-Density ERP Study," *Neuropsychologia*, 60, 39–45.
Young, L., Bechara, A., Tranel, D., Damasio, H., Hauser, M. and Damasio, A. (2010). "Damage to Ventromedial Prefrontal Cortex Impairs Judgment of Harmful Intent," *Neuron*, 65, 845–851.
Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A. and Saxe, R. (2010). "Disruption of the Right Temporoparietal Junction with Transcranial Magnetic Stimulation Reduces the Role of Beliefs in Moral Judgments," *Proceedings of the National Academy of Sciences*, 107, 6753–6758.
Young, L. and Dungan, J. (2012). "Where in the Brain Is Morality? Everywhere and Maybe Nowhere," *Social Neuroscience*, 7, 1–10.

Further Readings

For overviews of the highly influential "first wave" approaches to the cognitive science of morality, see Haidt, 2001 (for Haidt's social intuitionist model) and Graham et al., 2012 (for moral foundations theory, a development of the SIM approach); Greene (2008, 2016) (for Greene's dual-process theory and his argument that evidence for the theory has implications for normative ethics); and Dwyer, 1999, and Mikhail, 2007 (for introductions to the universal moral grammar approach). For arguments criticizing Greene's claims regarding the relevance of neuroscience to ethics, see Berker (2009) and Kamm (2009), as well as Kahane (2014, 2016). See Suhler and Churchland (2011) for the claim that neuroscientific and neurobiological evidence counts against psychological claims made by Haidt's work, including those regarding domain-specificity and innateness, and see Haidt and Joseph's response (2011) for the argument that neurobiological evidence cannot refute their psychological theory. See Huebner (2015) for an argument that the functions performed by the brain in moral judgment cannot be classed as either "emotional" or "cognitive". See Railton (2017) for an in-depth overview of current neuroscientific and other

evidence pertaining to new moral learning approaches and an argument that they may vindicate the rationality of moral judgment.

Related Chapters

Chapter 1 The Quest for the Boundaries of Morality; Chapter 2 The Normative Sense: What is Universal? What Varies? Chapter 5 Moral Development in Humans; Chapter 6 Moral Learning; Chapter 7 Moral Reasoning and Emotion; Chapter 8 Moral Intuitions and Heuristics; Chapter 9 The Evolution of Moral Cognition; Chapter 15 Relativism and Pluralism in Moral Epistemology; Chapter 16 Rationalism and Intuitionism: Assessing Three Views about the Psychology of Moral Judgment; Chapter 20 Methods, Goals, and Data in Moral Theorizing; Chapter 21 Moral Theory and its Role in Everyday Moral Thought and Action; Chapter 22 Moral Knowledge as Know-How.