Journal of the American Philosophical Association (2018) 411–431 © American Philosophical Association
DOI:10.1017/apa.2018.25

Normative Moral Neuroscience: The Third Tradition of Neuroethics

ABSTRACT: *Neuroethics is typically conceived of as consisting of two traditions: the ethics of neuroscience and the neuroscience of moral judgment. However, recent work has sought to draw philosophical and ethical implications from the neuroscience of moral judgment. Such work, which concerns* normative moral neuroscience *(NMN), is sufficiently distinct and complex to deserve recognition as a third tradition of neuroethics. Recognizing it as such can reduce confusion among researchers, eliminating conflations among both critics and proponents of NMN.*

This article identifies and unpacks some of the most prominent goals, characteristic assumptions, and unique arguments in NMN and addresses some of the strongest objections NMN faces. The paper synthesizes these considerations into a set of heuristics, or loose discovery principles, that can help overcome obstacles in and attenuate resistance to NMN. These heuristics may simultaneously help identify those projects in NMN that are most likely to be fruitful and help fortify them.

KEYWORDS: cognitive science, experimental philosophy, metaphilosophy, moral psychology, neuroethics, neurophilosophy

Introduction

The neuroscience of moral judgment is the study of the relationship between the nervous system—especially the brain—and moral cognition, emotion, motivation, and behavior. Normative ethics is the study of how human beings ought to think, feel, and act; it is the study of right and wrong. Metaethics, which is normative in the broader sense in that it concerns the establishment of prescriptive standards for what counts as ethics,[1] is the study of the nature of morality. It seems clear that the brain is largely responsible for producing moral cognition, emotion, motivation, and behavior—though partial responsibility for these functions may also lie in other areas of the nervous system (see, e.g., Bechara and Damasio 2005). Thus, the neuroscience of moral judgment is viable as a descriptive project.

I am grateful to Michael Davis, Elisabeth Hildt, Warren Schmaus, and two blind reviewers for their insightful comments on previous drafts of this manuscript.

[1] As Kauppinen (2014) points out, concepts are inherently normative, and it is this normativity that makes possible discussions of how our actual judgments or responses compare to those judgments or responses we should make—even if only according to our own lights. It is only because of this normative dimension of metaethics that heated debates about the competent or errant deployment of moral concepts (sections 2–3) are even possible.

https://doi.org/10.1017/apa.2018.25 Published online by Cambridge University Press

But can studying exactly how the brain shapes these functions teach us anything about what is right and what is wrong or about the nature of morality more broadly? That is, can neuroscientific study not only produce descriptions of how we *actually* think, feel, and act, but can it also help us understand how we *should* think, feel, and act?

If it can, then the neuroscience of moral judgment has normative implications. By this, I mean that the development of a normative moral neuroscience (NMN) could, at least in principle, help guide philosophers toward genuine applied or even theoretic knowledge in normative ethics and metaethics. If the neuroscience of moral judgment has what I call *negative* normative implications, then it could contribute such guidance by helping us understand certain human limitations that determine how we cannot and therefore (arguably) are not obliged to think, feel, and act. Moreover, if the neuroscience of moral judgment has *positive* normative implications, it could help determine more than just how and to whom theoretic moral knowledge can be normatively applied. Neuroscience could, if positive projects in NMN are feasible, help adjudicate between general theories in moral philosophy.

If none of this is possible, then the neuroscience of moral judgment does not have normative implications. It is not my aim in this essay to adjudicate between the possibility that the neuroscience of moral judgment has direct normative implications—of a negative kind, a positive kind, or both—and the possibility that it does not. Nor is it my aim to argue about the relative value (or lack of value) of neuroscience for applied ethics, normative ethics, and metaethics (Kauppinen 2014) though in this essay its value may be most apparent for metaethics. Instead of engaging in either of these secondary debates, I argue that neuroethicists can draw valuable lessons by thinking of NMN as a third tradition of neuroethics, distinct from both the neuroscience of moral judgment and the bioethics of neuroscience (section 1).

In fact, recognizing it as such can help substantially reduce confusion among researchers, eliminating conflations among both critics and proponents of NMN. If NMN is to succeed, an organized attempt to understand and draw lessons from its modi operandi and its previous accomplishments (section 2) as well as from its shortcomings and vulnerabilities (section 3) may prove crucial. It will also be valuable to synthesize these lessons into a set of loose criteria (section 4) that NMN should, for the time being at least, seek to satisfy in order to minimize its susceptibility to common objections. To conclude (section 5), I will point out that even if neuroscience has no *direct* bearing on normative ethics, neuroscientific findings may nonetheless provide important insights into a number of foundational concepts on which philosophical ethics rests. Thus, neuroscience has important implications for our philosophical understanding of normative ethics and metaethics, even if these implications may be *indirect*.

The most salient of the criteria that I discuss in section 4, *schematic neutrality*, aims to redress issues caused by a priori theoretic assumptions in supposedly objective assessments of (neuro)scientific data. In many cases in which NMN theorists purport to draw conclusions from empirical data, those conclusions are in fact determined to a large extent not by the data, but by the a priori interpretive

https://doi.org/10.1017/apa.2018.25 Published online by Cambridge University Press

framework within which the research is conducted. In the sciences, the presupposition of frameworks for interpreting data is a valuable and arguably necessary step in gaining a foothold on certain problems. However, such frameworks tend to exclude automatically many possible answers to research questions out of practical necessity, rather than because those answers are known to be incorrect. By relying on interpretive frameworks in analyzing data, researchers in a sense end up testing the ability of their a priori frameworks to accommodate those data rather than exhaustively searching for the best interpretation of those data. And so researchers in NMN often risk doing little more than circularly evaluating the research frameworks they presumed a priori.

While this risk may occur somewhat broadly in science (Quine 1951), it poses especially difficult challenges for NMN. Unlike, say, most biologists or even those working in other areas of neuroscience, researchers working within empirical philosophy—including within the third tradition of neuroethics—seek not only to evaluate data relative to their research frameworks, but simultaneously to evaluate those frameworks themselves. This constant risk of circularity and the need to minimize it are defining features of NMN and they are sufficiently important to and salient in NMN as to deserve significant consideration in their own right. To study NMN, one must simultaneously adopt and disavow some set of theoretic commitments. This dual obligation is largely unfamiliar in both philosophy and neuroscience, so NMN may not be best conceived or practiced either as traditional (neuro)science or traditional philosophy.

1. The Third Tradition of Neuroethics

Many questions within the first tradition of neuroethics are, fundamentally, about the function of the brain. Roskies (2002) defines this first tradition of neuroethics as concerning the neuroscience of moral judgment.[2] In contrast, questions within the second tradition are often about the practical application of moral principles—as Roskies puts it, 'the ethical implications of neuroscience' (2002: 21). But the questions I am concerned with in this essay cross these boundaries and are, fundamentally, about the nature of morality. Questions within what I am calling NMN concern 'the conceptual interconnections and feedback between the two main divisions of neuroethics' (Roskies 2002: 23). The third tradition of neuroethics concerns the philosophical and ethical implications of the neuroscience of moral judgment.

Although questions of this kind are often lumped in with those in the first tradition, this categorization is not always the most useful. It is for this reason that a clearer recognition of what I am calling the third tradition is called for. I therefore wish to outline this tradition and its distinction from the first two traditions here, as I believe that its recognition can be valuable to the advancement of at least some projects within neuroethics.

[2] Roskies (2002) refers to the first tradition as 'the neuroscience of ethics', but I will not use that phrase here. As I suggested in the introduction, it is clear that there exists a neuroscience of moral judgment and decision making, but it is not clear that there exists a neuroscience of ethics *per se*.

https://doi.org/10.1017/apa.2018.25 Published online by Cambridge University Press

First, many first-tradition questions regarding the neuroscience of moral judgment seek *descriptive* answers. Questions regarding how moral decisions are made, how values are represented, and how similar or dissimilar ethical decision making might be to other types of decision making (Roskies 2002) are, in a sense, all questions about how the brain actually generates our moral judgments. But to ask how neuroscientific understanding will modify our philosophical ethical framework is to ask a normative question, insofar as philosophical ethics is (by definition) bound by an imperative to seek out wisdom. It is to ask how neuroscientific understanding can fulfill this imperative and thus how it *should* modify our philosophical ethical framework. Thus, we cannot assume that studying the neuroscience of moral judgment will necessarily teach us anything about philosophical ethics. To assume as much is to conflate the etiology of (im)moral behavior with the ontology of morality. It is to mistake the way moral judgment *is* made for the way it *ought* to be made. This *is-ought problem*, as it is sometimes called, is at the heart of many poignant critiques of NMN (section 2.2).

Second, many of the best-known questions about the neuroscience of moral judgment can be answered within a single specific scientific paradigm (Kuhn [1962] 2012)—namely, the paradigm of contemporary neuroscience. But for findings in psychology or neuroscience or biology to have philosophical implications, those findings need to be meaningful across paradigms. This is because the goals of the physical (neuro)scientist and those of the working philosopher differ greatly (Einstein 1949: 683–84).

Finally, the *how* in many first-tradition questions is (ostensibly) well-grounded in a *that*. Perhaps we can be sure *that* moral decisions are made in the brain, *that* values are represented, and *that* there are either analogies or disanalogies or both between moral and nonmoral decision-making.[3] But it is not at all clear *that* biological understanding has any bearing on which philosophical ethical frameworks we should endorse.

Of course, the framing of NMN as a third tradition of neuroethics assumes that there are two other divisions of neuroethics distinct not only from the third tradition, but also from each other. But even Roskies, who introduced these two other divisions, acknowledges that the putative distinction between them may be a useful conceit rather than a precise description of the external world: ‘The conceptual interconnections and feedback between the two main divisions of neuroethics are dense enough that it may be that distinctions between them can only be made roughly, and only in theory’ (Roskies 2002: 23).

Like Roskies, I am content to set aside questions of ontology for now in order to facilitate greater understanding. If nothing else, an outline of the third tradition of neuroethics can play a valuable role in the scaffolding of future neuroethics research, just as Roskies's outlines of the first and second traditions have done for

[3] Even these ostensibly well-grounded claims are uncertain. One could argue, for instance, that moral values are never actually represented by decision makers and are only convenient fictions constructed post hoc to describe behaviors and dispositions whose true etiologies remain unknown. Likewise, it does not seem to be the case that moral decision making is encapsulated entirely in the brain (Bechara and Damasio 2005; Prinz 2004; Schnall et al. 2008).

https://doi.org/10.1017/apa.2018.25 Published online by Cambridge University Press

research to date. Moreover, the third tradition of neuroethics—regardless of whether its borders with the first and second traditions are real or artificial—encompasses many of the most novel questions in neuroethics. Much of the first tradition is focused on neuroscience per se, with only a secondary emphasis on understanding the moral phenomena associated with that neuroscience. Much of the second tradition can be thought of as a new but nonetheless subsidiary part of bioethics. But the third tradition truly focuses on the consequences of neuroscience for philosophical ethics, and it is therefore deeply and self-consciously interdisciplinary in a way that the first two traditions often are not.

It is largely for this reason that there is utility in thinking of the sort of inquiry I focus on here as constituting a third tradition of neuroethics, a tradition concerning the ethical implications of the neuroscience of moral judgment. The kinds of arguments mounted within this tradition and the problems and objections they face are, in an important sense, unique. But in another sense, they pervade almost all corners of neuroethics in some way. My goal in the remainder of this section—and indeed, one of my primary reasons for writing this essay—is to make this clear.

1.1. Two Interpretive Problems of the Third Tradition

When faced with evidence of the ways people process ethical quandaries, NMN theorists are often confronted with two interpretive choices. First, they must decide whether moral processing[4] tends to function effectively (reflecting core human *competence*) or is prone to dysfunction (leading to *error* in judgment). *Competence theorists* contend that ordinary moral judgment tends to point toward normative moral judgment. In contrast, *error theorists* contend that ordinary moral judgment tends either to point away from normative moral judgment or to have no reliable relationship with normative moral judgment at all.

Second, NMN theorists must decide whether the study of ordinary moral judgment can provide *positive* evidence in favor of some normative ethical theory or only *negative* evidence against some theory. *Positive projects* assume that competent moral judgment tends to align with, and therefore support, certain views in the canon of philosophical ethics. Thus, positive projects in NMN are about substantiating normative ethical claims on the basis of empirical discoveries, an endeavor that, as noted, runs the risks associated with the is-ought problem.

Negative projects, on the other hand, assume just the opposite (Knobe et al. 2012). Negative projects in NMN are essentially about holding normative claims accountable to the descriptive realm of possibilities. Implicitly working from the premise that ought implies can, such projects typically seek to invalidate claims that we should behave (act, judge) in certain ways by revealing that we cannot behave (act, judge) in those ways (e.g., Doris 1998). Perhaps the

[4] I use the term moral 'processing' to refer to the general phenomenon that some authors call moral 'cognition' (Greene and Haidt 2002). I do so to avoid confusion with a second, narrower sense of moral 'cognition', which is a specific form of moral processing that many theorists contrast with moral 'emotion' (e.g., Greene 2007).

https://doi.org/10.1017/apa.2018.25 Published online by Cambridge University Press

most notable advantage of negative projects over positive ones is their intuitive plausibility.

For instance, to many people it seems patently unfair to hold adolescents, the cognitively impaired, animals, or indeed anyone at all to standards of behavior that they are neurophysiologically incapable of upholding. This sole fact, it could be argued, might be sufficient to defeat any appearance of moral obligation among those individuals. By contrast, it seems patently untrue that we should hold anyone to a given standard of behavior solely because they are capable of—or perhaps even inclined toward—upholding it.

Of course, the principle that ought implies can is not without issue. Some philosophers deny that moral obligation is bounded in this way by physical or practical possibility (Kant [1785] 1956). Even without relying on exotic cases and thought experiments, one can see problems with this principle. For instance, if an individual is physiologically incapable of inhibiting his violent racist tendencies, this would not seem to excuse his violent racist behavior. One might argue that this person, like all people, *ought* to be egalitarian even though he *cannot* be egalitarian. Likewise, it could be argued that normative moralization of children is centrally concerned with the enforcement of standards that children are not yet equipped to uphold.

Despite the reasons to believe that descriptive facts about individuals' capacities cannot always be used to identify the boundaries of their normative obligations, negative projects seem at least to present narrower targets for objection, on average, than positive projects do. Whereas the distinction between competence and error theories concerns the (mis)alignment of ordinary moral judgment and normative moral judgment, the distinction between positive and negative projects concerns the (mis)alignment of normative moral judgment and specific views in philosophical ethics. However, researchers' evaluations of theoretic (mis) alignment are, in both cases, subject to the influence of a priori views held firmly by the researchers making those evaluations before any evidence has even been collected. Thus, views within NMN typically express two qualitative evaluations that cannot be based entirely on empirical evidence: an evaluation of ordinary moral judgment (competent or errant) and a stance toward some canonical or novel moral theory (positive or negative).

It is in large part against these two kinds of qualitative evaluations that NMN researchers receive pushback. In my concluding remarks (section 5), I will argue that NMN researchers can and should develop at least some research programs that avoid making any such a priori evaluative commitments. But for now, I wish to explore the two major types of pushback these evaluations tend to engender. This exploration may provide some insight into how NMN researchers might minimize the impact of such objections in the future.

1.2. Two Kinds of Objections to the Third Tradition

Objections to NMN seek to establish that the neuroscience of moral judgment is incapable of contributing meaningfully to normative understanding of philosophical ethics. Establishing this in the abstract would be rather difficult, so

https://doi.org/10.1017/apa.2018.25 Published online by Cambridge University Press

most critics have focused their attacks on specific projects within NMN. Objections to specific projects within NMN may be thought of as consisting of two main kinds: inferential and methodological.

Inferential objections claim that the gap between philosophical ethics (including metaethics) and the neuroscience of moral judgment is not (or cannot be) truly bridged by NMN. These objections rail against the relevance of neuroscience—which objectors view as a fundamentally descriptive discipline—to questions of normative ethics and metaethics. For instance, one inferential objection to NMN is that even if we knew for certain that all judgments of guilt were based in part on the perception of intent, this would not prove the *positive* thesis that intent is a necessary condition for actual guilt (more on this in section 2.2). Inferential objections concern issues such as the a priori adoption of normative ethical stances in the interpretation of empirical data, presumptions about the normative relevance of ordinary moral judgment, and prejudices about the relative worth of different kinds of neurological (and psychological) processes.

Methodological objections seek to discredit the neuroscience of moral judgment itself, at least as it is currently practiced. Such objections typically do not raise concerns about the relevance of neuroscientific descriptions to normative ethics—instead, they deny that those descriptions are accurate in the first place. For instance, Berker (2009) points out that a frequently used set of moral dilemmas meant to distinguish deontological from utilitarian judgment (Greene et al. 2001) fails to do so because those dilemmas are poorly designed (as I discuss in section 3.2). Methodological objections may be specific to the neuroscience of moral judgment, or they may critique neuroscience more generally. Among other things, such objections may impugn the soundness of neuroimaging techniques like fMRI, criticize the construction of experimental probes and materials, or question the ability to infer general principles from demographically homogenous samples (see Holtzman 2017).

While the distinction between methodological and inferential objections may be useful, it is not as hard-and-fast as the distinction between the two types of interpretive problems (section 1.1). This is because inferential and methodological objections sometimes overlap. A poorly designed questionnaire always presents a methodological problem, but if the poverty of its design lies in its conflation of, say, moral and conventional norm violations, then it is also inferentially problematic. Thus, the appropriate response to certain inferential problems may actually involve changes in methodology (section 4.2).

Furthermore, it is worth noting that while both kinds of objection may be especially challenging for NMN theorists, such challenges are not unique to NMN. For instance, philosophers wishing to leverage evolutionary theories for or against certain views in moral metaphysics (Street 2006) and epistemology (Kahane 2011) face similar inferential objections (Vavova 2014; Holtzman, forthcoming). And every science, not just neuroscience, faces some degree of scrutiny in the form of methodological objections. Still, since this essay aims to attenuate resistance to NMN, I will be returning to these themes throughout the essay.

https://doi.org/10.1017/apa.2018.25 Published online by Cambridge University Press

2. For and Against Competence Theories

2.1. In Favor of Positive and Negative Competence Theories

Gerd Gigerenzer touts the competence of ordinary moral judgment and believes that discoveries about the ways people make moral judgments weigh negatively against some canonical views in philosophical ethics. Thus, he is what I call a *negative competence theorist*. Gigerenzer (2008) argues that the human mind functions largely on the basis of heuristic processes that seek to satisfice, that is, to find good-enough solutions to problems. And these heuristic processes, he argues, are in many cases more effective than optimization procedures, which seek to find the best solutions. Unlike many of his contemporaries (e.g., Sunstein 2013), Gigerenzer does not just think that heuristics are more efficient than other approaches to decision making; he believes that in many cases, they are simply better (Gigerenzer and Selten 2002). For this reason, he thinks that we should view the brain as a paragon of moral decision making in at least some ways.

On these grounds, Gigerenzer mounts an empirical challenge to idealized utilitarian calculus as it is traditionally conceived (Posner 1979). Although Gigerenzer's (2010) 'aim is not to provide a normative theory' (530), he argues that findings in the cognitive sciences 'challenge the normative ideal that maximizing can generally define how people ought to behave' (529). Therefore, while his philosophical arguments are of a negative nature, Gigerenzer tends to view the processes that typically underwrite moral decision making as competent. Gigerenzer merely suggests that *if* utilitarianism is correct, *then* in many cases it may be morally better to satisfice (Simon 1956) than to maximize, but how best to satisfice and why we should believe utilitarianism to be correct, Gigerenzer does not say.

Kurt Gray is another noteworthy competence theorist in NMN, but unlike Gigerenzer, he takes research in neuroscience and psychology to reflect positively on at least some moral theories. His views center around two components of moral judgment: the perception of other minds and the ascription of mental states to them. In 'Mind Perception Is the Essence of Morality' (2012), Gray and his colleagues argue that 'moral judgments should be linked to perceptions of intention and suffering' (104) and that, at least in what they call prototypical cases, 'moral transgressions involve perceptions of suffering' (108). Gray and his colleagues express strong (albeit qualified, see section 2.2) views on what morality is and on the ways that normative ethical evaluations ought to be made. Thus, Gray suggests that we might be able to leverage descriptive neuroscientific findings about brain regions involved in ordinary moral judgment (e.g., Young and Saxe 2008) into dicta concerning normative moral theory.

In this way, Gray's *positive competence theory* differs importantly from Gigerenzer's negative competence theory. Both authors believe that ordinary moral judgment fundamentally reflects competent application of normative ethical concepts. But in Gigerenzer's negative view, the neuroscience of (competent) moral judgment can tell us only what kinds of moral judgments are not normative: maximizing happiness ≠ good, maximizing ≯ satisficing. In contrast, Gray's positive stance suggests that (competent) moral judgment may be evidence for what kinds of judgments are normative: intentionally causing suffering = bad, unintentionally causing suffering > intentionally causing suffering.

2.2. Against Positive and Negative Competence Theories

Considered together, the cases of Gigerenzer and Gray illustrate a pair of related points. First, positive projects within NMN are arguably less plausible in many cases than negative projects. Unlike Gigerenzer's negative project, Gray's positive project threatens to jump from neuroscientific evidence that certain factors affect the perception of moral worth to the suggestion that those same elements factor into normative moral worth. In that sense, Gray's argument may have far-reaching implications if one accepts it, but one might reject its implications due to inferential objections.[5] In this regard, his point is that whatever the philosophical merit of an emotion-versus-reason framework for *normative* moral evaluation might be, our *actual* moral evaluations operate largely on the basis of two kinds of perceptions of other minds. Gray argues that of these two kinds—the perception of intentional agents and the perception of suffering patients—each requires both emotion and reason (or, as he puts it, 'cooperation between affect and cognition' [Gray and Schein 2012: 419]). Thus, the common conception of moral judgment as emerging from a battle between emotion and reason cannot be brought to bear on empirical-descriptive studies of moral processing, and should (on Gray's view) be replaced by Gray's own mind-perception model.

The second point illustrated by this pair of cases is that negative projects within NMN are in many cases arguably more parochial—that is, limited in relevance to specific moral schemes—than positive projects within NMN. In contrast to Gray, Gigerenzer makes the somewhat staid claim that attempting felicific calculus (Bentham [1823] 2005) in the 'real, uncertain world' (Gigerenzer 2010: 536) is usually less efficient and sometimes even less accurate than undertaking what we might call 'felicific estimation' in such a world. In this regard, Gigerenzer's claims remain largely within the descriptive domain. Gigerenzer is right to say that 'whether optimizing or satisficing leads to better outcomes in the real, uncertain world is an empirical question' (Gigerenzer 2010: 536). But as Bruni and his colleagues (2014) point out, the extent to which normative ethics is concerned with better outcomes (whatever that might mean) is an open, nonempirical question. Whereas utilitarians think that normative ethics is entirely concerned with outcomes, deontologists, virtue ethicists, care theorists, and others disagree.

So, too, is it an open question whether normative ethics is concerned with the real, uncertain world (Nagel 1976), or only with idealized, nonstochastic expectations. If

[5] Fortunately, Gray clearly recognizes the obstacles he faces in drawing philosophical conclusions from neuropsychological findings. He therefore takes care to distance himself from any wholehearted endorsement of a normative philosophical interpretation of his work. Gray is careful to couch his work in the qualification that perhaps only its nonnormative conclusions are justified. He rightly concedes that his view may be most plausible as one about the inability to describe findings in moral neuroscience according to the frameworks provided by canonical moral theories—not as a view about the validity of those canonical moral theories themselves (Gray and Schein 2012). In this way, Gray preempts certain inferential objections to his work.

https://doi.org/10.1017/apa.2018.25 Published online by Cambridge University Press

philosophers like Kant ([1785] 1956) are right, then the real-world outcomes of our behaviors may be largely morally irrelevant. For Kant, all that matters are our intentions to act in accordance with the principles of duty. If Kantians—and many others—are right to take issue with utilitarianism, then Gigerenzer's point about the value of heuristics in improving real-world outcomes, even if correct from a practical standpoint, may have no bearing on normative ethics.

3. For and Against Error Theories

3.1. In Favor of Positive and Negative Error Theories

Whereas Gigerenzer and Gray both believe that ordinary moral processing tends be driven by underlying competencies that reflect normative ideals, the NMN researchers who have received the most attention have usually argued the opposite. They opine that ordinary moral processing is fundamentally flawed and that the moral faculties we have evolved are no more useful in many of today's contexts than, say, our appendices are. According to these researchers, a great many of the moral judgments people make tend to be erroneous. Nonetheless, these error theorists often think that empirical work can help us draw normative conclusions—either positive or negative. They believe that the neuroscience of moral judgment can reveal the poverty of ordinary moral processing. Negative error theorists argue that insofar as the conclusions reached by impoverished neural processes should not be trusted, neither should the normative ethical perspectives they underwrite. Positive error theorists push the point even further, claiming that by discounting errant moral judgments, we may be able to hone in on the correct, last-man-standing moral perspective (although see section 3.2 for objections; see also Kamm 2009; Meyers 2015).

Perhaps no NMN theorist has generated more controversy than Joshua Greene (2008, 2014a). At the turn of the century, Greene and his colleagues (2001, 2004) purported to show via fMRI that automatic processes such as emotion tend to underwrite rule-based ('deontological') moral judgments, whereas reason and controlled cognition tend to underwrite outcome-based ('utilitarian') judgments.[6] On these grounds and on the (rather large) assumption that reason is more trustworthy than emotion (cf. de Sousa 1990; Pham 2007), Greene has reputedly mounted the argument that normative moral judgment—and thus, moral right and wrong—tends (and ought) to be determined by outcomes rather than by rules. I am not convinced that Greene's argument is quite as simplistic as this, but this does seem to be the received view of his work.

Greene (2014a) is thus usually interpreted to be an error theorist, arguing that though our moral faculties are in many respects amazingly fast and accurate, they are far from perfect. Moreover, when engaged in situations that were common during the evolutionary history of humankind, our relatively efficient moral faculties are especially prone to misfire due to adaptations that were once an aid

[6] Greene and his colleagues only began to characterize judgments of their experimental materials as 'deontological' or 'utilitarian' in 2004, and as many authors have pointed out, these characterizations are, shall we say, more expedient than they are accurate (see, e.g., Meyers 2015).

https://doi.org/10.1017/apa.2018.25 Published online by Cambridge University Press

to survival but do not align with morality (Greene 2014b; see also Singer 2005). Needless to say, what many perceive to be Greene's presuppositions about just what qualifies as moral 'misfiring' (Sunstein 2005) have not gone uncriticized (section 3.2).

Regardless, for Greene the upshot of his *error theory* is *positive*. Once we recognize the evolutionarily familiar but morally irrelevant factors that can influence moral processing, we can begin to identify what sorts of judgments do not depend on those distractors. And once we do, Greene argues, we usually find that utilitarianism wins the day. From this point, Greene is often accused of jumping from the finding that a utilitarian framework usually *does* win the day (under what he calls 'impersonal' circumstances) to the conclusion that it *should* win the day (under any circumstances). It is not clear to me that Greene actually makes this leap[7] although anyone who does certainly opens herself up to certain strong criticisms (section 3.2).

One frequently overlooked but standout aspect of Greene's work is its clearly stated recognition of the perils of jumping 'from neural "is" to moral "ought"' (Greene 2003: 846). From the start, Greene and his colleagues acknowledge that their results 'raise but do not answer' (2001: 2107) questions about the link between philosophical ethics and the neuroscience of moral judgment. Greene (2004) also highlights at least one major philosophical implication of his findings that does not take on normative aspirations. He argues that for utilitarian inclinations to win psychological battles against deontology, controlled cognition must win out over emotion.

This is a purely descriptive claim. And if it is true, Greene points out, then we may need to rethink radically the metaethics we have developed over the past 350 years. If deontological principles arise not from the triumph of pure reason, as Kant claimed, but from its absence, poverty, or failure, then the (descriptive) nature of deontology's foundations—even if not the (normative) propriety of those foundations—may have been long misconstrued. It is this empirically contentious—but normatively neutral—proposition that Greene (2008) calls 'The Secret Joke of Kant's Soul'.

The joke, according to Greene, is that deontology and rationality, those two pillars of Kant's work, are in reality diametrically opposed. Rationality begets consequentialist judgment, according to Greene, whereas deontology is the product of emotional decision making. I take this to be Greene's most interesting point and think it unfortunate that it has largely been overshadowed by the normative arguments usually ascribed to him. While many researchers (including Gray) would disagree with this descriptive claim, it is ultimately an empirical question, not a normative one, and therefore immune to many of the objections to which other aspects of Greene's work have been subjected.

[7] To wit: in response to this statement, one well-informed blind referee has suggested to me that Greene's (2014) book very clearly makes this leap. The other blind referee, who also appears to be well-informed on this subject, has asserted that Greene's view is indeed not this simple and has urged me not to wade any further into this debate. I thank the reviewers for supporting my belief that there remains disagreement on this issue by disagreeing about it themselves.

https://doi.org/10.1017/apa.2018.25 Published online by Cambridge University Press

Like Greene, Kumar and Campbell (2012) also seem to be error theorists. However, after rejecting Greene's project on interpretive grounds (section 3.2), they argue that negative philosophical projects may benefit from neuropsychological insights into moral processing. Thus, they are *negative error theorists*. They claim that when sets of ordinary moral judgments violate what they call consistency reasoning—a notion similar to (though substantively different from) that of wide reflective equilibrium (Rawls 1971)—we can be sure that a flawed ethical judgment has been made. However, they say, we cannot be sure which of the coincident judgments is flawed and therefore cannot deduce a positive conclusion about which judgment ought to be embraced.

3.2. Against Positive and Negative Error Theories

Some authors have criticized error theorists on inferential grounds. For example, Meyers (2015) has pointed out that moral systems based on neither outcomes nor rule-following are entirely ignored by Greene and other dual-process theorists. As noted earlier (section 2.2), the existence of virtue ethics, the ethics of care, and other moral frameworks suggests that any description of moral processing cannot be mapped neatly onto just two moral systems without the a priori exclusion of many other potentially valid moral systems.

In their inferential critiques of Greene (2003, 2008) and Singer (2005), Kumar and Campbell (2012) hone in on another major vulnerability of error theories. As they point out, any assumption about which of two conflicting views results from 'misfiring' (Sunstein 2005: 541) can only be established on a priori, theoretical grounds. Therefore, neuroscience cannot identify what counts as an error in moral processing. However, Kumar and Campbell (2012) themselves also make a problematic assumption, albeit a much subtler one. Their negative error theory relies on the presupposition that whenever a person draws conflicting moral conclusions in equivalent cases, at least one of those conclusions must be mistaken. But this presupposition is problematic for at least two reasons.

First, it is not clear how we should define equivalent cases in a moral context. Like Greene, Kumar and Campbell treat moral prospects as roughly equivalent whenever they involve equivalent economic prospects—for example, number of lives lost and saved (Tversky and Kahneman 1981). But as Sunstein (2005) has pointed out, the link between economic prospects (net lives lost) and normative ethical behavior (obligatory, permissible, and forbidden acts) is far murkier than the link between economic prospects and normative economic behavior.[8]

Second, even if we grant that some pair of moral judgments presents a legitimate ethical conflict, it does not follow that at least one of these judgments must be wrong. Some philosophers, encouraged by evidence from cross-cultural studies and other research, suggest that not all of the potential building blocks of normative moral

[8] Interestingly, the roots of this objection are actually older than NMN itself. Many contrastive vignettes that NMN researchers assume should be judged equivalently, such as trolley problems (Foot 1967), were initially introduced (and contrasted) in order to show that economic prospects and moral prospects are *not* always equivalent (Thomson 1976).

https://doi.org/10.1017/apa.2018.25 Published online by Cambridge University Press

systems fit together (Prinz 2008). If these philosophers are right, then any coherent system of moral values will, by necessity, have to leave out certain incommensurable values that nonetheless would be appropriate in, and arguably necessary for, other equally valid systems of moral values (Wong 2006). Thus, two mutually exclusive moral judgments may both be valid because the appropriate system of moral values to adopt in any given situation may vary. Furthermore, even if two conclusions—moral or otherwise—really do strictly contradict one another, it may not necessarily be wrong to endorse both (Priest 1998).

Kamm (2009) has made a point similar to that made by Meyers (2015), and has suggested a methodological cure to what she believes ails Greene's project. Poor design of moral dilemmas used as experimental probes, she alleges, is ultimately what hamstrings Greene's attempted inferences. She gestures at a number of ways these probes might be improved in future work, but her suggestions are hampered by a reliance on the unrefined and inaccurate notion that there exist emotional and rational 'centers of the brain' (Kamm 2009: 332).

An especially damning methodological objection has been laid out by Berker (2009). He points out that many of Greene's probes—including one asking whether it is morally permissible to hire someone to rape your wife so that you can then console her (Greene et al. 2001)—do not seem to present moral dilemmas at all.[9] Gray further suggests that even if Greene's moral dilemmas were well-constructed, the resolution of moral dilemmas is so atypical a function of moral judgment that studying it cannot teach us much about ordinary moral processing. Gray insists that 'moral dilemmas in which wrongness (deontology) and harm (utilitarianism) conflict are unrepresentative of typical moral cognition' (Gray et al. 2014: 1600). Importantly, Greene's exact probes and others like them have been used many times by many researchers since 2001. Thus, Berker's and Gray's criticisms, if apt, could have a far-reaching implications.

One more methodological concern about error theories—and about NMN theories more generally—is the lack of ecological validity in the experiments by which they are often defended. For instance, many studies in NMN—especially those conducted by error theorists—use 'trolley problems' (Thomson 1976) and other contrived dilemmas like them. Much has been written questioning the legitimacy of treating abstruse cases like these as paradigms of moral decision making (e.g., Meyers 2015). And as just noted, Gray has argued that all moral dilemmas are so unusual that not only their philosophical but also their psychological relevance may be doubted.

4. Principles of Quality NMN

What lessons for the third tradition of neuroethics can we draw from these arguments within—and against—NMN? How, if we are to pursue NMN, can we

[9] Other examples from Greene and colleagues (2001) include moral-impersonal case 19: 'Is it appropriate for you to hire this surgeon to carve out a stranger's eye in order to help restore your vision?' and moral-personal case 13: 'Is it appropriate to throw your baby in the dumpster in order to move on with your life?' Additional probes include simple cases of cheating on taxes, falsifying a resume, killing a 'despicable' boss, and poisoning one's grandmother.

https://doi.org/10.1017/apa.2018.25 Published online by Cambridge University Press

do so in the most responsible, well-founded way? In this section, I synthesize some of the discussions above into three loose principles of quality NMN, which are meant to answer these questions. The first of these principles, schematic neutrality, is meant to help overcome inferential objections. The second principle, psychological minimalism, is meant to address inferential problems that may be resolved by changes in methodology. And the third principle, methodological integration, is geared primarily toward methodological objections, which may be addressed by both changes in, and the expansion of, the methodologies used by NMN researchers. These three suggestions are not intended as inviolable principles that NMN researchers must adhere to without exception. Rather, they are meant only as rough guidelines that NMN researchers might want to embrace in order to circumvent some of the theoretical roadblocks that are visible in the present literature and to illuminate the future directions in which emerging NMN projects might flourish.

4.1. Schematic Neutrality

The most obvious step in producing more robust work in NMN may be to strip the work of a priori normative schemes to whatever extent possible. As easy as this may be to recognize, it will be difficult to achieve. Indeed, the sophistication of many arguments against NMN reveals just how hard it can be even to recognize that one is making normative assumptions in this pioneering subdiscipline. As illustrated in my critique of Kumar and Campbell (2012) above (section 3.2), settling on a normative ethical scheme in which to interpret moral neuropsychological findings actually involves two assumptions, not just one. There is the more obvious assumption regarding which scheme should be adopted, and then there is the more insidious assumption that only one scheme can be right.

Alternatively, some work within NMN might successfully proceed with normative schematic assumptions so long as such work is tagged with the caveat that it will, by necessity, be relatively parochial. There is nothing wrong with deriving normative principles from, for example, Gigerenzer's cognitive science of moral heuristics (section 2.1) as long as it is acknowledged that these normative principles are contingent on certain assumptions about the nature of normative ethics—in Gigerenzer's case, on the presupposition of consequentialism (section 2.2).

It should also be noted that normative schematic presuppositions in NMN are not limited to moral presuppositions. Psychological presuppositions have colored NMN just as much as ethical and metaethical ones have. As discussed in my analysis of Greene's work (section 3.1), normative psychological assumptions, such as the assumption that controlled cognition is better than automatic processing, have also at times infected NMN. Of course, until the end of the nineteenth century, psychology fell entirely within the auspices of philosophy, so in some sense the psychological landscape of NMN has long been shaped—for better or for worse—by traditional moral theorists, including Hume and Kant. Thus, the sort of schematic neutrality I suggest is rather sweeping and might admittedly make it more difficult to draw normative conclusions on the basis of neuroscientific research. Nonetheless, I think it is an important principle to stick to: NMN should

https://doi.org/10.1017/apa.2018.25 Published online by Cambridge University Press

not presume the authoritativeness of any one ethical, metaethical, or psychological theory any more than is necessary for theoretic scaffolding. Furthermore, the limitations incurred by such schematic presumptions should be acknowledged and discussed in detail in those philosophical works that are not schematically neutral.

4.2. Psychological Minimalism

I have just argued that a priori evaluative stances toward psychological states—such as what many perceive to be Greene's a priori evaluation of controlled cognition as better than emotion—are a major (though perhaps inevitable) vulnerability of any NMN project of which they are a part. Such evaluative assumptions open projects to certain obvious interpretive criticisms. Thus, a very basic tenet to which NMN theorists might wish to adhere is that the normative superiority of certain psychological states over others—such as controlled cognition over emotion—should not be assumed whenever such assumptions can be avoided. Here, however, I want to discuss how even descriptive psychological claims within NMN can sometimes be deeply problematic.

Recent work has cast serious doubt on our ideas about which neural substrates are involved in which psychological processes (see, e.g., Klein 2011). In fact, some authors believe that no emotion-cognition dichotomy may be plausible at all (e.g., Dolan 2002), at least when it comes to classifying neural substrates (Holtzman 2018). And even if such a dichotomy may be plausible in some contexts, Gray and Schein (2012) provide reasons to doubt its plausibility within—or at least its relevance to—the moral domain (section 3.2). As such, it may be best, at least until neuroscience advances further, for NMN not even to attempt to distinguish emotional from controlled cognitive activity on the basis of the activation of specific neural substrates.

Gray's work (section 2.1) also reflects a more general problem with characterizing neurophysiological data in psychological terms. Like Young and her colleagues (2007; Young and Saxe 2008), Gray focuses on the perception of other minds as the core of moral judgment. But in doing so, Gray assumes that the involvement of brain regions that have been associated with certain psychological processes in some experimental paradigms demonstrates the involvement of those same psychological processes in different (moral) experimental paradigms. However, this kind of abductive reasoning, which in neuroscience is known as reverse inference, is deeply problematic (Poldrack 2006). There is not a one-to-one or even a one-to-many relationship between neural activity and psychological states and processes; that relationship is more likely many-to-many (Longino 2013). Even the implications of double dissociations[10] between neural activity and

[10] In its narrow, original usage (Teuber 1955), a *double dissociation* is a case in which damage to brain region *A* is associated with the impairment or loss of capacity *X* while capacity *Y* remains intact, whereas dysfunction in brain region *B* is associated with the impairment or loss of capacity *Y* while capacity *X* remains intact. In its broader usage, which emerged during the 'cognitive turn' in neuropsychology (e.g., Marin, Saffran, and Schwartz 1976), the term refers to any case in which capacity *X* remains intact while capacity *Y* is severely impaired in some patients, whereas capacity *Y* remains intact while capacity *X* is severely impaired in others (Shallice 1988: 248).

https://doi.org/10.1017/apa.2018.25 Published online by Cambridge University Press

psychological processes are uncertain (Van Orden et al. 2001), and so it should be clear that single associations between neural activity and psychological processes do not prove the engagement of those processes whenever that activity is observed (Jacoby 1991).

Thus, not only moral-philosophical but also psychological inferences from neuroscientific findings should be made cautiously and scrutinized carefully to ensure that such inferences are fully warranted. This is not to say that we should (or even meaningfully could) do away with all such inferences and assumptions. Still, we might adopt a somewhat conservative version of the ideas just developed, as follows: Psychological characterizations of neuromoral processing should be kept to a minimum and flagged as such whenever possible.

4.3. Methodological Integration

Whereas schematic neutrality and psychological minimalism are oriented (to different degrees) toward overcoming interpretive objections to NMN, this last recommendation is directed primarily against methodological objections. To begin, we should acknowledge that NMN could benefit greatly from a shift toward more ecologically valid work. Such work is especially challenging to design and interpret because of the number of uncontrolled variables and potential ambiguities it can involve. However, if NMN theorists hope to draw normative philosophical conclusions from the neuroscience of moral judgment, it may be especially incumbent upon them to grapple with these challenges and to work within the limitations that ecological validity may place on the interpretation of their work. Although Kamm (2009) is correct in noting that many experimental prompts lack philosophical rigor in their design, Berker (2009), Gray and colleagues (2012), and others are right to point out that any use of moral dilemmas or of contrived moral vignettes may put researchers 'on the wrong track' (Kahane 2012: 519). While both ethical and practical considerations may necessitate that the majority of new research in NMN falls short of perfect ecological validity, keeping this ideal in mind may nonetheless prove valuable.

While increased philosophical scrutiny of experimental materials may be necessary, vignette-driven lab research, if not supplemented by other kinds of research, remains at risk to miss the forest for the trees. In addition to figuring out how to refine the sorts of experiments NMN researchers have been conducting, a broader reflection on the kinds of research methods NMN should use is in order. The third tradition could benefit especially from studying cases in which between-group differences (and similarities) are incontrovertible, in which observable behavioral outcomes obviate notoriously troublesome (Friborg et al. 2006) Likert-type self-reports,[11] and in which the decisions participants make are

[11] In addition to its usual problems, reliance on self-reports faces a unique problem in the context of moral judgment: the objection from motivational internalism. According to motivational internalism, a genuine moral judgment must be accompanied by a disposition to act in accordance with that judgment. If internalism is correct, then the judgments expressed verbally in self-reports regarding moral problems may not always be genuinely moral judgments.

https://doi.org/10.1017/apa.2018.25 Published online by Cambridge University Press

both realistic and typical of those that people encounter in their daily lives. Additionally, increased study of (im)moral behavior in social situations, rather than among isolated individuals, may be especially valuable, given the social nature of moral decision making.

There is already a great deal of literature in fields cognate to NMN whose potential normative-ethical implications has yet to be explored by neuroethicists. Such work covers topics ranging from the apparent potential of opioid antagonists to treat kleptomania (Grant and Kim 2002) to the automatic down-regulation of affective inhibitory responses among surgeons viewing images of patients that the rest of us would find extremely aversive (Decety et al. 2010). In addition to suggesting new directions for future NMN studies, extant work may on its own provide meaningful insights and ideas for those working within the third tradition. By way of summary: Insights into the nature and implications of neuromoral processing should be sought using all the tools at our disposal, which means significant diversification of our methods and the inclusion of methods and findings from other fields.

5. Conclusion: The Value of NMN

Normative moral neuroscience may best be thought of as a unique subdiscipline, a third tradition of neuroethics. It is its own tradition in at least as strong a sense as the neuroscience of moral judgment and the ethics of neuroscience are their own distinct traditions. NMN has come into its own in the past fifteen years, but it is still quite young. Its present shortcomings doubtless reflect the field's youth and so should not be overblown. Nonetheless, NMN by its nature faces certain hard problems not encountered in other, less obviously interconnected areas of neuroethics. The need to grapple with these problems opens the field to a cadre of potential attacks and objections.

In particular, NMN presents researchers with a recurring trade-off between the philosophical impact of their theories and the plausibility of their arguments in favor of those theories. By studying and learning from the field's past successes and failures, we may be able to optimize this trade-off. Moreover, by proliferating a variety of programs within the third tradition, each of which recognizes its own limitations, NMN researchers may be able to make the trade-offs of any given research program less salient in the big picture. By recognizing and taking into account the sacrifices in either impact or plausibility that NMN projects make in service of the other, and by pursuing many of these projects at once, NMN researchers may be able to enrich NMN significantly.

Many—but not all—who object to NMN do so by reference to the is-ought problem (section 1) or the related naturalistic fallacy (Moore 1903). These objectors argue that as a descriptive discipline, neuroscience can have no impact on normative ethics. But normative principles do not operate alone: a great deal of ethics and metaethics is founded upon nonnormative philosophical theory. Among other things, normative principles depend on descriptive theories regarding the nature of rationality, the nature of agency, and the nature of the human mind. Even if neuroscientific research on moral judgment does not directly

https://doi.org/10.1017/apa.2018.25 Published online by Cambridge University Press

bear on normative questions, neuroscientific research may nonetheless have immediate implications for philosophical understanding of reasoning, free will, and personal identity. Inasmuch as these fields bear directly on normative ethics, so too does neuroscience bear on normative ethics, at least indirectly. Thus, neither a priori normative evaluation of participant competence or error in moral judgment nor the presupposition of controversial ethical or metaethical theories is required for the neuroscience of moral judgment to have an impact on philosophical understanding of normative ethics.

I have advocated for an approach to NMN that is both broader and more cautious than current work tends to be. This may sound like a contradiction in terms, but breadth and caution are allies, not enemies. A proliferation of approaches may increase the number of initial missteps made in NMN—but as in any experimental field, missteps are natural and important parts of progress. Thus, NMN research should continue on the track that it is on, but it should also explore more literature from other fields and should work to develop projects in new directions. Normative moral neuroscience is an interdisciplinary field that goes out on a limb and sets its goals high. Even if it falls short of some of those goals, its explorations along the way will surely enrich psychology, neuroscience, and philosophy.

GEOFFREY S. HOLTZMAN
FRANKLIN & MARSHALL COLLEGE
geoffreyholtzman@gmail.com

References

Bechara, Antoine, and Antonio R. Damasio. (2005) 'The Somatic Marker Hypothesis: A Neural Theory of Economic Decision.' *Games and Economic Behavior*, 52, 336–72. doi:10.1016/j.geb.2004.06.010.

Bentham, Jeremy. ([1823] 2005) *An Introduction to the Principles of Morals and Legislation.* New York: Oxford University Press.

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

Bruni, Tommaso, Matteo Mameli, and Regina A. Rini. (2014) 'The Science of Morality and Its Normative Implications.' *Neuroethics*, 7, 159–72. doi:10.1007/s12152-013-9191-y.

De Sousa, Ronald. (1990) *The Rationality of Emotion.* Cambridge, MA: MIT Press.

Decety, Jean, Chia Y. Yang, and Yawei Cheng. (2010) 'Physicians Down-Regulate Their Pain Empathy Response: An Event-Related Brain Potential Study.' *NeuroImage*, 50, 1676–82. doi:10.1016/j.neuroimage.2010.01.025.

Dolan, Raymond J. (2002) 'Emotion, Cognition, and Behavior.' *Science*, 298, 1191–94. doi:10.1126/science.1076358.

Doris, John M. (1998) 'Persons, Situations, and Virtue Ethics.' *Nous*, 32, 504–30.

Einstein, Albert. (1949) 'Remarks Concerning the Essays Brought Together in This Cooperative Volume'. In Paul A. Schilpp (ed.), *Albert Einstein: Philosopher-Scientist* (Chicago: Open Court Publishing), 673.

Foot, Philippa. (1967) 'The Doctrine of Double Effect'. *Oxford Review*, 5, 5–15. doi:10.1002/9781444323528.ch41.

Friborg, Oddgeir, Monica Martinussen, and Jan H. Rosenvinge. (2006). 'Likert-Based vs. Semantic Differential-Based Scorings of Positive Psychological Constructs: A Psychometric Comparison of Two Versions of a Scale Measuring Resilience'. *Personality and Individual Differences*, 40, 873–84.

https://doi.org/10.1017/apa.2018.25 Published online by Cambridge University Press

Gigerenzer, Gerd. (2008) 'Why Heuristics Work'. *Perspectives on Psychological Science*, 3, 20–29. doi:10.2307/40212224.

Gigerenzer, Gerd. (2010) 'Moral Satisficing: Rethinking Moral Behavior as Bounded Rationality'. *Topics in Cognitive Science*, 2, 528–54. doi:10.1111/j.1756-8765.2010.01094.x.

Gigerenzer, Gerd, and Reinhard Selten. (2002) *Bounded Rationality: The Adaptive Toolbox*. Cambridge, MA: MIT Press.

Grant, Jon E., and Suck W. Kim. (2002) 'Adolescent Kleptomania Treated with Naltrexone: A Case Report'. *European Child & Adolescent Psychiatry*, 11, 92–95. doi:10.1007/s007870200016.

Gray, Kurt, and Chelsea Schein. (2012) 'Two Minds vs. Two Philosophies: Mind Perception Defines Morality and Dissolves the Debate between Deontology and Utilitarianism'. *Review of Philosophy and Psychology*, 3, 405–23. doi:10.1007/s13164-012-0112-5.

Gray, Kurt, Chelsea Schein, and Adrian F. Ward. (2014) 'The Myth of Harmless Wrongs in Moral Cognition: Automatic Dyadic Completion from Sin to Suffering'. *Journal of Experimental Psychology: General*, 143, 1600–15. doi:10.1037/a0036149.

Gray, Kurt, Liane Young, and Adam Waytz. (2012) 'Mind Perception Is the Essence of Morality'. *Psychological Inquiry*, 23, 101–24. doi:10.1080/1047840X.2012.651387.

Greene, Joshua. (2003) 'From Neural "is" to Moral "Ought": What Are the Moral Implications of Neuroscientific Moral Psychology?' *Nature Reviews Neuroscience*, 4, 846–49. doi:10.1038/nrn1224.

Greene, Joshua D. (2007) 'Why Are VMPFC Patients More Utilitarian? A Dual-Process Theory of Moral Judgment Explains'. *Trends in Cognitive Sciences*, 11, 322–23. doi:10.1016/j.tics.2007.06.004.

Greene, Joshua D. (2008) 'The Secret Joke of Kant's Soul'. In W. Sinnott-Armstrong (ed.), *Moral Psychology: The Neuroscience of Morality*, vol. 3 (Cambridge, MA: MIT Press), 35–80.

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

Greene, J. D. (2014b) *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. New York: Penguin Books.

Greene, Joshua D., R. Brian Sommerville, Leigh E. Nystrom, John M. Darley, and Jonathan D. Cohen. (2001) 'An FMRI Investigation of Emotional Engagement in Moral Judgment'. *Science*, 293, 2105–08. doi:10.1126/science.1062872.

Greene, Joshua D., and Jonathan Haidt. (2002) 'How (and Where) Does Moral Judgement Work?' *Trends in Cognitive Sciences*, 6, 517–23. doi:10.1016/s1364-6613(02)02011-9.

Greene, Joshua D., Leigh E. Nystrom, Andrew D. Engell, John M. Darley, and Jonathan D. Cohen. (2004) 'The Neural Bases of Cognitive Conflict and Control in Moral Judgment.' *Neuron*, 44, 389–400. doi:10.1016/j.neuron.2004.09.027.

Holtzman, Geoffrey S. (2017). 'Neuromoral Diversity: Individual, Gender, and Cultural Differences in the Ethical Brain'. *Frontiers in Human Neuroscience*, 11, 501.

Holtzman, Geoffrey S. (2018). 'A Neuropsychological Challenge to the Sentimentalism/Rationalism Distinction'. *Synthese*, 195, 1873–89.

Holtzman, Geoffrey S. (Forthcoming). 'Famine, Affluence, and Intuitions: Evolutionary Debunking Proves Too Much'. *Disputatio*.

Jacoby, Larry L. (1991). 'A Process Dissociation Framework: Separating Automatic from Intentional Uses of Memory'. *Journal of Memory and Language*, 30, 513–41. doi:10.1016/0749-596X(91)90025-F.

Kahane, Guy. (2011) 'Evolutionary Debunking Arguments'. *Nous*, 45, 103–25.

Kahane, Guy. (2012) 'On the Wrong Track: Process and Content in Moral Psychology'. *Mind and Language*, 27, 519–45. doi:10.1111/mila.12001.

Kamm, Frances M. (2009) 'Neuroscience and Moral Reasoning: A Note on Recent Research'. *Philosophy & Public Affairs*, 37, 330–45.

Kant, Immanuel. ([1785] 1956) *Groundwork of the Metaphysic of Morals*. Translated by H. J. Paton. New York: Harper Torchbooks.

Kauppinen, Antti. (2014). 'Ethics and Empirical Psychology: Critical Remarks to Empirically Informed Ethics'. In Markus Christen, Carel van Schaik, Johannes Fischer, Markus

Huppenbauer, and Carmen Tanner (eds.), *Empirically Informed Ethics: Morality between Facts and Norms* (Dordrecht: Springer), 279–305.

Klein, Colin. (2011) 'The Dual Track Theory of Moral Decision Making: A Critique of the Neuroimaging Evidence'. *Neuroethics*, 4, 143–62. doi:10.1007/s12152-010-9077-1.

Knobe, Joshua, Wesley Buckwalter, Shaun Nichols, Philip Robbins, Hagop Sarkissian, and Tamler Sommers. (2012) 'Experimental Philosophy'. *Annual Review of Psychology*, 63, 81–99.

Kuhn, Thomas S. ([1962] 2012) *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kumar, Victor, and Richmond Campbell. (2012) 'On the Normative Significance of Experimental Moral Psychology'. *Philosophical Psychology*, 25, 311–30. doi:10.1080/09515089.2012.660140.

Longino, Helen E. (2013) *Studying Human Behavior: How Scientists Investigate Aggression and Sexuality*. Chicago: University of Chicago Press.

Marin, Oscar S. M., Eleanor M. Saffran, and Myrna F. Schwartz. (1976) 'Dissociations of Language in Aphasia: Implications for Normal Function'. *Annals of the New York Academy of Sciences*, 280, 868–84.

Meyers, Chris D. (2015) 'Brains, Trolleys, and Intuitions: Defending Deontology from the Greene/Singer Argument'. *Philosophical Psychology*, 28, 466–86. doi:10.1080/09515089.2013.849381.

Moore, George E. (1903) *Principia Ethica*. Cambridge: Cambridge University Press.

Nagel, Thomas. (1976) 'Moral Luck'. *Proceedings of the Aristotelian Society*, 50, 115–51.

Pham, Michel T. (2007) 'Emotion and Rationality: A Critical Review and Interpretation of Empirical Evidence'. *Review of General Psychology*, 11, 155–78. doi:10.1037/1089-2680.11.2.155.

Poldrack, Russell A. (2006) 'Can Cognitive Processes Be Inferred from Neuroimaging Data?' *Trends in Cognitive Sciences*, 10, 59–63. doi: 10.1016/j.tics.2005.12.004.

Posner, Richard A. (1979) 'Utilitarianism, Economics, and Legal Theory'. *The Journal of Legal Studies*, 8, 103–40.

Priest, Graham. (1998) 'What Is so Bad about Contradictions?' *The Journal of Philosophy*, 95, 410–26.

Prinz, Jesse J. (2004) *Gut Reactions: A Perceptual Theory of Emotion*. New York: Oxford University Press.

Prinz, Jesse J. (2008) *The Emotional Construction of Morals*. New York: Oxford University Press.

Quine, Willard V. O. (1951) 'Two Dogmas of Empiricism'. *The Philosophical Review*, 60, 20–43.

Rawls, John. (1971) *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Roskies, Adina. (2002) 'Neuroethics for the New Millennium'. *Neuron*, 35, 21–23. doi:10.1016/S0896-6273(02)00763-8.

Schnall, Simone, Jonathan Haidt, and Alexander H. Jordan. (2008) 'Disgust as Embodied Moral Judgment'. *Personality and Social Psychology Bulletin*, 34, 1096–1109. doi:10.1177/0146167208317771.

Shallice, Tim. (1988) *From Neuropsychology to Mental Structure*. Cambridge: Cambridge University Press.

Simon, Herbert A. (1956) 'Rational Choice and the Structure of the Environment'. *Psychological Review*, 63, 129–38.

Singer, Peter. (2005) 'Ethics and Intuitions'. *Journal of Ethics*, 9, 331–52.

Street, Sharon. (2006) 'A Darwinian Dilemma for Realist Theories of Value'. *Philosophical Studies*, 127, 109–66. doi:10.1007/s11098-005-1726-6.

Sunstein, Cass R. (2005) 'Moral Heuristics'. *Behavioral and Brain Sciences*, 28, 531–42. doi:10.1017/S0140525X05000099.

Sunstein, Cass R. (2013) 'Is Deontology a Heuristic? On Psychology, Neuroscience, Ethics, and Law'. Available at SSRN: https://ssrn.com/abstract=2304760.

Teuber, Hans-Lukas. (1955) 'Physiological Psychology'. *Annual Review of Psychology*, 6, 267–96.

Thomson, Judith J. (1976) 'Killing, Letting Die, and the Trolley Problem'. *The Monist*, 59, 204–17.

Tversky, Amos, and Daniel Kahneman. (1981) 'The Framing of Decisions and the Psychology of Choice'. *Science*, 211, 453–58.

Van Orden, Guy C., Bruce F. Pennington, and Gregory O. Stone. (2001) 'What Do Double Dissociations Prove?' *Cognitive Science*, 25, 111–72. doi:10.1016/S0364-0213(00)00036-7.

Vavova, Katia. (2014) 'Debunking Evolutionary Debunking'. In Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics*. (Oxford: Oxford University Press), 9: 76–101.

Wong, David B. (2006) *Natural Moralities: A Defense of Pluralistic Relativism*. New York: Oxford University Press.

Young, Liane, Fiery Cushman, Marc Hauser, and Rebecca Saxe. (2007) 'The Neural Basis of the Interaction between Theory of Mind and Moral Judgment'. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 8235–40. doi:10.1073/pnas.0701408104.

Young, Liane, and Rebecca Saxe. (2008) 'The Neural Basis of Belief Encoding and Integration in Moral Judgment'. *NeuroImage*, 40, 1912–20. doi:10.1016/j.neuroimage.2008.01.057.

https://doi.org/10.1017/apa.2018.25 Published online by Cambridge University Press