

A fallacious jar? The peculiar relation between descriptive premises and normative conclusions in neuroethics

Nils-Frederic Wagner¹ · Georg Northoff¹

© Springer Science+Business Media Dordrecht 2015

Abstract Ethical questions have traditionally been approached through conceptual analysis. Inspired by the rapid advance of modern brain imaging techniques, however, some ethical questions appear in a new light. For example, hotly debated trolley dilemmas have recently been studied by psychologists and neuroscientists alike, arguing that their findings can support or debunk moral intuitions that underlie those dilemmas. Resulting from the wedding of philosophy and neuroscience, neuroethics has emerged as a novel interdisciplinary field that aims at drawing conclusive relationships between neuroscientific observations and normative ethics. A major goal of neuroethics is to derive normative ethical conclusions from the investigation of neural and psychological mechanisms underlying ethical theories, as well as moral judgments and intuitions. The focus of this article is to shed light on the structure and functioning of neuroethical arguments of this sort, and to reveal particular methodological challenges that lie concealed therein. We discuss the methodological problem of how one can—or, as the case may be, cannot—validly infer normative conclusions from neuroscientific observations. Moreover, we raise the issue of how preexisting normative ethical convictions threaten to invalidate the interpretation of neuroscientific data, and thus arrive at question-begging conclusions. Nonetheless, this is not to deny that current neuroethics rightly presumes that moral considerations about actual human lives demand empirically substantiated answers. Therefore, in conclusion, we offer some preliminary reflections on how the discussed methodological challenges can be met.

✉ Nils-Frederic Wagner
nils-frederic.wagner@web.de

¹ Mind, Brain Imaging and Neuroethics Research Unit, University of Ottawa Institute of Mental Health Research, Ottawa, Canada

Keywords Neuroethics · Ethical theory · Neuroscience of morality · Moral psychology · Naturalistic fallacy · Normative fallacy · Normative indeterminacy · Norm-fact linkage

Introduction

There is a contemporary zeitgeist reflecting the tendency to bring together philosophy and the empirical sciences, in particular neuroscience. Inspired by the rapid advance of modern brain imaging techniques, such as functional magnetic resonance imaging (fMRI), philosophers who engage with neuroscience have begun to work on what was long believed (and is still believed by some) to be the last secure stronghold of a purely conceptual discipline: ethics.¹ So far as this is concerned, scholars from different disciplines look toward neuroscience for guidance in moral questions. This often naturalistically minded endeavor has led to the rise of neuroethics² as a fairly novel branch of interdisciplinary research.³ Yet, there is a difficulty in giving a straightforward all-encompassing definition of neuroethics, since the work done under this label is not monolithic in its methodology. What we are concerned with in what follows is a naturalistic form of the neuroscience of ethics which is the predominant, though not only, school within neuroethics. In further characterizing the burgeoning field of neuroethics, we are following the lead of Adina Roskies, who asked,

Will the biologizing of the moral undermine its status as moral? ... It is clear that as such questions are approached scientifically, the answers we get will shape our ethical views and, thus, will affect how we approach the ethics of neuroscience. As we learn more about the neuroscientific basis of ethical reasoning, as well as what underlies self-representation and self-awareness, we may revise our ethical concepts. [3]

Along those lines, Michael Gazzaniga affirms: ‘Cognitive neuroscience has valuable information to contribute to the discussion of certain topics that have traditionally been taken up by bioethicists, namely, those issues in which brain science has relevant knowledge that should impact the ethical questions being debated’ [4].

Such approaches to neuroethics can be labeled as ‘robustly naturalistic’ in their methodology. Attending to neuroscientific evidence, then, is believed not only to be expository as to how human beings *de facto* act but also to reveal tenets of normative ethical theories. Doing neuroethics in this way can be roughly delineated

¹ Recently, this conviction has been opposed by experimental approaches to philosophy in general and experimental ethics in particular. Major proponents endorsing experimental philosophy are, amongst others, Joshua Knobe and Shaun Nichols (for a theoretical and methodological justification of their approach, see [1]).

² For a discussion of the theoretical and methodological hallmarks of neuroethics, see [2].

³ The same holds for the ever-growing research conducted under the umbrella term of moral psychology. Unlike in neuroethics, however, moral psychologists are less concerned with making normative claims, but rather aim at describing human functioning in moral contexts. This is not to ignore that the findings of moral psychology are frequently invoked into ethical controversies; sometimes they are consulted to serve as a ‘tie-breaker’ between conflicting theories.

as the attempt to draw conclusive relationships between neuroscientific observations and normative ethical theories—consequently aiming to suggest concrete prescriptions in applied ethics. The general aim is to merge scientific descriptions and normative evaluations of human (and occasionally non-human) life. Neuroethics so defined investigates both the significance of neuroscientific findings for the understanding of morality, and the relevance of ethics for determining the normative import of evidence from neuroscience.⁴ The former is concerned with neural and psychological mechanisms underlying ethical concepts and judgments; whereas the latter is concerned with the implications of those findings for moral practice. Accordingly, Adina Roskies introduced the clarifying distinction between ‘neuroscience of ethics’ and ‘ethics of neuroscience’ [3].

On a related note, neuroessentialist views asserting that, ‘for all intents and purposes, we are our brains’ [6] have become increasingly popular. Such views have found their way into ethical discourses, creating new methodological difficulties emerging from a neuroscientific naturalism that materializes in the structure of neuroethical arguments that, more often than not, rely effusively on empirical evidence. While there is much debate over the ‘is-ought divide’ in empirically informed ethics in general [7], the underlying structure of neuroethical arguments in particular remains elusive.

Current debates in neuroethics mostly focus on the application of neuroscientific findings to concrete ethical questions, such as free will [8], moral responsibility and psychopathy [9], and impacts of brain interventions on personal identity [10]. Little, however, has been said about the general structure of how these arguments operate. In addressing this important and underappreciated methodological concern, the focus of this article is to reveal the implicit metaethical premise that underlies a great deal of neuroethical arguments and to showcase specific methodological predicaments that result thereof. In that way, we seek a better grip on assessing merits and demerits of neuroethical work and point to possible limitations and fruitful applications.

In what follows, we will focus on two particular methodological challenges in the neuroscience of ethics. First, we will tackle the methodological problem of how one can—or, as the case may be, cannot—validly infer normative conclusions from neuroscientific observations. Second, we will shed some light on how preexisting normative ethical convictions threaten to invalidate the interpretation of neuroscientific data, and, by so doing, arrive at question-begging conclusions. Thereby, the guiding methodological questions will be: Can neuroscience contribute to definitions of moral reasoning and moral behavior? Can empirical findings that explain how we actually behave or reason, support claims about how we ought to behave and reason? Are neuroethical investigations unconsciously biased by preexisting normative convictions? If so, is it feasible to invoke those theoretical assumptions to justify neuroethical conclusions?

In targeting these issues, our goal is not to give an exhaustive overview of work done in neuroethics but, rather, to examine paradigmatic examples of important

⁴ In current medical ethics, decision making and informed consent are hotly debated topics. For a recent investigation of these issues from a neuroethical perspective, see [5].

neuroethical arguments as present in the work of Joshua Greene and Patricia Churchland.⁵ Here, a note of caution is appropriate: the aim of our article is not to assess Greene's or, for that matter, anyone else's *arguments* for their normative ethical convictions—this has been done thoroughly elsewhere.⁶ Our aim, rather, is to reveal the *underlying structure* of how neuroethical arguments work when they appeal to empirical evidence from neuroscience; hereby, Greene's and Churchland's views are dealt with as paradigmatic examples, not as specific targets. The central point we want to address is the issue of how empirical evidence is invoked to play a decisive role in reaching neuroethical conclusions claiming to be morally significant. The role that empirical evidence plays is believed to be decisive insofar as it is appealed to in order to tip the balance in favor of one position or another. In order to reveal the methodological predicaments in such an approach, we proceed in three main steps:

- (1) Examine how the naturalistic fallacy applies to robust naturalistic neuroethical arguments: do these arguments directly infer an 'ought' from an 'is'—or is their force rather based on, as it were, smuggled-in normative assumptions? We will call these assumptions 'semi-normative' claims.
- (2) Introduce the normative fallacy: do these arguments further infer an 'is' from an 'ought'—thus presupposing the conclusion in the premises? We will call these 'result-closed' arguments.
- (3) Sketch some possible solutions to the diagnosed methodological predicaments: namely, avoid biased interpretations of empirical data by starting neuroethical investigations with a normative indeterminacy leading to what we will describe as 'result-open' arguments. The aim here is to achieve sound neuroethical arguments that cohere best with empirical evidence due to what we indicate as a 'norm-fact linkage'.

The structure of neuroethical arguments: naturalism and methodological predicaments

To be clear from the outset, we believe that current neuroethics rightly presumes that moral considerations about actual human lives demand empirically substantiated answers. Yet, the naturalistic conjecture that inquiry into the natural world can increase moral knowledge in just the same way as it increases scientific knowledge seems rather contentious. Along these lines, a good deal of current neuroethics proposes (sometimes implicitly, i.e., neither defended nor even stated as such) a naturalistic form of moral realism, according to which there are objective moral truths, or moral facts and moral properties. And these moral facts, as

⁵ There are, of course, a great many other scholars doing important work in neuroethics and moral psychology that we are not discussing here; see, e.g., [11–15].

⁶ For an ingenious analysis of Greene's normative ethical arguments, see [16].

naturalists believe, are, at the same time, natural facts and properties.⁷ This is frequently called ‘moral naturalism’, which is, at least in its most robust form, committed to a rejection of the fact-norm distinction.⁸ Neuroethics thus seems to be grounded in some form or other of ‘metaphysical naturalism’, which is the conviction that all facts and properties are natural, even if we are unable to presently recognize them as such. Accordingly, non-naturalistic forms of normativity are seen, in this view, as unfounded or founded upon illusory beliefs. Neuroethical arguments of this sort typically appear in a form like this:

- (1) (Implicit) metaethical premise: there are (depending on the metaethical conviction) absolute (context independent) or relative (context dependent) normative truths which are, in either case, natural facts or properties that can be discovered scientifically.⁹
- (2) Empirical claim: observation of some fact *x* about brain activity during a morally significant judgment or conduct.

Therefore,

- (3) Normative conclusion: according to the empirical evidence about fact *x*, the moral judgment or conduct in question is inferred to be either right or wrong (good or bad).

It is quite evident that a particular methodological challenge in these kinds of neuroethical endeavors lies in the relation between facts and norms. More precisely, there are crucial steps to be taken to get from empirical claims, derived from observations of brain activity underlying moral intuitions and capacities, to normative conclusions of how to get things right. This holds both in principle, i.e., according to normative ethical theories, and in concrete settings of applied ethics. In this regard, Guy Kahane emphasizes the difference between an investigation into underlying psychological and neural mechanisms that may account for moral

⁷ Many ‘traditional’ contemporary moral philosophers endorse one form or another of non-naturalistic moral realism (most prominently Derek Parfit, Tim Scanlon, and Thomas Nagel). On the other hand, proponents of naturalistic neuroethics are mostly either neuroscientists, like Sam Harris, who endorses a naturalistic form of moral realism, or neurophilosophers—some of whom hold opposing metaethical views, such as Jesse Prinz’s non-cognitivism and Patricia Churchland’s naturalistic moral realism. Joshua Greene is metaethically agnostic and lately considers himself to be a moral skeptic. Although he has frequently asserted the supremacy of utilitarianism as a normative view based on neuroimaging studies that he interprets according to a coherentist moral epistemology.

⁸ For recent reflection on the plausibility of moral naturalism, see [17].

⁹ Henceforth, we assume this metaethical premise to be present in the exemplary neuroethical arguments we consider. We are aware of the fact that not all neuroethical arguments are ipso facto committed to moral naturalism. However, the arguments we are using as paradigmatic examples throughout this article attempt to draw more or less direct normative conclusions from descriptive premises that describe brain activity, and such arguments are likely to presuppose moral naturalism. Our claim is just that the argumentative force of endeavors like this depends on moral naturalism, not that these are the only sorts of arguments available or that all proponents of neuroethics explicitly endorse or implicitly commit themselves to a robust form of moral naturalism.

competence and normative ethical theories that seek to provide principled correct answers to moral questions. Kahane writes:

[T]he aim of ethical theory is surely not to investigate moral competence, or people's psychology or capacities, but to answer substantive normative questions. Moral intuitions may be evidence for or against possible answers to these questions, but they are not data that moral theories seek to causally explain. The aim of ethical theory is to get things right, not to explain why we have a certain set of beliefs (let alone of intuitions). [18]

Kahane's reflections call into question the feasibility of inferences from factual descriptions (based on the observation of brain activity underlying moral beliefs and intuitions) to normative conclusions that purport to either vindicate or impugn certain forms of judgment or conduct.

We now turn to shed some light on how neuroethical arguments that deal with this issue are threatened by two different (albeit intertwined) forms of fallacies: the 'naturalistic fallacy' and the 'normative fallacy'.

The threat of naturalistic and normative fallacies

Naturalistic proponents of neuroethics frequently endorse a reductionist form of inferring normative conclusions from factual descriptions. In so doing, the normative level is mostly disregarded; that is, the normative level is reduced to the factual level. Ethical norms are thus unilaterally replaced by neural facts. Hence, the aim is to, as it were, 'neuronalize' the ethical concept, which means to reduce it to neuronal facts.¹⁰ This, however, involves the danger of an uncritical acceptance of empirical presuppositions and resulting terminological definitions. Traditionally, this form of inference has even been presented as fallacious.

Ever since David Hume [19] and George Edward Moore [20], philosophers take issue with drawing normative conclusions from factual descriptions. These famous non-cognitivists argue, in slightly different ways, against naturalistic forms of moral realism and ethical rationalism by stating that ethical conclusions cannot be drawn validly from premises which are in themselves non-ethical. It is not valid, as they say, to infer an 'ought' from an 'is'; to infer from fact to value, from descriptive to normative ethical propositions, from visibility to desirability. In a nutshell, evaluative conclusions cannot be drawn from non-evaluative premises. This has been labelled as the naturalistic fallacy. Moore puts it as follows: 'I have thus appropriated the name Naturalism to a particular method approaching Ethics. ... This method consists in substituting for "good" some property of natural object or of a collection of natural objects' [20, pp. 91f.]. In another passage of *Principia Ethica*, Moore writes, 'But if [one] confuses "good," which is not ... a natural object, with any natural object whatever, then there is a reason for calling that a naturalistic fallacy' [20, p. 65]. Norms and values are regarded as distinctively

¹⁰ For a discussion of this issue, see [2]. Another form of neuroethical investigation (with which we are not concerned here) is merely to reveal the relevance of ethical concepts for neuroscientific research.

different properties than facts and descriptions; so each set of properties belongs to a different realm. The implication of this view is the invalidity of unilateral inferences from one realm to the other. For example, it is a non sequitur to infer from the premise that if someone observes his fellow students skipping class today that she is morally right to skip class herself as well. What is at stake as to norms and facts, Hume and Moore say, is a difference in kind, not merely a difference in degree.

There is much philosophical controversy about the naturalistic fallacy and accordingly a substantial body of literature concerning this major metaethical issue.¹¹ One can, and justifiably, be undecided about whether or not there can in principle be a way to derive an 'ought' from an 'is'. However that may be, if a robust naturalistic neuroethical argument that aims at prescribing how to get things right is pursued, a methodological step to bridge the gap between the different realms is needed. This holds true even if one assumes that there is no such thing as a naturalistic fallacy, since one still has to explain how to get from observations to prescriptions, for fallacies are not the only source of gaps. As long as the debate over the divide between 'is' and 'ought' is not dissolved, the claim that there is a peculiar relation (or a possible gap) between facts and norms, stands. Ignoring this would be to confuse metaphysical with epistemic and linguistic differences and, therefore, neglecting the need for distinct methodological approaches when relating 'is' and 'ought' claims, either in a metaphysical or in an epistemological or in a linguistic sense. That is to say, even if one accepts the contentious assertion that there is no deep metaphysical difference between facts and norms, and thus believes that it is not fallacious to infer from 'is' to 'ought', it does not follow that there cannot be crucial methodological differences in the ways that conclusions are inferred from these two sorts of statements. For the epistemic differences between facts and norms—accompanied by the linguistic differences in uttering 'is' and 'ought' sentences (for examples of the linguistic formulation of the is-ought gap, see [33, 34])—cannot be denied. In other words, granted for the sake of the argument that there is no such thing as a naturalistic fallacy, it would still be unclear how facts relate to norms for the simple epistemic reason that 'ought' statements make claims about how the world *should be*, whereas 'is' statements are descriptions of how the world *is*. Here, it is worth emphasizing the different ways in which evidence is gathered in support of these two sorts of statements. Evidence for 'is' claims is gathered by observation, whereas evidence for 'ought' claims is gathered by arguments from principles that, more often than not, appeal to consequences. Thus, even if one denies the metaphysical difference between facts and norms, or is agnostic as to how 'is' and 'ought' claims are metaphysically related, the epistemic and linguistic difference still holds and has to be methodologically accounted for.

There are intelligible arguments on either end of the 'is-ought' debate, and, as we will attempt to illuminate, there are also some viable intermediate positions. Without trying to resolve the naturalistic fallacy (which seems to be a matter of trying to square the circle), what the concerns initiated by Hume and Moore illustrate is the *peculiarity* in the metaphysical relation between descriptive

¹¹ For pertinent recent discussions, see [21–31]. Since Frankena [32], some people have called into question the existence of a naturalistic fallacy.

statements and normative conclusions¹² or, more broadly construed, between facts and norms.¹³ In current neuroethics, however, this relation is often implicitly assumed to be a straightforward naturalistic obviousness. Normative facts are believed to be, both in principle and in particular, more or less directly inferable from neuroscientific facts.

Keeping these methodological considerations in mind, we will propose that it is important to ask the following two questions when confronted with naturalistic neuroethical arguments that attempt to draw normative conclusions from descriptive claims.

- (1) Are the descriptive claims correct? This involves asking whether the experimental designs of neuroscientific studies are actually significant and thus able to capture what they aim to investigate.¹⁴
- (2) Do the normative conclusions really follow? This involves asking whether the interpretation of the empirical data, given that the designs are significant, is sound.

Granting for the moment that there might be a principled way to overcome the naturalistic fallacy, one can nonetheless be skeptical about how descriptive claims could have, even *prima facie*, any normative significance whatsoever. Considering the following basic structure of a naturalistic neuroethical argument shall help to further illustrate this point.

Descriptive claim: the amygdala *is* firing when ϕ ing.

Therefore,

Neuroethical conclusion: we *ought* (or, depending on the interpretation of the descriptive claim, *ought not*) to ϕ .

Left in this basic structure (admitted, of course, that this presentation is quite simplified; yet, for our purposes, there is no need to consider empirical details at this point), arguments of this sort appear hopelessly flawed, for they are based on what we will call a *strong* form of the naturalistic fallacy. By ‘strong’, we here refer to a direct, i.e., unilateral normative inference from a descriptive claim.

To avoid such a strong form of the naturalistic fallacy, it seems as though an *extra premise* is needed that aims at connecting the descriptive claim to the

¹² As we said before, if one is skeptical about that, at least the epistemic and linguistic difference between facts and norms is rather uncontentious and needs to be methodologically accounted for.

¹³ The relation between facts and norms is peculiar because there is no straightforward, i.e., direct, way of drawing conclusive arguments from one realm to the other. That is, neither from facts to norms, nor from norms to facts. We will further explain this point in what follows. The general idea is that in current neuroethics, investigators frequently form conclusions according to a suite of convictions that covertly inherits some content from the investigators’ preexisting normative convictions.

¹⁴ Asking whether the empirical claim is correct is, of course, an empirical question and can thus be tackled through the assessment of the data. More important for present purposes is the second claim, which questions the normative significance of the experimental design. In the debate on free will, for example, an often invoked criticism of the Libet experiments is that they do not properly investigate free will, but rather, as Markus Schlosser puts it, ‘freedom of indifference’ [35].

normative conclusion. It is difficult, however, to find extra premises that can support valid inferences from descriptive claims to normative conclusions; especially without presupposing the result, or, for that matter, begging the question. Consider the following integration of what we will call a *semi-normative* extra premise (for lack of a better term)¹⁵ into an exemplary neuroethical argument:

- (1) Empirical claim: the amygdala is known to be integral for emotional processing.
- (2) Semi-normative claim: emotions are an unreliable source for moral judgments.¹⁶

Therefore,

- (3) Neuroethical conclusion: when the amygdala is firing during moral decision making, the resulting moral judgments—including whatever implications may follow for particular normative ethical theories—should not be trusted.¹⁷

When evaluating this argument, what immediately comes to mind is the question of where exactly the justification of the semi-normative claim (2) comes from. And what, after all, is a semi-normative claim? As we will argue hereafter, in answering this question, there may be concealed a source of yet another form of a fallacious inference.

Empirical claim (1)—that the amygdala is firing during emotional processing—aims to causally explain psychological states and not to evaluate them. That much is clear. Whereas the semi-normative claim (2)—that emotions are an unreliable source of moral judgments—bears in itself a normative assumption. It is, therefore, not a purely empirical statement. This is so because the assumption that emotions/sentiments are an unreliable source of moral judgments lies at the core of certain ethical theories and is as a result bound to their normative implications, as in various forms of consequentialism. Therefore, semi-normative claims aim to justify—and consequently to prescribe—morally correct principles that are in accordance with (or follow from) said normative assumptions.

A paradigmatic example of this kind of neuroethical reasoning can be found in the teaming up of neuroscientist Joshua Greene and philosopher Peter Singer. The team argues that the fact that consequentialist moral theories sometimes prescribe actions that are at odds with widely held moral intuitions/emotions cannot by itself count against these theories. On the contrary, it rather increases the plausibility of consequentialist moral theories, they say, since when following consequentialists

¹⁵ In what follows, we will elaborate more on this and aim at making the point that this can also be thought of as a kind of ‘meta-normative’ conviction. These sorts of claims take a stand on the viability of certain forms of normativity (e.g., emotional arousal in moral judgments), but without really providing any metaethical reasoning for this conviction.

¹⁶ This, again, is merely illustrative and not meant to be a statement that we endorse or believe to be true.

¹⁷ When considering this example it is not our purpose to argue for or against any particular normative ethical theory. Rather, we attempt to show how the persuasiveness of neuroethical arguments with this sort of structure is threatened by the particular methodological challenges we discuss.

frameworks, we arrive at conclusions through, as it were, ‘cold’ (prudential) reasoning, which is a more reliable source with a higher likelihood to arrive at sound moral judgments than emotions are.¹⁸ This claim has for a long time been invoked in moral philosophy by proponents of consequentialism arguing against deontology, among others, initially detached from any empirical considerations. Greene argues along those lines, as we will show in what follows, but supplements his case with empirical evidence from fMRI studies he conducted on trolley dilemmas.

Neuroscience of morality and trolley dilemmas

Trolley dilemmas consist of a series of thought experiments originally developed by Philippa Foot [41] to trigger and analyze moral intuitions. Ever since, these hypothetical cases are extensively discussed in normative ethics and have recently attracted much attention in experimental approaches to ethics. The popularity of these dilemmas is particularly due to the debate ignited by Judith Jarvis Thomson, whose widely cited formulation of the trolley dilemma is as follows:

Suppose you are the driver of a trolley. The trolley rounds a bend, and there come into view ahead five track workmen, who have been repairing the track. The track goes through a bit of a valley at that point, and the sides are steep, so you must stop the trolley if you are to avoid running the five men down. You step on the brakes, but alas they don’t work. Now you suddenly see a spur of track leading off to the right. You can turn the trolley onto it, and thus save the five men on the straight track ahead. Unfortunately, there is one track workman on that spur of track. He can no more get off the track in time than the five can, so you will kill him if you turn the trolley onto him. [42]

A slight variation of this case in which you are a bystander that can detour the trolley by pulling a switch, is often referred to as ‘pulling the switch case’, or, for short, ‘switch’. Another vexed variant is the so called ‘footbridge case’. As before, a trolley is running down a track towards five people. You are on a footbridge under which it will pass, and you can stop the trolley by pushing a heavy weight person onto the track. As it happens, there is a person next to you that appears to be sufficiently weighty to stop the trolley—that is, to the best of your knowledge (however that may have come about). Your only way to stop the trolley is to push him over the bridge onto the track, killing him in order to save five, thereby using him as a means to an end.

According to Greene (see [36] and elsewhere), it is especially the footbridge scenario that showcases how the often observed emotionally based moral judgment

¹⁸ Admittedly, Greene’s picture is a bit more complex than we describe it here for the sake of simplicity. Greene argues for what he calls a ‘dual-process theory of moral thinking’, according to which there are two distinct psychological systems forming moral judgments. The ‘automatic mode system’ generates moral judgments based on emotional responses to morally significant situations and leads to rapid ad hoc reactions. In contrast, the ‘manual mode system’ generates moral judgments based on conscious deliberation, thereby possibly overriding the automatic mode system in favor of what reason tells us to be the morally right conduct. For further details, see [36–39]. For Singer’s treatment of moral intuitions, see [40].

to refrain from pushing the weighty man over the bridge is problematic. The reason is that, as Greene puts it, an ‘up close and personal’ involvement in directly harming a stranger likely leads to an emotionally based moral judgment—in this case, the deontological conclusion that it is intrinsically wrong to push the stranger over the bridge onto the track even if it would save five workmen. This is so because it would be to use the weighty man as a means to an end, rather than an end in himself. In contrast, according to the consequentialist’s calculus, it is morally right to push the stranger in order to save the five workmen, since it leads to a favorable consequence: saving five lives at the cost of sacrificing one person’s life. From a consequentialist point of view, whether or not this involves up close and personal harming, or treating someone merely as a means to an end, is neither here nor there. Greene asserts that based on the fact that there is no inherent moral difference in either pulling a switch or pushing a weighty man in the different versions of the trolley dilemma, one can, and in fact *ought to*, infer that it is unreasonable and unjustified to conceive that pushing the weighty man in the footbridge case is morally wrong while pulling the switch is morally right.¹⁹ After all, Greene believes, the only difference between the footbridge and the switch case lies in a ‘morally insignificant’ fact: the involvement of up close and personal harming of a stranger. Accordingly, he believes that this ‘debunks’ deontological moral judgments.²⁰ For these deontological judgments are based on moral emotions which are contingent reactions shaped by the details of our evolutionary history, and thus *should* be divested of their moral significance, as far as this is possible.

Here, it becomes apparent that the earlier discussed semi-normative claim that emotions are an unreliable source for moral judgments, although somewhat disguised as an empirical statement, is for the most part a presupposed hidden normative conviction. Considering that the empirical part of this claim appears as a rather idiosyncratic attempt to support the normative claim, Greene’s claim ought to be seen as contentious.²¹ The inference between the invoked empirical evidence and the normative conclusion is often very difficult to draw, and should not be subject to an author’s undefended normative convictions. Even if these normative convictions are defended, it is important—though often omitted—to show exactly what role the invoked empirical evidence plays in reaching the conclusion. If the invoked empirical evidence was not the backbone of the argument, then why not refrain from it altogether? In the most recent defense of his view, Greene explicitly asserts that ‘moral psychology matters, not because it can generate interesting normative conclusions all by itself, but because it can play an essential role in generating interesting normative conclusions’ [44]. In footnote 68 of the same paper, Greene further opposes Selim Berker’s claim that ‘nonempirical normative assumptions “do all the work” in the above arguments, rendering the science “normatively

¹⁹ Here, we should add that this judgment is based on the outcome, thereby presupposing a consequentialist framework. In light of this, one wonders whether this begs the question; however this is a question for another day.

²⁰ Elsewhere, Greene elaborates more on why he believes Kantian ethics to be grounded in emotions and why this is not a plausible basis for ethical reflection on his view [43].

²¹ For a similar claim, see Berker’s critique of Greene’s fMRI based discussions of said trolley cases [16].

insignificant”’, arguing instead: ‘Normative assumptions do some work, but empirical evidence does essential work as well’.

How the ‘essential work’ that empirical evidence does in these arguments is contentious can be made particularly clear by looking at opposing interpretations of similar empirical data. This phenomenon is most plausibly explained by differences in preexisting normative commitments that consequently lead to different conclusions, and thus, they undermine the point of empirical evidence being crucial in reaching the conclusion. Gerd Gigerenzer, for example, holds a view that is exactly opposed to Greene’s and Singer’s consequentialist convictions. He asserts that people should not rely on conscious deliberation when making moral judgments but, rather, on moral intuitions, or, as Gigerenzer calls them, ‘fast and frugal heuristics’ [45, 46].

With that said, the attempt to support normative ethical views by invoking empirical evidence is not prevalent in consequentialists’ theories alone, but also in deontologists’ theories. For example, contrary to Greene’s (seemingly) empirically supported endorsement of consequentialism when faced with trolley dilemmas, Patricia Churchland believes there is empirical evidence in similar studies supporting deontology.²² A simplified version of her argument goes as follows:

- (1) Empirical claim: the amygdala is known to be integral for emotional processing.
- (2) Semi-normative claim: emotions are a reliable source for moral judgments.

Therefore,

- (3) Neuroethical conclusion: when the amygdala is firing during moral decision making, we should trust the resulting moral judgments—including the supportive implications that may follow for deontology.

Result-closed arguments

Neuroethical arguments of the kind discussed above might be described as *result-closed* arguments. The purpose of adding a semi-normative claim is to provide a ‘smuggled-in-assumption’ that presupposes the intended conclusion in a question-begging way, thereby trying to bring home the preexisting commitment. Arguably, the addition of a semi-normative claim intensifies the gap between descriptive statements and normative conclusions because it is empirically more sophisticated, but ultimately the gap remains. It remains because there is no bridging connection between the normative assertion and the empirical data. The conclusion is already present in the semi-normative claim and therefore the result of the argument is predetermined, or closed. The empirical data are merely used to justify the preexisting normative commitment by asserting that the normative force of the claim is present in a sound interpretation of the empirical data. This interpretation, however, presupposes the normative commitment and thereby begs the question. In

²² See [47] for details.

contrast to the aforementioned strong naturalistic fallacious neuroethical arguments that directly infer normative conclusions from descriptive premises, the empirically more sophisticated variant that draws on a semi-normative claim involves what can be described as a *weak* naturalistic fallacy. Those arguments still derive an ‘ought’ from what they posit to be an ‘is’ but do so by combining the normative and the empirical premise into one, albeit question-begging, claim.

Normative fallacy

We now turn to arguing that these forms of argument beg the question in yet another way that is more difficult to see. Besides the danger of invalidly inferring from facts to norms, there is an additional danger involved in making inferences in the other direction: the drawing of seemingly factual conclusions from hidden normative premises. In what follows, we first shed light on how such reasoning begs a methodological question, demonstrating that some neuroethical arguments make empirical claims that are covertly shaped by undefended normative positions. In a second step, we raise these methodological considerations in further addressing the result-closed quality of neuroethical arguments as discussed above.

The general methodological issue of drawing factual conclusions from normative premises has first been discussed by Tom Campbell [48] and has ever since been underappreciated.²³ We call this methodological difficulty in neuroethical arguments, in line with Campbell’s general conceptual analysis, a ‘normative fallacy’. Campbell argues that philosophical analyses of concepts and the logic of discourse are, despite denials, either descriptive or normative in nature. He believes this is at least in part due to the fact of an indistinct boundary between philosophy and social science. If philosophy is seen as a merely conceptual endeavor (being solely committed to following the rules of logic and internal coherence), and therefore believed to be independent of empirical facts, then there is seemingly no need to involve empirical evidence in arguing for theoretical conclusions or normative commitments. Thus, some philosophers believe themselves to be guarded from the necessity to consider empirical evidence. In Campbell’s words, ‘this frees [philosophers] from the responsibility of providing empirical evidence to support their conclusions and also wards off the accusation that they are parading subjective preferences as if they were rationally justifiable propositions’ [48]. This belief has led to the undesirable consequence that some philosophical analyses contain empirical generalizations that reflect hidden normative assumptions or convictions of the philosopher rather than carefully interpreted empirical evidence. Campbell states further:

In doing this they may be said to reverse the naturalistic fallacy, and, by arguing from ‘ought’ to ‘is’, commit what I shall call the normative fallacy. This fallacy consists of arguing from propositions which are themselves normative, or could count as evidence only for normative propositions, to conclusions which contain factual assertions. The error of such reasoning is

²³ Despite the fact that a lot of philosophical discussions are affected by some sort of a normative fallacy, as of yet, Campbell’s paper is the only published work elaborating on this issue.

obvious, but it is frequently masked by confusions about the precise nature of philosophical analysis. [48]

The same general point made about the naturalistic fallacy can thus in reverse be applied to the normative fallacy. Nothing can appear in the conclusion of a valid deductive inference which is not already implicit in the conjunction of the premises.

Now, of course, by their very nature, neuroethical arguments rely heavily on empirical evidence; but taken closely into consideration, these supposedly factual generalizations sometimes turn out to contain hidden preexisting normative assertions. According to Campbell, it is a confusion about what is actually going on in many philosophical arguments that leads to the sort of conjunction of normative arguments and factual conclusions that constitutes the normative fallacy. Campbell posits that it is easy enough for philosophers to consider their analyses of concepts to constitute factual discoveries based on a certain understanding of philosophical activity. Needless to say, there is a rival view of what it is to do conceptual analysis that derives from Wittgenstein [49]. Rather than discovering the nature of a given concept, it can be said that philosophers are in the business of recapitulating often elusive and occasionally profound, but nevertheless normative assertions about acceptable meanings of words, or the way concepts feature in ordinary discourse. The normative arguments that sometimes invisibly buttress philosophers' apparently factual conclusions about concepts are normative claims about those meanings or uses of words and concepts of which the philosopher approves. What follows is that in many cases, the techniques of inquiring after what is meant by certain phrases or arguing about what they should mean can be regarded as appealing to normative opinions rather than empirical evidence.

So, Campbell convincingly argues that we cannot make valid claims about facts of the world by means of inferring from claims that are based on normative grounds. Descriptive conclusions cannot be drawn validly from normative premises.

Result-closed neuroethical arguments

How does this general methodological point apply specifically to neuroethical arguments? The above discussed example that we *ought* to follow a certain normative ethical theory, which is seemingly supported by neuroscientific evidence describing brain functions, is a case in point threatened by the normative fallacy. For it does not entail that there really *is* anything normatively significant in the invoked empirical evidence that can support these normative assumptions, unless the normative commitment that leads to the conclusion is already implicit in the premise and therefore presupposed.

In order to illustrate what we mean by that, consider again the neuroethical argument based on fMRI studies on trolley dilemmas that have been consulted to debunk deontological moral judgments. Recall Greene's argument regarding the footbridge case, which roughly says that moral judgments resulting from moral intuitions based on emotional reactions are unreliable because they are at odds with prudential reasoning (which in turn is believed to be a reliable source of moral judgments). However, nothing in the empirical observation of the amygdala firing

when people are presented with the footbridge case provides in itself any evidence for the rightness or wrongness of the resulting moral judgment. What the observation of the amygdala firing merely shows, is that people are emotionally aroused when confronted with these scenarios. The claim leading to the conclusion that deontological judgments are unreliable because of emotional involvement is based on the preexisting normative assertion that moral emotions (or intuitions) should not be trusted because they may be at odds with what reason recommends. However, this normative assertion in itself cannot be directly supported by the observation of brain activity, since those empirical observations can only reveal emotional arousal and do not themselves render evaluations as to the reliability of the resulting moral judgments. The force of the argument, therefore, must come in elsewhere; that is, it must derive from its normative presupposition. In Greene's case, the presupposition seems to be that moral judgments based on emotions are wrong because one *ought* not to trust in moral emotions according to consequentialism. This presupposition is a case in point: it shows that Greene's consequentialism predates and animates his conclusions from the fMRI studies he conducted on trolley dilemmas. (It goes without saying that these convictions have been asserted long before people even dreamt about modern brain imaging techniques.)

Clearly, this is a form of a normative fallacy and amounts to what we earlier called a result-closed argument. How so? In starting from the normative assumption that emotions are an unreliable source of moral judgments, then going on to conclude that one therefore ought not to trust them, the result of the neuroethical argument, which is supposed to be based on empirical evidence, is already pre-established by the hidden normative commitment. As a consequence, the pre-established normative commitment significantly determines the interpretation of the data. From the normative conviction that, say, emotions are an unreliable source of moral judgments, Greene and others infer that there de facto must be something empirically observable (namely, in this example, the firing of the amygdala during moral reasoning) corresponding to the unreliability of emotions. Thereby, an 'is' is inferred from an 'ought'. The fallacious inference runs as follows: one 'ought' not to trust in emotions as a source of moral judgments, thus there must be some empirical observable 'is' that affirms this 'ought'. This, then, reverses the naturalistic fallacy, since it consists in inferring from the 'ought' of not trusting moral judgments when based on emotions to the 'is' of the amygdala firing that is invoked to support the earlier mentioned 'ought'. One ought not to trust moral judgments based on emotions, consequentialists posit, therefore, something in empirically observable brain functioning shows that emotions are unreliable when it comes to moral reasoning.

It is, however, highly contentious whether there is any equivalent of the normative commitment to be discovered in the data itself. This can, for example, be seen in opposing normative interpretations of the same data. If one happens to be a proponent of deontological theories, believing that moral intuitions actually are a reliable source of moral judgments, then the very same data appear in a completely different light and, as a consequence, are interpreted in the opposite way. Some deontologists argue, as we observed earlier, that one really ought to follow one's moral intuitions in the footbridge case precisely because the amygdala is firing,

which indicates emotional involvement and therefore, since deontologists believe in the reliance on moral emotions, not pushing the weighty man is the right thing to do. This normative interpretation of the amygdala firing in the footbridge case, by the same token, would also be a result-closed argument, equally based on a normative fallacy. Therefore, it seems as though conclusions of neuroethical arguments of this sort are crucially based on preexisting normative convictions, and the empirical data in their own right play no significant role in the conclusion; rather, the empirical data is consulted in order to bring home the preexisting normative ethical commitment.

Concluding remarks and proposed solutions: normative indeterminacy and norm-fact linkage

The main aim of this article has been to reveal methodological challenges that threaten to invalidate conclusions of robust naturalistic neuroethical arguments. Both the naturalistic and normative fallacies suggest that there are methodological predicaments in neuroethical arguments that seem like solid brick walls, difficult to overcome. This, as the title of the article suggests, can also be seen as, figuratively speaking, a ‘fallacious jar’ in which one is stuck between methodological predicaments that involve the danger of falsely inferring both from facts to norms and from norms to facts. Nevertheless, we believe that the overall project of neuroethics can be fruitful and is indeed needed, since moral considerations about actual human lives demand empirically substantiated answers. With this in mind, the consultation of modern brain imaging techniques (and other empirical methods) can be extremely valuable. For this reason, we do not want to remain entirely destructive, but also want to provide a few pertinent (admittedly preliminary) ideas for how some of the discussed problems can be tackled and might lead to a means of escaping the fallacious jar.

Pertinent to our reflections on how to tackle some of the diagnosed problems that occur in robust neuroethical research, there is a substantial body of thought coming from philosophy of science that addresses the complex ways in which scientific practices, and the products of science, are interwoven with values. Since the 1950s and 1960s, it has been argued that science is inevitably, at least to some extent, governed by value judgments [50]. For one, the application of scientific methods is value-laden. In the quest for empirical discoveries, methods are restrained according to (often implicitly presupposed) normative ethical convictions. For example, invasive or potentially harmful experiments on healthy human participants are disallowed even at the expense of potentially finding a cure for cancer. Accordingly, it has been argued that besides logic and evidence, science is in need of additional guidance for theory choice [51]. In order to account for this, the term ‘epistemic values’ was introduced to encompass the values that were seen as acceptable in guiding scientific research and theory building [52].

More recently, Heather Douglas [53] has opted to abandon the ideal of value free science, particularly if value free science is meant to include the rejection of epistemic values. We agree that there need not be (and maybe cannot be) a science

that is freed of epistemic values, and we concur that it would be good for science to allow for ‘more open discussion of the factors that enter into scientific judgments and the experimental process’ [53]. Douglas acknowledges the methodological predicament of the naturalistic fallacy, alongside the difference between descriptive and normative statements, but she contends that ‘this does not mean ... that a descriptive statement is free from values in its origins. Value judgments are needed to determine whether a descriptive label is accurate enough and whether the errors that could arise from the description call for more careful accounts or a shift in descriptive language. Evidence and values are different things, but they become inextricably intermixed in our accounts of the world’ [53]. In order to make sense of empirical discoveries, we are dependent on scientific interpretations of the acquired data, and these interpretations inevitably have a normative dimension.

Now, how does this apply to the discussed methodological predicaments in robust naturalistic forms of neuroethical research? It is not only the threat of preexisting normative commitments that makes these sorts of neuroethical arguments contentious, but also the frequent neglect of the contingencies governing the *de facto* norms and social structures of everyday life. Any scientific endeavor inevitably presupposes certain epistemological and metaphysical commitments since agents are shaped by a particular context—perceiving and interpreting the world around themselves in a great many different ways. In trying to determine the normative significance of neuroscientific evidence, the relatively austere individual and idiosyncratic social points of view cannot be altogether disregarded. In other words, the social and political contexts in which neuroethical questions are posed and answers are proposed need to be taken into consideration.

Normative concepts like moral judgment and related theoretical reflections, such as the moral status of persons, cannot be comprehensively understood if the social and political context of these concepts is not accounted for.²⁴ This calls for an explicit discussion of the often implicit assumptions of these context-shaped concepts and the need to situate them in both their social and political contexts. If normative ethical convictions are taken for granted and seen as non-negotiable points of reference, the threat of falling for the normative fallacy arises. If, however, these normative ethical convictions are explicitly discussed and critically engaged, there is warranted hope for minimizing this predicament.

On the other hand, the recognition that normative convictions always stand in relation to social and political contexts, provides also a reason for why a ‘neuralization’ (i.e., a reduction of ethical concepts to descriptive facts) falls short. This is so because descriptive facts are, presumably, context-independent; whereas, normative convictions are context-dependent. Such a ‘neuralization’ may then even have fatal implications, especially within applied questions of neuroethics, as they are inevitably shaped by their current social and political context. On a more positive note, these shortcomings suggest once more the need to

²⁴ For an illuminating theory of how personhood and personal identity are fundamentally based on dynamic interactions among biological, psychological, and social attributes and functions of a person’s life that are mediated through a social and cultural infrastructure, see Marya Schechtman’s ‘person life view’ [54]. A great merit of this view is that it coheres very well with recent evidence from social neuroscience and developmental psychology.

complement robust naturalistic forms of neuroethics, and, for that matter, all sorts of empirically informed ethics, with a thoroughly argued conceptual analysis that does justice to carefully situating ethical concepts within their relevant social and political contexts (as a source of norms).

We have argued that there are systematic methodological predicaments in robust naturalistic neuroethical endeavors. For this reason, the development of a methodological tool that enables the avoidance of fallacious inferences from facts to norms and vice versa is needed. One obvious answer to the above described challenge of the normative fallacy is to start neuroethical investigations without preexisting normative convictions, thereby keeping the ultimate result open (rather than result-closed arguments that derive from certain presupposed normative convictions right from the beginning). This helps to avoid biased interpretations of empirical data that are invoked, as it were, to justify these preexisting convictions ‘postmortem’.

Instead, we propose that starting with what we call a *normative indeterminacy* leads to a result-open argument. A normative indeterminacy, of course, does not imply a normative indifference. On the contrary, such a starting point asks for a thorough assessment of the normative ethical convictions that are at stake. In other words, without having a certain normative ethical theory in mind (either in order to support or to debunk said theory), all sorts of neuroethical investigations should at a first stage remain uncommitted to any theory while considering a variety of possible candidates. In so doing, it is crucial, first, to explicitly reveal possible candidates, and second, to discuss the merits and demerits of these normative ethical convictions within the relevant social and political context. This methodological strategy, however, does not remain on the level of revealing merely the particular normative ethical convictions themselves—say, the consequentialists take on endorsing rational moral judgments—but also the implicit presuppositions of these normative ethical convictions that need to be uncovered and discussed. That is to say, the role particular normative ethical convictions play within their social and political context needs to be taken into consideration.

Western societies may, for example, be more driven by consequentialist convictions, whereas eastern societies may be more driven by deontological or virtue ethical convictions. Acknowledging this, then, might also call for further interdisciplinary collaborations with scholars from the social sciences. Such an approach leads to result-open arguments, because it is based on a context-sensitive normative pluralism. A commitment to discussing presuppositions of norms, as well as appreciating that facts have different normative implications depending on which norms are presupposed, allows for comparing and matching of facts and norms in a bilateral way. In following such a methodological strategy, these kinds of neuroethical arguments are more likely to remain open to what happens to result from bringing together ethical questions and empirical evidence. Not only are such neuroethical investigations likely to lead to unbiased results, but they can be reasonably anticipated to have a higher explanatory value since they would be based on a more fine-grained and context-sensitive methodology that captures better the complexities of moral reasoning.

More generally, in order to overcome the barrier between the normative and descriptive realms—in either direction, i.e., from norms to facts and from facts to

norms—some way to introduce a reciprocity of influence between the two realms is needed. We call this desideratum a *norm-fact linkage*.²⁵ Such a notion presupposes a vantage point from within the interface between the descriptive and the normative realm that allows for a development of a mutual linkage between the two. This is needed in order to arrive at theories that are both normatively and empirically plausible without giving precedence to either dimension of the neuroethical endeavor. In other words, a normatively plausible empirical foundation is needed to show that one is neither falsely inferring an ‘ought’ from an ‘is’, nor an ‘is’ from an ‘ought’. How can such a norm-fact linkage that makes possible a close intertwining of normative concepts and neuroscientific observations get off the ground?

The aforementioned examples of moral judgments in trolley dilemmas point to the fact that descriptive change may entail normative change, and vice versa. Therefore, a methodological strategy needs to be sensitive to the close interdependency between norms and facts, while not neglecting their difference. One way is the just mentioned norm-fact linkage; consisting in going back and forth between normative concepts and neuroscientific findings. The usual starting point of empirical neuroethics is to scrutinize an ethical concept that is linked to neuroscientific observations. As said before, the aim is to either ‘neuronalyze’ the ethical concept, or to nail down its philosophical content in order to showcase its relevance for neuroscientific research. One neglected aspect of first encounters between normative concepts and neuroscientific observations is the matter of what neuroscientific observations imply for the normative concept itself. What, for example, is implied for the concept of moral judgments if it is shown empirically to be driven by emotions rather than by reason? What does the fact of the involvement of emotions or intuitions in moral judgments imply for the norms inherent in said judgments? Does the linkage between norms and facts in moral judgments need to be conceptualized differently if emotions or intuitions, rather than reason, are predominant? This may lead to conceptual modifications in neuroethical theories depending on the neuroscientific findings; thus, the initial ethical concept becomes neuroethical in a way that is sensitive to the context-dependent norm-fact linkage.

An interdependent linkage between neuroscientific findings and normative concepts, however, needs to go a step further. What is needed is a method that details the different dynamics of influence between facts and norms at different points in the neuroethical pursuit of such kind, clarifying when and why the friction goes one way or the other. This sort of pursuit cries out for a much greater focus on methodology than is presently typical in the different approaches to neuroethics. However divergent these neuroethical pursuits may be, in any case, there is a need to be self-consciously agile, capable of forging and refining diverse linkages between norms and facts in which both levels are accountable to the other. By following such a method, one may be able to find normative concepts that cohere best with empirical evidence and, thereby, to further the understanding of how scientific observations and normative evaluations of human life are tied together.

²⁵ For further details, see [2].

Acknowledgments We owe special thanks to Pedro L. Chaves, Gabriele Contessa, Lucas Jurkovic, Daniel T. Kim, Gregory J. Walters, Katherine Wayne, Annemarie Wolff, and three anonymous referees for valuable feedback on earlier drafts of this article. For financial support, we are grateful to the Canadian Institutes of Health Research (CIHR), Michael Smith Chair in Neurosciences and Mental Health (EJLB-CIHR), Michael Smith Foundation, and the Hope of Depression Foundation (HDRF/ISAN) to Georg Northoff, and for a Postdoctoral Fellowship from the University of Ottawa and a research grant from Taipei Medical University-Shuang Ho Hospital, Brain and Consciousness Research Center to Nils-Frederic Wagner.

References

1. Knobe, J., and S. Nichols. 2007. An experimental philosophy manifesto. In *Experimental philosophy*, ed. J. Knobe and S. Nichols, 3–14. Oxford: Oxford University Press.
2. Northoff, G. 2009. What is neuroethics? Empirical and theoretical neuroethics. *Current Opinion in Psychiatry* 22(6): 565–569.
3. Roskies, A. 2002. Neuroethics for the new millenium. *Neuron* 35(1): 21–23.
4. Gazzaniga, M.S. 2005. Facts, fictions and the future of neuroethics. In *Neuroethics: defining the issues in theory, practice, and policy*, ed. J. Illes, 141–148. Oxford: Oxford University Press.
5. Northoff, G. 2006. Neuroscience of decision making and informed consent: An investigation in neuroethics. *Journal of Medical Ethics* 32(2): 70–73.
6. Reiner, P. 2011. The rise of neuroessentialism. In *Oxford handbook of neuroethics*, ed. J.I.B. Sahakian, 161–177. New York: Oxford University Press.
7. Schleiden, S., M.C. Jungert, and R.H. Bauer. 2010. Mission: Impossible? On empirical-normative collaboration in ethical reasoning. *Ethical Theory and Moral Practice* 13(1): 59–73.
8. Libet, B.W. 1985. Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences* 8(4): 529–566.
9. Maibom, H.L. 2014. To treat a psychopath. *Theoretical Medicine and Bioethics* 35(1): 31–42.
10. Lipsman, N., and W. Glannon. 2013. Brain, mind and machine: What are the implications of deep brain stimulation for perceptions of personal identity, agency and free will? *Bioethics* 27(9): 465–470.
11. Haidt, J. 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 108(4): 814–834.
12. Haidt, J. 2007. The new synthesis in moral psychology. *Science* 316(5827): 998–1002.
13. Greene, J., and J. Haidt. 2002. How (and where) does moral judgment work? *Trends in Cognitive Sciences* 6(12): 517–523.
14. Huebner, B., S. Dwyer, and M.D. Hauser. 2009. The role of emotion in moral psychology. *Trends in Cognitive Science* 13(1): 1–6.
15. Hauser, M.D. 2008. When your moral organ is right! *Think* 7(19): 17–21.
16. Berker, S. 2009. The normative insignificance of neuroscience. *Philosophy and Public Affairs* 37(4): 293–329.
17. Kitcher, P. 2014. Is a naturalized ethics possible? *Behaviour* 151(2–3): 245–260.
18. Kahane, G. 2013. The armchair and the trolley: An argument for experimental ethics. *Philosophical Studies* 162(2): 421–445.
19. Hume, D. 1978. *A treatise of human nature*. Oxford: Oxford University Press.
20. Moore, G.E. 1993. *Principia ethica*, ed. T. Baldwin. Cambridge: Cambridge University Press.
21. Tanner, J. 2006. The naturalistic fallacy. *Richmond Journal of Philosophy* 13. http://www.richmond-philosophy.net/rjp/rjp13_tanner.php. Accessed May 11, 2015.
22. Hudson, W.D. (ed.). 1969. *The is-ought question: a collection of papers on the central problem in moral philosophy*. London: Macmillan.
23. Searle, J.R. 1964. How to derive “ought” from “is”. *Philosophical Review* 73(1): 43–58.
24. Wilson, D.S., E. Dietrich, and A.B. Clark. 2003. On the inappropriate use of the naturalistic fallacy in evolutionary psychology. *Biology and Philosophy* 18(5): 669–681.
25. Harman, O. 2012. Is the naturalistic fallacy dead (and if so, ought it be?). *Journal of the History of Biology* 45(3): 557–572.
26. Dodd, J., and S. Stern-Gillet. 1995. The is/ought gap, the fact/value distinction and the naturalistic fallacy. *Dialogue* 34(4): 727–746.

27. Walsh, F.M. 2008. The return of the naturalistic fallacy: A dialogue on human flourishing. *Heythrop Journal* 49(3): 370–387.
28. Baumrin, B.H. 1968. Is there a naturalistic fallacy? *American Philosophical Quarterly* 5(2): 79–89.
29. Landeweerd, L. 2004. Normative-descriptive and the naturalistic fallacy. *Global Bioethics* 17(1): 17–23.
30. Sinnott-Armstrong, W. 2008. *The neuroscience of morality: emotion, brain disorders, and development*. Vol. 3, *Moral philosophy*. Cambridge, MA: MIT Press.
31. Rescher, N. 1990. How wide is the gap between facts and values? *Philosophy and Phenomenological Research* 50: 297–319.
32. Frankena, W.K. 1939. The naturalistic fallacy. *Mind* 48(192): 464–477.
33. Stevenson, C.L. 1944. *Ethics and language*. New Haven: Yale University Press.
34. Hare, R.M. 1991 [1952]. *The Language of Morals*. Oxford: Oxford Clarendon Press.
35. Schlosser, M.E. 2014. The neuroscientific study of free will: A diagnosis of the controversy. *Synthese* 191(2): 245–262.
36. Greene, J.D., R.B. Sommerville, L.E. Nystrom, J.M. Darley, and J.D. Cohen. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293(5537): 2105–2108.
37. Greene, J.D., L.E. Nystrom, A.D. Engell, J.M. Darley, and J.D. Cohen. 2004. The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44(2): 389–400.
38. Greene, J.D., S.A. Morelli, K. Lowenberg, L.E. Nystrom, and J.D. Cohen. 2008. Cognitive load selectively interferes with utilitarian moral judgment. *Cognition* 107(3): 1144–1154.
39. Greene, J.D., F.A. Cushman, L.E. Stewart, K. Lowenberg, L.E. Nystrom, and J.D. Cohen. 2009. Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition* 111(3): 364–371.
40. Singer, P. 2005. Ethics and intuitions. *Journal of Ethics* 9(3–4): 331–352.
41. Foot, P. 1967. The problem of abortion and the doctrine of double effect. *Oxford Review* 5: 5–15.
42. Thomson, J.J. 1985. The trolley problem. *Yale Law Journal* 94: 1395–1415.
43. Greene, J.D. 2008. The secret joke of Kant’s soul. In *The neuroscience of morality: emotion, brain disorders, and development*, vol. 3 of *Moral psychology*, ed. W. Sinnott-Armstrong, 35–79. Cambridge, MA: MIT Press.
44. Greene, J.D. 2014. Beyond point-and-shoot morality: Why cognitive (neuro) science matters for ethics. *Ethics* 124(4): 695–726.
45. Gigerenzer, G. 2008. Moral intuition = fast and frugal heuristics? In *The cognitive science of morality: intuition and diversity*, vol. 2 of *Moral psychology*, ed. W. Sinnott-Armstrong, 1–26. Cambridge, MA: MIT Press.
46. Gigerenzer, G. 2010. Moral satisficing: Rethinking moral behavior as bounded rationality. *Topics in Cognitive Science* 2(3): 528–554.
47. Churchland, P.S. 2012. *Braintrust: What neuroscience tells us about morality*. Princeton: Princeton University Press.
48. Campbell, T.D. 1970. *The normative fallacy*. *Philosophical Quarterly* 20(81): 368–377.
49. Wittgenstein, L. 2009 [1953]. *Philosophical Investigations*. Ed. and trans. P.M.S. Hacker and J. Schulte. Oxford: Wiley.
50. Rudner, R. 1953. The scientist qua scientist makes value judgments. *Philosophy of Science* 20(1): 1–6.
51. Churchman, C.W. 1956. Science and decision making. *Philosophy of Science* 23(3): 247–249.
52. McMullin, E. 1982. Values in science. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1982(4): 3–28.
53. Douglas, H. 2007. Rejecting the ideal of value-free science. In *Value-free science? ideals and illusions*, ed. H. Kincaid, J. Dupr’E and A. Wylie, 120–141. Oxford: Oxford University Press.
54. Schechtman, M. 2014. *Staying alive: Personal identity, practical concerns, and the unity of a life*. Oxford: Oxford University Press.