

Greene's Metamorality (Really) without Moral Truth

Steven R. Kraaijeveld

Master's Thesis
Research Master Philosophy
Utrecht University

Supervisor: Dr. Hanno Sauer



Universiteit Utrecht

Acknowledgements

I would like to express my sincere gratitude to Dr. Joel Anderson and Dr. Hanno Sauer, for allowing me to rework some ideas from an essay that I wrote for one of their courses and for their respective guidance throughout this project.

As always, I have my parents to thank for being able to study and to do what I love.

Contents

Introduction	1
I. Background: Developments in Empirical Philosophy	5
I.1. Experimental Philosophy	5
I.2. Metaethics	9
I.3. Moral Psychology.....	12
I.4. Moral Neuroscience.....	14
II. Greene's Dual-Process Theory	18
III. Moral Tribes and Deep Pragmatism.....	23
IV. Automaticity, Happiness, and Truth: Three Arguments against Deep Pragmatism	27
IV.1. Automatic Mode Objection	28
IV.2. Happiness Objections	30
IV.3. Moral Truth Objections.....	32
IV.4. What Is To Be Done?	37
V. Conclusion	42
Bibliography	43

Mere knowledge of human psychology would in itself infallibly make us despondent if we were not cheered and kept alert by the satisfaction of expressing it.

— Thomas Mann¹

Introduction

An empirical turn has recently been taken in moral philosophy (Sinnott-Armstrong 2008a). The emergence and rapid development of fields like experimental philosophy, metaethics, moral psychology, and moral neuroscience has led to the inclusion of empirical findings in discussions of normative ethics that might otherwise have remained restricted to the realm of pure theory, to the so-called and often-maligned² philosophical armchair. The trend has not merely been one of inclusion; researchers have also employed experimental findings in order to challenge long-standing philosophical claims and theories.³ Whether merely to inform existing debates, or to actively undermine specific assumptions and theories, empirical evidence has increasingly been brought to bear on issues not only within normative ethics, but also within other philosophical sub-disciplines like epistemology or metaphysics, as well as on philosophical intuitions as such.

Joshua D. Greene has been at the forefront of a movement to bring empirical evidence into debates that are traditionally reserved for speculative moral philosophy. His research is characterized by multiple aims, involving and ranging across both negative (debunking) and positive (normative) arguments. In his early work, Greene's evinces an interest in the underlying human brain structures involved in making different kinds of moral judgments (Greene et al. 2001; 2005a; 2005b). On this and related research, he subsequently builds the negative argument that deontological moral judgments are primarily driven by knee-jerk emotional responses and therefore should be mistrusted (Greene 2008). Greene also makes a positive case for the need of a metamorality, which he calls deep pragmatism and which he develops in his book *Moral Tribes* (2013). This metamorality is meant to arbitrate when the values of different 'moral tribes' clash, in a manner similar to the way morality tempers selfish impulses within tribes. While Greene's negative program, the debunking of deontological moral judgments, has received ample attention,⁴ there is scant critical literature on his positive contribution in the form of

¹ From "Tonio Kröger" in *Death in Venice and Other Stories* (1998, 154).

² See Sytsma and Livengood's (2016) introduction for a summary of this state of affairs, and for their suggestion of exchanging the metaphor of a burning armchair, often associated with experimental philosophy, with that of a toolbox.

³ See, e.g., Alexander (2012).

⁴ See, e.g., the recent series of essays in response to Greene's views on deontological and utilitarian moral judgments in Liao (2016a).

deep pragmatism. In fact, to the best of my knowledge, there have been no article-length discussions of either *Moral Tribes* or deep pragmatism. That Greene's argument should be sound and seriously considered is important, given that what he seeks through his metamorality is nothing less than to solve "the central tragedy of modern life, the deeper tragedy behind the moral problems that divide us" (2013, 5). The ideal of constructing a metamorality that can settle public moral controversy is eminently worth pursuing, and has potential ramifications that extend beyond the purely academic.

The main objective of this thesis is to fill this gap in the literature by critically examining Greene's argument for the need of metamorality and his specific proposal of deep pragmatism. In order to properly engage his argument, however, one must first carefully unpack Greene's larger research program. This is precisely what I will do. I will examine Greene's research and some of the most pertinent and pressing philosophical questions that surround it, in order to lay the groundwork for an adequate picture of deep pragmatism. Having achieved this, I will be in a position reconstruct in their strongest form and subsequently address a number of scattered objections that have been leveled, in short reviews of *Moral Tribes*, against Greene's case for deep pragmatism. I will show that none of these is detrimental to Greene's argument, except for one form of objection concerning moral truth that reveals an internal contradiction between two lines of argument that Greene takes towards deep pragmatism. I will argue that Greene must necessarily deny the existence of knowable moral truth, which means that he has to abandon one of his arguments, namely the one with which he seeks to undermine antiutilitarian intuitions. I will show that there is a way for Greene to save this argument, which he can do by recasting the unreliability of antiutilitarian moral intuitions as undesirable with reference to cooperation. In this way, he avoids unwarranted reliance upon moral truth. Nevertheless, this way out blunts the force of Greene's argument for deep pragmatism; it also suggests that cooperation is, in the end, the superior metamorality.

The structure of this thesis is as follows. In Chapter I, I survey the most important contemporary empirical developments in relation to moral philosophy generally and to Greene's research program specifically. The chapter will largely be descriptive and clarificatory in nature, paving the way for a more focused subsequent discussion of some of the claims made by Greene. I will outline developments in the fields of experimental philosophy (I.1), metaethics (I.2), moral psychology (I.3), and moral neuroscience (I.4) that are germane to Greene's arguments. In the latter section, I also briefly address the question of what significance neuroscientific research has or might have for normative ethical theory. Drawing in part on work by Jesse Prinz, I argue that cognitive neuroscience can inform normative ethics, albeit in a more circumscribed manner than Greene's optimism suggests. What

ultimately ties the four disciples together is the positive valuation of empirical research in addressing perennial questions concerning morality. Greene's research covers each of the four fields under my analysis; situating his work within each field—drawing boundaries—will serve to enlighten subsequent discussions of Greene's research and normative claims, especially as they serve his arguments for deep pragmatism.

Chapter II is devoted to a detailed examination of Greene's dual-process theory of moral judgment. Particular attention will be paid to Greene's SLR camera analogy and to what he calls the Central Tension Principle, which holds that characteristically deontological judgments tend to be supported by automatic emotional responses, while characteristically consequentialist judgments tend to be supported by more deliberate reasoning and related cognitive processes (Greene 2016a). This principle, and the evidence on which it is based, is crucial to Greene's later argument for deep pragmatism as the best candidate for a metamorality.

In Chapter III, I consider in detail the culmination of Greene's research program and intellectual development thus far, namely his theory of deep pragmatism as covered in his monograph *Moral Tribes*. Focusing on what Greene calls the Tragedy of Commonsense Morality, which expresses the clashing of values between different groups or tribes, as well as on his No Cognitive Miracles Principle, which gives preference to more controlled reasoning over automatic emotional responses in the face of unfamiliar moral problems in evolutionary terms, I will detail why Greene thinks we need a metamorality. I will also assess the different lines of argument that Greene employs towards his specific rendering of a metamorality.

In Chapter IV, the final substantive and pivotal chapter of this thesis, I develop three lines of objection against Greene's reasoning behind deep pragmatism as metamorality. In the first section (IV.1.), I examine what I call the Automatic Mode Objection, which suggests that, rather than manual mode reasoning, automatic judgments may ultimately serve deep pragmatism better. In the second section (IV.2.), I focus on another line of criticism that I refer to as Happiness Objections. There are two versions: the first casts doubt on why tribes should renounce established systems of meaning for the sake of happiness as defined by Greene's deep pragmatism, while the second questions the pragmatic goal of happiness maximization, which is not as unproblematic as Greene makes it out to be. In the third section (IV.3.), I advance a last line of objections that I call Moral Truth Objections. There are also two forms here: the first questions why we ought to adopt a metamorality when Greene does not commit himself to moral truth, while the second reveals a contradiction in Greene's argument for deep pragmatism based on his use of competing claims about moral truth.

I will show that, while all objections must be taken seriously, only the second form of objection based on moral truth seriously threatens Greene's project. In fact, it requires Greene to give up one of his arguments for deep pragmatism. In the final section (IV.4.), I will demonstrate the necessity for Greene to deny moral truth using a (meta)science analogy. The consequence of this necessary denial, I will show, is that Greene has to drop his argument for the unreliability of antiutilitarian intuitions—based, as it is, on the affirmation of moral truth. I will suggest a way out for Greene, which is based on the reevaluation of his argument against antiutilitarian intuitions with reference to cooperation. Yet this turns out, in the final analysis, to be a double-edged sword. While it saves Greene from relying on an incoherent argument, it also opens the way for a simpler and more fundamental metamorality based on human cooperation.

I. Background: Developments in Empirical Philosophy

The aim of this chapter is to provide a backdrop to Greene's work, by sketching some developments in four distinct yet related disciplines: experimental philosophy, metaethics, moral psychology, and moral neuroscience. The respective fields are vast, and a complete overview far exceeds my scope. I will therefore focus on the specific aims and methodologies that are directly relevant to understanding Greene's project. The overarching and unifying theme across the disciplines, I will argue, is the application of empirical methods and findings to philosophical concepts, questions, and problems generally, and to ethics specifically. An understanding of some of the manifestations of this theme across apparently divergent disciplines will put us in a better position later on to critically examine the details of Greene's account of deep pragmatism—forged, as I will show it is, out of the following approaches.

I.1. Experimental Philosophy

Early experimental philosophy may be distinguished by the explicit goal to empirically investigate philosophical intuitions (see, e.g., Alexander 2012; Knobe 2007; Knobe and Nichols 2008). As Alexander and Weinberg put it, experimental philosophy is "unified behind a common methodology and a common aim: the application of methods of experimental psychology to the study of the nature of intuitions" (2007, 56). This investigation was prompted by the perceived⁵ widespread use of intuitions in philosophical theorizing. Reliance on intuitions in thought experiments and theory building was considered by early proponents of experimental philosophy to be a ubiquitous feature of philosophical practice and, importantly, it was at the same time viewed as a practice with potentially problematic limitations (of generalizability, for instance) that could and should be tested empirically.

As an illustration, consider the following. Traditional philosophical discussions of freedom and moral responsibility, as Alexander (2012) points out, were often based on the assumption that people are natural incompatibilists, believing that causal determinism is ultimately incompatible with free will

⁵ The prevalence of intuitions—and, importantly, reliance on intuitions—in (analytic) philosophy has been a matter of considerable debate. For instance, while George Bealer (1996; 1998) has argued that philosophical intuitions belong to philosophy's "standard operating procedure" (1996, 122), others, like Timothy Williamson (2004) and Herman Cappelen (2012; 2014), have argued that philosophy does not rely on intuitions as commonly conceived. Max Deutsch (2015) provides an extensive account of the current and complex state of intuitions in (experimental) philosophy. I do not take a stance on the issue; my goal here is to position Greene in the debate.

or moral responsibility. This assumption was then used to place the argumentative burden on philosophical compatibilists rather than incompatibilists. In an early and important set of studies on moral responsibility and free will, Eddy Nahmias, Stephen Morris, Thomas Nadelhoffer, and Jason Turner (2004; 2005; 2006) empirically investigated the assumption that people are natural incompatibilists. Using a series of vignettes describing different scenarios, they demonstrated that people are actually more likely to be natural compatibilists than incompatibilists. To give a more concrete example, in a study of what they call *Jeremy Cases*,⁶ Nahmias et al. (2005) presented participants with the following scenario:

Imagine that in the next century we discover all the laws of nature, and we build a supercomputer which can deduce from these laws of nature and from the current state of everything in the world exactly what will be happening in the world at any future time. It can look at everything about the way the world is and predict everything about how it will be with 100% accuracy. Suppose that such a supercomputer existed, and it looks at the state of the universe at a certain time on March 25, 2150 AD, 20 years before Jeremy Hall is born. The computer then deduces from this information and the laws of nature that Jeremy will definitely rob Fidelity Bank at 6:00 pm on January 26, 2195. As always, the supercomputer's prediction is correct; Jeremy robs Fidelity Bank at 6:00 pm on January 26, 2195. (566)

They subsequently asked participants to suspend disbelief and to imagine the details of the scenario to be true. Crucially, they then asked participants whether they thought that, when Jeremy robs the bank, he does so of his own free will. What they found is that a significant majority of participants (76%) reckoned that Jeremy robs of his own free will. The normative conclusion that Nahmias and his colleagues draw from this finding is that, if the burden of proof should be on anyone, it ought to be on those who wish to assume that people tend to be natural incompatibilists—contrary to what the evidence shows about people's actual, observed beliefs about (in)compatibilism.

Thus, philosophical intuitions about the kinds of beliefs that people actually hold, in this case concerning free will and moral responsibility, were challenged by means of empirical findings. Nichols and Knobe (2007) subsequently took up this line of research to show that *affect* plays a crucial role in people's intuitions in response to the sorts of vignettes used by Nahmias and colleagues, so that people

⁶ For the sake of concision, I omit some of the details of the Jeremy Cases as well as the other studies that are included in Nahmias et al.'s (2005) paper.

in fact have both compatibilist and incompatibilist intuitions that are generated by different underlying psychological processes. Taken together, this research demonstrates 1) a tendency to put philosophical intuitions and assumptions about these intuitions to the test, and 2) an impetus towards investigating the basic psychological processes behind these intuitions.

Another trend has been to examine not just the details of philosophical intuitions that people have as such, but also whether there are systematic differences (or, more strongly put, predictable biases) when it comes to these intuitions. Examples of this kind of research include the probing of intuitive differences across genders (Buckwalter and Stich 2014), cultures (Haidt and Joseph 2007; Buchtel and Norenzayan 2008), and personality traits (Bartels and Pizarro 2011). The underlying thought, and the predominant conclusion, is that philosophical intuitions are not uniform across human beings; there are real differences in the sorts of intuitions that different people (or, more specifically, people with different, what may be taken as morally irrelevant characteristics) have. In another yet methodologically related line of research, moral intuitions were also found to be sensitive to such phenomena as order effects (Wiegmann, Okan, and Nagel 2012), which casts doubt not just on the inter-personal but also on the intra-personal stability of intuitions. The supposed expertise of professional philosophers has also been studied and challenged, with philosophical expertise having been found to provide no guarantee of stability against order effects or other biases (Schwitzgebel and Cushman 2012; see also Weinberg, Gonnerman, Buckner, and Alexander 2010). The nature of folk concepts and intuitions, and the predictions that philosophers make about them, have also been the subject of study, criticism, and subsequent refinement (Nichols 2004b; Kauppinen 2007).

Later developments in experimental philosophy have moved beyond purely considering intuitions;⁷ a recent self-definition, for instance, broadly characterizes its aim as to "systematically collect and analyze empirical data in attempting to answer philosophical questions or solve philosophical problems" (Sytsma and Livengood 2016, 5). To date, research under the label of experimental philosophy⁸ has been conducted in many different traditional areas of philosophy, including, *inter alia*, metaphysics (Nichols and Bruno 2010), philosophy of language (Machery, Mallon, Nichols, and Stich 2004; Haukioja 2015), philosophy of mind (Knobe and Prinz 2008; Sytsma 2014), intentionality (Knobe 2010), and epistemology (Buckwalter 2010; Feltz and Zarpentine 2010).

⁷ Though the field is no longer restricted to intuitions, and non-intuitional work is being conducted, the study of intuitions still has a dominant role in experiment philosophy (see, e.g., Sytsma and Livengood 2016).

⁸ It should be noted that there is some debate about what precisely should count as experimental philosophy. For instance, ought it to include only the work of those self-consciously active within the field, or should the work of non-self-identifying researchers also be embraced? Furthermore, if the latter, then what are the criteria that should make it count? See, for an empirical study of philosophers' views on these matters, and for a summary of the debate, Sytsma and Livengood (2016).

Sytsma and Livengood (2016) have offered a helpful classification scheme for work that has been and is currently being done in experimental philosophy.⁹ They first differentiate between two general programs—the intuitional and the non-intuitional—according to whether research is or is not, respectively, concerned directly with intuitions. Considering that the foundational work in the field was mostly constrained to intuitions, the authors refrain from further subdividing the somewhat eclectic—at least presently—program of non-intuitional experimental philosophy.

Within the intuitional program, however, they further distinguish two approaches—the evidential and the neutral—according to whether, respectively, research involves prescriptive or descriptive questions and aims. On the evidentiary, prescriptive side, normative questions about intuitions predominate (e.g., should intuitions play a role in philosophical theorizing?), while on the neutral, descriptive side the prevailing concern is broadly with what intuitions are and how they operate.

Inside the evidentiary intuitional program, there are two additional subdivisions: the negative program, which seeks to undermine evidentiary faith in intuitions, and the positive program, which lends at least partial support to the evidential use of intuitions. While advocates of the negative program are generally "pessimists about the use of intuitions as evidence or think that we should suspend judgment about the reliability of our intuitions until we know what factors affect them" (40), advocates of the positive program are "more optimistic ... [and] tend to think that at least some intuitions track the truth" (41). As Sytsma and Livengood emphasize, the main differences between the negative and positive camps concern expectations and use: whereas so-called 'positives' anticipate finding "an interesting signal in the noise of human intuitions" (41), professed 'negatives' often overtly run empirical studies to "raise doubts about appeals to intuition in philosophy" (40).

Finally, within the neutral intuitional program, there are two more subcategories: the cognitive program, where the principal aim is to identify the "psychological and neurological underpinnings of our intuitions" (43), particularly as these relate to traditional philosophical subjects, and the descriptive program, which is not interested in psychological mechanisms, but instead seeks to account for how people "actually talk about philosophically interesting topics, including how they use intuition-talk and related speech-acts" (43). Figure 1 provides a schematic overview of the aforementioned projects within experimental philosophy, adapted from Sytsma and Livengood (2016, 36).

⁹ Subsequent page references to this work are included parenthetically in the text.

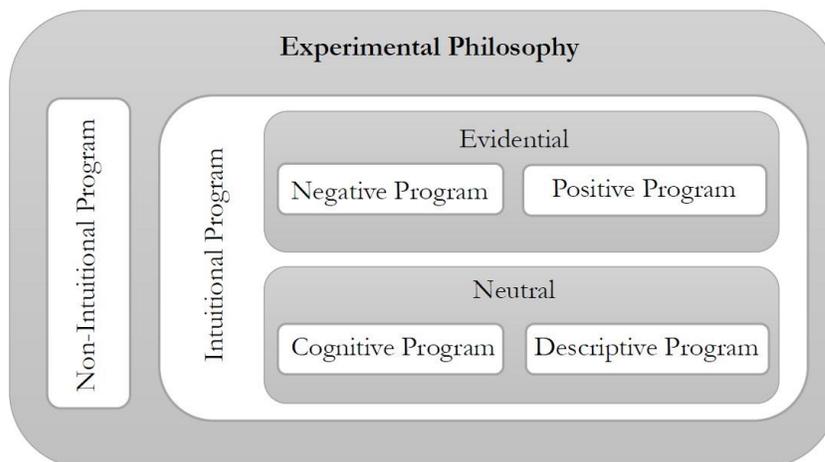


Figure 1: Classification scheme for research programs within experimental philosophy.

At this juncture, the question is how Greene's research fits within the larger scheme of work being done in experimental philosophy. As Sytsma and Livengood admit, researchers rarely explicitly describe themselves as engaged in one of the programs as opposed to another, nor does their work tend to fit neatly within the boundaries of any one of the programs. Greene is no exception. A predominant concern for him, especially in his early work, has been to undermine certain kinds of intuitions—more specifically, deontological ones, or ones that are justified in terms of rights and duties (2008; 2016a). These intuitions, he has argued, are principally driven by emotional responses, which leads him to question reliance upon them when it comes to moral reasoning. From the beginning, Greene has valued utilitarian intuitions—those justified by the consequences that derive from a particular action—but by the time of *Moral Tribes*, Greene will make a full-blown positive case for a utilitarian ethic. In this way, then, Greene may be situated within the evidential intuitional program, contributing both to the negative program by devaluing deontological intuitions, and to the positive program by championing utilitarian intuitions.

I.2. Metaethics

Even though Greene's research does not explicitly address either the field of metaethics as a whole or the details of its distinct debates, metaethical issues are clearly at issue in his work—especially in *Moral Tribes*, when he turns his attention to the metaethical approach he calls deep pragmatism.

Curiously, though, while he will go on in *Moral Tribes* to defend the need for a metamorality, there is no mention in this work of metaethics; nor are particular developments or academic debates within the field of metaethics addressed. Greene concedes at the outset of *Moral Tribes* that "[t]he idea of a metamorality" is not wholly new," and that concern with specifying universal moral principles has been "a dream of moral philosophy since the Enlightenment" (26). Yet he leaves out entirely the scholarly tradition that has honed in on distinctively metaethical questions, although he at least appears to pay lip-service to the tradition when he urges that what we need to solve contemporary moral problems is something more than first-order morality (26). It is precisely the departure from first-order ethics that marks the field of metaethics. As Fisher (2011) puts it, metaethics takes a "bird's-eye view on the practice of ethics: the metaethicist peers down as intently as possible and tries to make sense of what is going on" (2). Similarly, Van Roojen (2015) holds that metaethics "involves reflecting on the nature of ethics," which it does at a more general, abstract, and reflexive way than, for instance, first-order normative or applied ethics does (1). Miller (2003) comes even closer to expressing Greene's proposal in metaethical terms, when he distinguishes between first-order questions like "Ought one to do X?", which constitute the domain of normative ethics, and second-order questions like "What are people doing when they are debating whether one ought to do X?", which are within the realm of metaethics (1). This latter question is nonetheless exactly the one that Greene asks in the beginning of *Moral Tribes*, and which he will pursue using some of his own research.

In short, Greene has at his disposal a tradition of thought concerned with higher-order ethics, namely the field of metaethics, which he chooses to ignore. What is clear is that he builds his metamorality of deep pragmatism out of what he terms basic morality; differently put, first-order ethics will serve to inform his second-order metaethical account. One possible reason for the neglect of metaethics proper might be the intended audience of his work, since *Moral Tribes* is a book intended for the general public. However, Greene does not otherwise shy away from engaging with and explaining topics of considerable complexity. There therefore does not seem to be, *prima facie*, a strategic or presentational reason why he should have avoided drawing upon at least some of the most relevant academic strands of metaethics.

Precisely because metaethics is relevant to the nature of *Moral Tribes*, and to the metamorality that Greene develops in it, it is worth sketching some relevant debates in the field. Particularly significant is the issue of cognitivism versus non-cognitivism with regard to psychological states. For any given moral judgment, one might ask what sort of psychological state it expresses. On the one hand, there are cognitivists who hold that moral judgments express beliefs, which by their nature can

be true or false: "they are *truth-apt*,¹⁰ or apt to be assessed in terms of truth and falsity" Miller (2003, 3). That is, according to cognitivists, moral judgments are beliefs that lend themselves to being true or false. There are many approaches to cognitivism, and while a full exploration lies outside my scope, two broad forms must be distinguished, as they will serve to inform later discussions of Greene's metamorality: strong cognitivism and weak cognitivism.¹¹ For *strong* cognitivists, moral judgments are both apt for evaluation in terms of truth and falsity, and they can be the outcome of "cognitively accessing the facts which render them true" (Miller, 2003, 4). So, not only can moral judgments be true or false, but this evaluation can be the result of having access to that which makes the judgment true or false. For *weak* cognitivists, moral judgments are also apt to be assessed in terms of truth and falsity; however, this cannot be the result of having access to the facts that so render them (Miller, 2003, 5-6). To illustrate the difference between a weak and a strong theory of cognitivism, Miller (2003) writes:

[A] weak cognitivist theory would be one which held that our best judgements about morals determine the extensions of moral predicates, rather than being based upon some faculty which tracks, detects or cognitively accesses facts about the instantiation of moral properties. (6)

In its weak form, then, moral judgments can be true or false—without thereby being based, as in the strong form, on "a faculty with a tracking, accessing or detecting role" (Miller, 2003, 6).

On the other hand, noncognitivists espouse the negative claim that moral judgments as such are not cognitive, and they generally deny that moral judgments are cognitive states (van Roojen, 2015). Instead, non-cognitivists tend to hold that moral judgments are states of emotion or desire. They are therefore not truth-apt; one cannot express them, nor do they allow for evaluation, as true or false.

These and related questions are also found in discussions of moral realism, which, broadly taken, is the view that there are "moral properties and moral facts (in which those properties figure) whose existence and nature are independent of the stances of individuals and groups" (Horgan and Timmons, 8). Moral realism, in turn, is part of a larger debate in metaphysics and logic, as well as in ethics, between realism and antirealism (Butchvarov 2006).

The answer to why Greene effectively ignores metaethics appears to lie in his disinclination to get involved with metaphysics. He explicitly refuses to take a stance on whether or not there is moral

¹⁰ Emphasis in the original.

¹¹ Within these forms there are many further distinctions, such as between naturalism and non-naturalism in the strong cognitivism version (Miller, 2003); my aim, again, is not to be exhaustive, but merely to delineate the relevant opposing sides as an anchor for Greene's position.

truth in *Moral Tribes*, choosing to "remain agnostic" on the matter instead (188). In the end, his refusal to engage with the metaphysical aspects of metamorality may be why he skirts some of the more technical philosophical (i.e., metaethical) questions and implications that arise in his treatment of deep pragmatism. That this stance is problematic will become clear later, when I develop Greene's arguments in greater depth. However, what should be apparent now is that, whether or not Greene wants to entertain metaphysical notions, his empirical research project generally and his theory of deep pragmatism specifically draw on claims that commit him to a specific position in at least some of the aforementioned metaethical debates. Aside from Greene's problematic use of moral truth, a large portion of his research is devoted to studying the emotional and cognitive neurophysiological underpinnings of moral judgments, from which he will also draw conclusions to support his metamorality. When Greene makes inferences from the kinds of psychological and neurophysiological mechanism supposed to feed into different kinds of moral judgment—taking a bird's-eye view of what is going on (in the brain) when people make moral judgments and evaluating the normative ethical validity of these judgments based on details concerning underlying brain activity—he is actively involved in metaethics.

I.3. Moral Psychology

Moral psychology is sometimes treated under the umbrella of metaethics (e.g., Fisher 2011). There is a certain amount of thematic overlap between the two, and psychological research—particularly as it pertains to morality—has immediate implications for metaethics, which is a view that is embraced by the so-called 'new moral psychologists' (Tiberius 2015, xi). However, moral psychology can also be seen as a more or less independent discipline with its own historical questions and aims that dates back in early modern times to philosophers like Samuel Clarke, Francis Hutcheson, and David Hume (Nadelhoffer, Nahmias, and Nichols 2010), while its roots lie—one might say of course—in ancient times with the reflections of Plato and Aristotle on the roles of reason and emotion in moral judgments (Sinnott-Armstrong 2008b). The contemporary field, as I will treat it, grew out of a relatively recent interdisciplinary movement that has wrested the study of morality away from its privileged place within philosophy, towards and in a spirit of integration with psychology and related empirical sciences (Doris 2012). As Sinnott-Armstrong (2008a, xiii) explains, G. E. Moore's

introduction and indictment of the 'naturalistic fallacy'¹² in his 1903 *Principia Ethica* did much to convince later thinkers within moral philosophy that empirical science is irrelevant to their subject.

A concerted effort by philosophers to "mine cognitive psychology and brain science, as well as evolutionary biology, for general philosophical lessons" may be roughly identified in the 1990s (Sinnott-Armstrong 2008a, xiii). Moral psychology as it is today may be viewed as "a hybrid theory, informed by both ethical theory and psychological fact," which is best exemplified by the kinds of questions it asks and the sorts of responses that are given (Doris 2012, 1). For instance, consider one of the central issues in moral psychology: What is the nature of moral judgment? As Doris points out, an inquiry of this kind demands empirically informed answers, while "developing these answers in theoretically sophisticated ways requires delicate inquiry in philosophical ethics" (2012, 1). One needs both the psychological facts as well as robust ethical theories to adequately address the predominant questions of moral psychology.

As a case in point, consider the situationism debate, which concerns the degree to which moral behavior is influenced by (morally irrelevant) situational factors as opposed to stemming from more robust or global character traits or moral character (Upton 2009). Proponents of situationism in what is known as the second phase of the debate have argued that situationist findings undermine traditional virtue ethics (Harman 1999; Harman 2000; Doris 1998). Doris uses findings from studies within psychology on mood and helping behavior to argue just how sensitive the latter is to situational factors (2002). Isen and Levin's (1972) classic study demonstrated that inducing a good feeling in people, by giving them cookies or by having them find a dime in a phone booth, significantly affected their subsequent helping behavior; when people were made to 'feel good' they were more helpful afterwards compared to control subjects. Doris reasons, from many related and similar findings, that "characterological moral psychology is an empirically inadequate approach to the determinants of helping behavior" and uses Milgram's studies on obedience (2009) and Zimbardo's prison experiments (2007), among others, to argue that the same holds for the determinants of destructive behavior (2002, 38). The details of the debate extend beyond what I can convey here, and Doris's account has received its fair amount of criticism (e.g., Sabini and Silver 2005; Kleingeld 2015). What the situationist debate

¹² As Oliver Curry (2006) points out, there are historically many arguments that go by the name of 'naturalistic fallacy'. Moore's original concern was with "identifying good with its object"; later philosophers have adjusted the meaning(s) of the fallacy, so that Peter Singer has taken it as moving from facts about the worlds to values, while Simon Blackburn has held it as misidentifying an ethical with a natural concept (Curry, 2006, 236). A genealogy of the natural fallacy lies outside my scope; what is important is the effect that Moore's original criticisms had on the later development (or stagnation) of empirical science in moral philosophy.

and Doris's particular contribution brings to light is how psychological and behavioral findings may be used in the generation and defense of moral theory.

Greene's research falls squarely within moral psychology. The central issue for much of the work within the field—the nature of moral judgment—characterizes a large segment of his research program. His early research involved an fMRI investigation of emotional engagement while people were confronted with moral dilemmas, in which he set out to distinguish the roles of reason and emotion in moral judgment (Greene et al. 2001). In the article, Greene supplements his empirical findings with philosophical speculation when he asks, as a concluding remark: "How will a better understanding of the mechanisms that give rise to our moral judgments alter our attitudes toward the moral judgments we make?" (Greene et al. 2001, 2107¹³). Within this question lingers a normative ethical claim: In such a case as the psychological mechanism of moral judgment(s) being X, one must—from a normative ethical perspective—take towards it stance Y or Z. As will become clear in later sections, the hallmark of Greene's program is to combine moral psychology (i.e., the psychological mechanisms that underlie judgments and reasoning within the moral domain) with philosophical implications based on its particular characteristics.

I.4. Moral Neuroscience

As S. Matthew Liao (2016b) points out, the neuroscientific study of emotions in decision making—and later work on emotions in moral judgment—is indebted to a pioneering series of experiments by Antonio Damasio and colleagues in the 1990s. These studies, in particular of persons with damage to the ventromedial prefrontal cortex (VMPFC), brought attention to the crucial role of emotions in decision making (Damasio 2006). People with damage to the VMPFC seem to have many of their cognitive abilities intact, yet they are notoriously bad at making real-life decisions—a fact that may be attributed to impaired emotional functioning (Damasio 2006; Bechara et al. 1994). For instance, in a card game people with intact VMPFCs tend to show anticipatory skin conductance responses (SCRs) to cards associated with consistent earlier rewards or penalties, while people with damage to the VMPFC do not evince such anticipatory SCRs, thereby showing an indifference to future outcomes (Bechara et al. 1996).

¹³ This is the correct page number.

Greene took an early interest in the identification of certain brain structures underlying processes related to emotion and cognition in an attempt, furthermore, to extend these more specifically to moral judgment (Greene 2005a). In this work he applied the methods of cognitive science to moral judgment (Greene et al. 2001; Greene et al. 2004), sought to classify how and where moral judgment works (Greene and Haidt, 2002; Greene 2005b), and tried to clarify the implications of neuroscientific moral psychology for ethics (Greene 2003) as well as for the law (Greene and Cohen 2004). Greene also turned his attention to the VMPFC specifically by linking it to utilitarian kinds of moral judgment (Greene 2007). This latter move is related to the dual-process model that he will go on to develop, and which I will discuss in detail in the next chapter.

The question concerning the extent to which emotion and reason underlie moral judgment has historical roots beyond the emergence of neuroscience, for instance in an important debate in the 17th and 18th centuries between moral rationalists like Ralph Cudworth, Samuel Clarke, and John Balguy, who held that morality resided in reason alone, and moral sentimentalists, like the Third Earl of Shaftsbury, Francis Hutcheson, and David Hume, who considered morality to originate at least partially with human sentiment (Gill 2007). The rationalist-sentimentalist debate has taken many forms over time; what matters for present purposes is that Greene's work can be placed in relation to modern adherents of sentimentalism like Prinz (2006; 2016) and Nichols (2004a; 2008), who argue that moral judgments are emotional in nature.¹⁴ Much of the debate between sentimentalists and rationalists is waged within moral psychology and metaethics; to the extent that brain studies are used to inform such debates, however, these may also be classified as moral neuroscience. Greene's concern with the extent to which moral dilemmas engage emotional processing, as opposed to processes related to more explicit reasoning, certainly falls within this larger debate.

One important point must be addressed here, namely the relevance of cognitive neuroscience for normative ethics. It should be clear, through my earlier discussion of experimental philosophy for example, that I consider empirical data to be philosophically meaningful at least potentially. However, Selim Berker (2009) has launched an influential critique of Greene's research in which he not only challenges the latter's methodology and statistical analyses (and, accordingly, his conclusions), but in which he also questions the significance of neuroscience for normative ethics per se. There is no space to go into the specifics of Berker's criticisms of Greene's analyses; Greene has responded meticulously to Berker concerning those matters (2010) and has either deflated the criticisms or used them to

¹⁴ Their arguments are importantly different, and sentimentalist theories more generally take on many shapes; I oversimplify somewhat in order to sketch the background to Greene's research without losing sight of larger aims.

sharpen his procedures or even to re-run earlier statistical analyses. The question of the significance of neuroscience to normative theory is decisive, however, for if it is true that, as Berker puts it, "once we separate the bad arguments for why Greene et al.'s empirical research has normative implications from the better arguments for that conclusion, we can see that the neuroscientific results are actually doing no work in those better arguments" (2008, 294), then Greene's research project appears to be significantly impaired.

Not necessarily, though. Greene, as he emphasizes in his response to Berker, has "never attempted to draw normative conclusions from brain scans alone, or even from neuroscientific data alone," and believes that the implications of neuroscientific data for normative ethics are indirect and "depend on further non-neuroscientific ideas and assumptions" (2010¹⁵). Prinz (2016) holds a similar view, arguing that, taken on their own, "extant neuroimaging studies leave classic debates unsettled, and require other evidence for interpretation," which is why philosophical and psychological accounts are needed to settle the meaning of brain scans—rather than the other way around (45). Prinz has it right, and Greene, if he did not realize it before, has by the time of *Moral Tribes* understood that one cannot argue for an ethical theory purely from empirical neuroscientific data. Evidence of this understanding comes in the form of the multiple approaches (empirical and theoretical) that Greene takes towards the defense of his metamorality in that work.

The Berker-Greene debate has, in any case, led Greene to refine some of his arguments. His latest statement (2016) on the role of science in normative ethics is as follows:

Science can advance ethics by revealing the hidden inner workings of our moral judgments, especially the ones we make intuitively. Once those inner workings are revealed, we may have less confidence in some of our judgments and the ethical theories that are (explicitly or implicitly) based on them. (119)

This claim is strikingly close to the kinds made in the negative evidentiary program of experimental philosophy, as outlined by Sytma and Livengood (2016). It also sidesteps unwarranted reliance on neuroscience, for Greene—as will become clear in later chapters—will bring in philosophical argumentation to supplement empirical research. While cognitive neuroscience can inform philosophical questions generally and normative ethics particularly, it does so in a more

¹⁵ Page numbers are not indicated in the text; quotations are from 'page 8' of the online document (Greene 2010).

circumscribed manner than Greene's initial optimism suggests—tempered by Prinz's warning. Whether Greene's conclusions based indirectly on neuroscientific data will hold therefore ultimately depends on the specific inferences that Greene draws from them and on their supplementary philosophical argumentation.

Before moving on, it should be emphasized that the neuroscience behind phenomena like emotion, reason, and moral judgment is by no means settled; interpretations of extant results remain contested as they become subjected to renewed interpretation, while novel studies and findings are continually emerging. As such, an asterisk should be placed behind many of the findings to indicate their precariousness. This is as it should be, however, within the hardly linear trajectory of scientific progress. What can be done at any point in time is to make use what is known, and this is what Greene does, as will become clear in the next chapter. The neuroscientific details of his account are likely to change over time, but since Greene's theory uses but does not rise or fall solely on these details, he might well be able to accommodate such change. In any case, that neuroscience be significant to ethical theory is a triumph for his cause.

An underlying theme of experimental philosophy, metaethics, moral psychology, and moral neuroscience is that the empirical study of moral thought and behavior—and their neurological underpinnings—is relevant to philosophical questions and problems. Greene picks up on the developments and affordances provided by each field of study in formulating his own line of research, in particular his dual-process modal of moral judgement to which I will now turn.

II. Greene's Dual-Process Theory

Dual-process theories have become a common explanatory model across different fields of psychology (Chaiken and Trope 1999; Brand 2016; Frankish 2010). Such models all tend to be indebted to pioneering work by Daniel Kahneman on the distinction between two kinds of thinking and decision-making: one fast, implicit, effortless, and unconscious; the other slow, explicit, effortful, and conscious (see, for a summary, 2011). Kahneman connected the fast and unconscious process to intuitions and the slow and conscious process to reasoning; from there, dual-process models began to be developed in moral psychology to illuminate different kinds of moral judgments and reasoning processes (Brand 2016). An early and well-known example of the application of this model to moral judgment is Jonathan Haidt's (2001; 2013) social intuitionist model, which holds, generally, that when people engage in moral judgment, intuitions come first and strategic reasoning second; put more strongly, Haidt has argued that moral reasoning is but a post-hoc rationalization of prior intuitive judgments.

Greene has proposed his own version of a dual-process model of moral judgment (Greene et al. 2001; Greene et al. 2004; Greene 2007; Greene et al. 2008; Greene 2008). The details of his model have changed over time, as criticisms and additional research has led to refinements and reformulations; the most recent expression of Greene's model, on which I will focus, was published in 2014.¹⁶ The first thing to discuss in relation to the model is Greene's SLR camera analogy, which he also makes use of in *Moral Tribes*. The analogy goes as follows. An SLR camera has two complementary modes: a point-and-shoot automatic mode and an adjustable manual mode. This dual-mode design, Greene argues, "exemplifies an elegant solution to a ubiquitous design problem, namely, the trade-off between *efficiency* and *flexibility*"¹⁷ (2016, 120). The automatic setting is highly efficient but inflexible, while the opposite holds for the manual setting. Greene, drawing on Kahneman (2003), maintains that the human brain has the same overall design and associates the automatic settings of the brain—the "reflexes and intuitions that guide our behavior"—with predominantly emotional responses, while he links manual mode—a "general purpose reasoning system"—with behaviors and goals that are typically longer-term, non-automatic, conscious, effortful, and rule-based (120). Before moving on to the principle that Greene seeks to derive from his analogy, he cautiously articulates two points of

¹⁶ It was reprinted as Greene (2016); this is the version to which I will refer, with page numbers in the following section indicating the location in this text.

¹⁷ Emphasis in the original.

disanalogy between SLR cameras and the human brain, namely 1) a camera must be in one mode or the other, while the brain's automatic settings are always operating, and 2) there is asymmetrical dependence in the human brain: while one can have automatic settings without manual mode (as in most animals), the reverse is not viable, whereas in a camera the two modes can and do function entirely independently (121).

Having developed this analogy as groundwork for his dual-process model, Greene goes on to specify the model's "more interesting, and correspondingly more controversial" tenet in relation to moral judgment, which he calls The Central Tension Principle (CTP) and is stated as follows:

CTP: Characteristically deontological judgments are preferentially supported by automatic emotional responses, while characteristically consequentialist judgments are preferentially supported by conscious reasoning and allied processes of cognitive control. (121)

The CTP is the culmination of Greene's work on moral judgment and moral dilemmas, in which he has examined, among other things, the neural substrates of utilitarian and deontological moral judgments. The tension between utilitarianism and deontology within ethics, Greene stresses, is rooted in the previously discussed Kahnemanian tension in cognitive design between efficiency and flexibility. While Greene's earlier treatment of deontology in particular tended to be crude and was duly criticized, his use of 'deontological' and 'utilitarian' are still, as he admits, not wholly consistent with standard philosophical use (121). He therefore applies the qualifier *characteristically* and explains that characteristically deontological judgments are ones that are "naturally justified in deontological terms ... and that are more difficult to justify in consequential terms," while characteristically utilitarian judgments are those "naturally justified in consequentialist terms ... and that are more difficult to justify in deontological terms" (122). As examples of the former he evokes appeals to rights and duties, while for the latter he suggests appeals to impartial cost-benefit reasoning.

As evidence for the CTP, Greene mobilizes a large body of research both of his own and conducted by others.¹⁸ The most significant and frequently engaged line of evidence comes from experiments with different versions of a moral dilemma that is known as the trolley problem (Foot 1969; Thomson 1985). The trolley problem directly inspired Greene's dual-process model (2016c). Although there are various versions of the problem,¹⁹ I will focus on two trolley cases—the *switch* case

¹⁸ For a summary, see Greene (2016, 122-126).

¹⁹ For a summary, see Edmonds (2014).

and the *footbridge* case—as these are also Greene's main concern and form the bulk of his argument. In the *switch* case, a wayward trolley is heading towards a group of five people that it will hit and kill. The dilemma consists in whether or not one ought to divert the trolley to head it towards another track containing only a single person, who will then inevitably be killed instead. The *footbridge* case has a very similar logic, except that in order to halt the trolley heading towards the group of five, there is the option to drop a person with a heavy backpack (or in variations, a fat man) onto the path of the trolley. This action will stop the trolley and therefore save the five people—yet it will kill the person who was dropped. Both cases operate on the same calculus: to save five lives at the expense of a single other. Nonetheless, studies have revealed that people tend to endorse this calculus significantly more in the *switch* case than in the *footbridge* case; redirecting the trolley is consistently judged to be more morally permissible, on average, than stopping the trolley by means of pushing a person onto its path (Greene 2008). This finding is curious due to the fact that, if people consistently employed an ethical theory, then moral judgments should be aligned across the two dilemmas. If one were operating from a utilitarian ethic, then in both cases one would be expected to judge that the action with the best consequences (saving five lives over one) ought to be endorsed. On the other hand, from a deontological ethic, in both cases one would be expected to strictly oppose the use of a person as a means to saving others—no matter what the consequences.

The articulation of possible reasons for this difference in intuitions across different versions of the trolley cases has a formidable philosophical history, beginning with Foot (1967). Important for present purposes, however, is what Greene makes of it. According to Greene's CTP, deontological judgments are preferentially supported by automatic emotional responses; or, as he has also put it, deontological moral judgments are driven by what he has called 'alarm bell' emotion, while consequentialist judgments are supported by more controlled reasoning (2008). In the *footbridge* case, for instance, the physical nature of pushing a man off the bridge signals an alarming emotional response, which has to be overridden in order to arrive at a utilitarian judgment. Greene traces this emotional response to our evolutionary past, where 'up close and personal' violence posed a problem for life within groups, while relations between more distal groups, unlike in today's globalized state, did not (see also Singer 1972; 2005; 2011). Greene's argument here is an example of what have become known as evolutionary debunking arguments (EDAs), which appeal to the etiology of beliefs—in this case evolutionary origins—to undercut their justification (Kahane 2011; Wielenberg 2016; Nichols 2014). That evolutionary explanations for moral beliefs may threaten the view that human beings possess moral knowledge has a considerable history, with influential EDAs having been proposed in

various forms (Wielenberg 2010; Singer 1982; Street 2006; Joyce 2006). Greene is aware of this tradition; he asks, for instance, in his 2008 article, how far the empirical debunking of human moral nature can go, and identifies this question as among the most fundamental moral questions we face today. He also explicitly partakes in this kind of argumentation when he insists on both the evolutionary origin and the moral inconsequence of certain features of problems on which deontological judgments rely. The distinction between personal and impersonal forms of harm, Greene maintains, is not just explained (away) in evolutionary terms, but it is also contingent and morally irrelevant. The inconsistency in responses to the *switch* and *footbridge* cases is based on a morally irrelevant feature: whether one physically pushes a person off a bridge to save five lives or whether one presses a switch to achieve this end is not in itself a morally relevant component of the problem (Greene 2013, 217). Furthermore, deontological judgments, as they are based on more automatic emotional responses, appear to be especially sensitive to this; in fact, they can be oversensitive, "responding to things that, upon reflection, don't seem to be morally relevant," or they can be undersensitive, "failing to respond to things that, upon reflection, do seem to be morally relevant" (Greene 2013, 212). In the trolley cases, this means for Greene that deontological moral judgments are unreliable, for example to the extent that they respond to personal force and are thus tracking an irrelevant moral feature.

One further and related way that Greene attempts to explain intuitional difference between the trolley cases is with what he calls the Modular Myopia Hypothesis (MMH). He elaborates this hypothesis in *Moral Tribes*, where he offers the MMH as a "synthesi[s of] the dual-process theory of moral judgment with a theory about how our minds represent actions" (2013, 224). Succinctly put, Greene postulates a system located in the brain that inspects action plans and that is highly sensitive to, and alarms us of, acts of violence within the plan's primary causal chain (2013, 226). The MMH is supposed to map onto some of the features of the trolley problem, as when personal force is encountered and has to be accounted for in the *footbridge* case as opposed to the *switch* case. The presence of an act of physical violence (i.e., pushing the fat or backpacked man onto the tracks) in the primary action plan of the *footbridge* case rings the alarm—in Greene's language—leading to deontological moral judgments, while the absence of violence accordingly leads to a less emotion-driven response in the form of utilitarian moral judgments.

It must be noted, finally, that dual-process theories of cognition as well as having been defended (e.g., Evans and Stanovich 2013) have also been criticized (e.g., Gigerenzer 2010). Greene's dual-process model is no exception; it, too, has also been challenged from more than one source (e.g.,

Kahane 2012; Darwall 2016; Driver 2016). Most of the arguments have centered on issues concerning the interpretation of empirical data and the extent to which the available neuroscientific evidence in fact supports a dual-process model of moral judgment. These arguments tend to become technical quickly, and I will not deal with them here. Suffice it to say, for present purposes, that Greene is aware of the criticisms leveled against his model and the interpretations on which it relies, and has accordingly defended his case (Greene 2016b).

With a more thorough understanding of Greene's dual-process model of moral judgment, we are in a position to examine the theory which it informs, namely deep pragmatism. I will treat Greene's theory and the details of his arguments for deep pragmatism in the following chapter.

III. Moral Tribes and Deep Pragmatism

Greene argues in *Moral Tribes* that we need a metamorality to solve what he views as a new kind of moral problem. He argues for his preferred form of metamorality, which he calls deep pragmatism, from a host of different sources; it is informed by, among other things, his own neuroscientific research, evolutionary psychology, and consequentialist moral philosophy. His most sustained treatment of deep pragmatism is found in *Moral Tribes*, which is why I will focus on the arguments from that book. Where needed, however, I will bring in additional formulations and arguments that Greene makes outside of his monograph.²⁰

Moral Tribes opens with a double tragedy that sets the tone and agenda of the book. Greene uses the metaphor, or rather allegory, of different tribes coming to live on new pastures²¹ to illustrate and differentiate between the two tragedies. The first, the Tragedy of the Commons (ToC), occurs as in the classic case²² when there is intratribal conflict—when there is conflict between, as Greene characterizes it, "Me versus Us," or selfishness set against concern for others (14). The ToC is the classic dilemma of cooperation, of how to coordinate group-level interests among individuals each with their own self-regarding interests, which our brains were "designed to solve" through various mechanisms that in the end tip the balance towards cooperation (14). In fact, Greene defines morality as such within these terms, namely as "a set of psychological adaptations that allow otherwise selfish individuals to reap the benefits of cooperation" (23). Morality has evolved, in Greene's account, in order to provide a solution to the problem of cooperation—that is, as a way of 'averting' the ToC. This means that humans beings were "biologically speaking...designed for cooperation," but the problem, so argues Greene, is that this design encompasses and extends only to members of one's own tribe (23).

While morality thus evolved for intragroup cooperation, providing a way out of the bind posed by the ToC, it did not evolve for intergroup cooperation because evolution is "an inherently competitive process," and competition between groups—and the inevitable presence of conflict where competition exists—is essential to the survival of groups in their struggle to secure finite resources for

²⁰ To avoid redundancy, this chapter will serve as a general introduction of deep pragmatism, in order to set the stage for the more critical chapter that follows. There, as the occasion demands, I will explore several strands of Greene's argument in greater depth and detail.

²¹ He calls this the Parable of the New Pastures; it is meant to illustrate how different groups or tribes came to live together on the same stretch of land, and the inter-group conflict that emerges as a result.

²² The tragedy of the commons, an influential theory in the field of economics, was originally formulated by Garret Hardin (1968).

themselves (23-24). This intertribal conflict constitutes the second tragedy, the Tragedy of Commonsense Morality (ToCM), which Greene portrays in terms of "Us versus Them," or "our interests and values versus theirs" (14). Unlike the ToC, this tragedy is not a matter of curbing selfishness; instead, Greene argues that conflict between tribes occurs and is sustained because different tribes have "incompatible visions of what a moral society should be" (4). Each particular tribe has "its own version of moral common sense," which explains much of what one sees in terms of intergroup conflict and violence (4). This tragedy of conflicting commonsense morality is therefore "the larger problem behind the moral controversies that divide us" (15) and is at the core of *Moral Tribes*. For Greene is not intent on merely voicing the tragedy as he perceives and conceives it—he also seeks a solution.

Yet how is one to go about solving the ToCM? Greene stresses that, while morality largely solves the ToC, nature has in its arsenal no expedient solution to the ToCM. Since we cannot rely on any previously evolved capacity or system in this case, what is required is what Greene refers to as an "unnatural" solution (147). The designated unnatural solution for Greene is the instatement of a metamorality; this higher-level moral system would then function to "adjudicate among competing tribal moralities ... just as a tribe's morality adjudicates among competing individuals" (147). Having identified the need for a metamorality along these lines, Greene turns to the question of what the metamorality should be.

This question is initially cast by Greene in terms of moral truth; he writes, somewhat playfully, that "[i]f we're feeling metaphysically ambitious, we may seek the moral truth:²³ universal principles that tell us how we ... ought to live, what rights and duties we truly have" (177). He proposes and discusses at some length three different (historical) approaches to moral truth, each of which has truth "imposed on us from outside" (178): the religious model of truth imposed by God, the mathematical model of truth imposed by Reason, and the scientific model of truth imposed by Nature. He ultimately rejects each of these candidates for a metamorality because none of them, to Greene's understanding, provide unequivocal access to moral truth(s)—neither religion, nor mathematics, nor science is able to offer a clear and indisputable system of moral truth that can adjudicate in cases of conflicting values.

Hence we are "thrown back on the morass of competing values" (188). What is needed according to Greene is a "more modest metamorality" that is based not on moral truth, but on shared values that he also calls a common currency (178). Greene declares to remain agnostic about whether

²³ Emphasis in the original.

there is such a thing as moral truth; in any case, he denies that there is "direct, reliable, non-question-begging access to [it]" (188). This is a crucial point, for if there were accessible moral truth, then Greene's entire project of looking for a workable metamorality would be subservient to—if not entirely bypassed by—whatever this accessible truth would be, as I will demonstrate more formally in the last chapter. Thus, one can take Greene's agnosticism as denial, especially in light of the fact that Greene has elsewhere argued against the existence of moral truth (Greene 2002).

From the idea that moral truth does not exist or, because if it does exist we cannot access it, it follows for Greene that we must "capitalize on the values we share" (189). These values do not demand to be "perfectly universal," but they must be "shared by members of different tribes whose disagreements we might hope to resolve by appeal to a common moral standard" (191). This common currency of shared values, which is one of Greene's recurring metaphors, should serve to arbitrate in cases of conflict between values across different tribes. At this point, having dismissed alternatives, Greene is ready to propose his own candidate for a metamorality: deep pragmatism.

In order to fully appreciate Greene's argument for deep pragmatism, one must keep in mind what, for him, is at stake in understanding morality in the first place. It requires two things: 1) an understanding of "the structure of modern moral problems" and how these diverge from the sorts of problems that our brains have evolved to solve, and 2) an understanding of "the structure of our moral brains" and how different types of thinking are apt to solve different problems (14). These two features of morality turn out to directly inform Greene's account of metamorality. For Greene will return to his dual-process model, which he previously developed based on the features of our moral brains, to argue for a utilitarian kind of metamorality.

Greene associates the automatic setting of the brain with deontological moral judgments and manual mode with utilitarian moral judgments. As outlined in the previous chapter, while deontological judgments in the automatic setting are highly efficient but inflexible, the contrary is true of utilitarian judgments in manual mode. This is important in light of what Greene elsewhere (2016) calls the No Cognitive Miracles Principle (NCMP):

NCMP: When we are dealing with unfamiliar²⁴ moral problems, we ought to rely less on automatic settings (automatic emotional responses) and more on manual mode (conscious, controlled reasoning), lest we bank on cognitive miracles. (131)

²⁴ In evolutionary terms.

Since we cannot have been prepared, evolutionarily speaking, for modern moral problems, it follows that what is needed is flexibility in the face of them—and utilitarian moral judgment is what allows this flexibility, at least significantly more than deontology moral judgment. Add to this the ideal of maximizing happiness, which for Greene is something that everyone understands, and what we get is deep pragmatism.

IV. Automaticity, Happiness, and Truth: Arguments against Deep Pragmatism

What I seek to do in the following sections is to (re)construct in their strongest possible form several arguments brought against Greene's account of deep pragmatism²⁵ and to offer respective counterarguments. A consequence of the fact that so little has been written about or against deep pragmatism since its introduction in *Moral Tribes* in 2013, is that Greene has had no dialogical reasons to refine and sharpen his theory. What inspires the following sections, then, is the idea that Greene's account is worth taking seriously and that it deserves detailed attention, despite the fact that there are, as will be seen, significant weaknesses in his arguments for deep pragmatism.

I will focus on three lines of objection against Greene's reasoning for deep pragmatism as favored metamorality. In the first section (IV.1.), I will examine what I call the Automatic Mode Objection (AMO), which is based on Tobia (2015). In the second section (IV.2.) I will focus on another line of criticisms, which I refer to as Happiness Objections (HOs). There are two forms of HO. The first, derived from Conning (2015), questions why tribes should surrender established systems of meaning for the sake of happiness as defined by Greene's account of deep pragmatism. The second type of HO is based on Rosenqvist (2017) and casts doubt on the pragmatic goal of happiness maximization, which is not as clear-cut as Greene makes it seem. Disagreements between tribes over what is likely to produce happiness as well as over what happiness entails are likely to ensue when happiness maximization is established as metamorality. In the third section (IV.3.), I will develop the final line of objections that I call Moral Truth Objections (MTOs). It, too, has two forms. Rosenqvist (2017) voices the first kind when he asks why we ought to adopt a metamorality in order to solve the TCM in the first place when Greene does not commit himself (i.e., remains agnostic) about moral truth. If the TCM is not morally bad, then why do we need to solve it? The second type of MTO is based on Wielenberg's (2014) exposition of incoherent commitments to moral truth in Greene's argument for deep pragmatism.

²⁵ To date (September 1, 2017), as I stated in the Introduction, there have been no article-length discussions of Greene's *Moral Tribes* or his theory of deep pragmatism. There have been several book reviews which have contained relevant points of criticism. Given the brief and limited nature of these publications, however, the critical points are undeveloped and underexplored; furthermore, the issues that are touched on in the reviews have not been taken up (to the best of my knowledge) either by the respective reviewers or by others, evidence of which is the lack of subsequent critical scholarship on deep pragmatism. One of my aims is to remedy this situation by giving deep pragmatism sustained critical attention.

I will argue that, while all objections must be taken seriously, only the second form of MTO seriously threatens Greene's project. Greene has to drop one of his arguments concerning moral truth; but there is, in the end, a way out for him. It requires the reformulation of unreliable intuitions as undesirable. I will explore this option, and where it leaves Greene and deep pragmatism, in the final section (IV.4.).

IV.1. Automatic Mode Objection

In his review of *Moral Tribes*, Tobia (2015) offers a few critical points in response to Greene's case for metamorality. The most important criticism for present purposes begins with a questioning of the impartiality of deep pragmatism. Greene argues that "everyone's happiness counts the same" (2013, 163), which leads Tobia to wonder how inclusive his notion of 'everyone' really is. For instance, does it include animals? Are future persons taken into consideration? It is true that Greene does not address these issues in *Moral Tribes*, and it is therefore fair to point them out. The question of inclusiveness is, it must be said, one that any moral theorist faces—especially those of the utilitarian persuasion who must clarify the boundaries of their calculations. If happiness ought to be maximized impartially, it makes good sense to ask *whose* happiness counts. More will be said about this point in the next section. What matters now is what Tobia does with this concern, which is—perhaps somewhat surprisingly—to relate the issue to emotions and automatic judgment. His earlier critical remarks about happiness maximization culminate in a criticism of Greene that can be summarized as follows:

AMO: Deep pragmatism is ultimately best served not through manual mode reasoning but by means of more emotion-based automatic settings.

That is, accepting Greene's deep pragmatism as metamorality, the theory might nonetheless be turned on its head if it turns out that the sort of reasoning it champions (i.e., manual mode) might ultimately serve it *less well* than the kind of reasoning it seeks to overcome (i.e., automatic mode). This thought is at the heart of the AMO. Tobia reasons as follows:

If maximizing happiness impartially involves considering future persons, and if particular automatic processes (emotions) are good at solving cooperation problems, the best way to solve

new inter-tribal cooperation problems (in the *long* term for *lots* of future people)²⁶ may be to develop and inculcate automatic processes adept to handle these problems. (749)

Greene is, after all, concerned with solving moral problems on the 'new pastures', and the question of whether or not, and to what extent, moral concern should be extended to (potential) agents outside of one's group of kin and other members of one's tribe (like, e.g., other animals, persons very far away, and future persons) is a decidedly modern issue from an evolutionary perspective (Singer 2011). From this line of thought, Tobia concludes that it may in the end be best—that is, better than employing manual mode reasoning—to "bestow upon our descendants the best inflexible and efficient emotions" (749). Better, in other words, to instill the right emotions, to be effortlessly and suitably directed, than to have to engage in slow and effortful reasoning.

It seems that Tobia has missed the point. In order for automatic settings to develop outside of whatever natural course they might otherwise have taken over time, one needs changes in (the employment of) manual mode thinking first—which is precisely what Greene is urging and trying to achieve. Automatic responses are shaped and fine-tuned through being consistently beneficial to survival over extended periods of time. If this is Tobia's point, then Greene would not disagree with the spirit of the AMO; it would be a good thing to have the right kind of automatic settings in the first place. However, what is needed, Greene would argue, in evolutionarily unfamiliar terrain and with new moral problems is a reasoned response; relevant automatic processes cannot suddenly arise of their own accord for these problem (i.e., as stipulated by the NCMP). The latter can, however, at least potentially, be shaped by the first, when one decides *how one ought to feel* about the problem. The issue of future persons, for instance, was not a concrete problem in our evolutionary past; as such, there are no specific innate automatic responses available. In order to inculcate or bestow upon future generations, then, "the best inflexible and efficient emotions," one must first decide which emotions they ought to be. And this can only be done through some form of reasoning. In this way, Tobia's criticism is easily countered by Greene.

There is, however, another way to interpret the AMO. As Greene himself has argued, automatic settings do not exclusively have to be innate or hardwired: they may also be "acquired through cultural learning ... and through individual experiences" (2016, 121). Reexamining the AMO, one might argue that either of these two—cultural learning or individual experience—may impart the

²⁶ Emphasis in the original.

appropriate automatic settings and thereby constitute better alternatives for deep pragmatism than manual mode reasoning. Yet it is still far from clear that these forms of learning and experience can occur before the exercise of some sort of controlled reasoning. In fact, Greene has addressed this issue in response to a defense of moral intuitions by Peter Railton. Railton (2014) developed the idea that moral intuitions need not always be biased or short-sighted; they may be acquired through individual experience and a sophisticated learning process. Greene (2017) has argued in turn that acquiring good moral intuitions requires both representative data and value-aligned training, which is problematic precisely when it comes to intertribal disagreement (i.e., public moral controversy) because training processes may ultimately serve simply to reinforce extant tribal differences.

Taking Greene at face value, he is interested in how to tackle novel moral problems, and he has developed deep pragmatism towards this purpose. As such, cultural learning and individual experience still presuppose familiarity with the problematic of the pertinent moral issue; one still needs to formulate a problem—especially one as intangible as, for instance, the moral status of future persons—before one can have an automatic emotional response to it that is something other than vacuous or the reinforcement of prevailing tribal values.

In the end, then, the AMO offers nothing against Greene's account of deep pragmatism that he cannot agree with or accommodate.

IV.2. Happiness Objections

The next set of criticisms revolve around Greene's choice of happiness maximization as the ultimate goal of a metamorality. Greene uses his dual-process model to argue that utilitarianism is the "native philosophy of the human manual mode," which everyone 'gets' because all human beings "have more or less the same manual-mode machinery" (2013, 194). The first criticism to this comes from Conning (2015), who hones in on the question of why tribes should trade existing sources of meaning for the sake of happiness ideals as construed by Greene. This objection may be articulated in the following way:

HO1: If tribes derive a sense of meaning from uncooperative (in intertribal terms) beliefs and practices, then they have little, if any, incentive to exchange these for abstract happiness ideals.

The problem, for Conning, is that the deeply pragmatic solution of maximizing happiness would be recognized by communities as providing "a narrower meaning than that which is already available from their own heritage" (2015, 121). So why would they give up their tribal values?

What must be recognized is that what Conning emphasizes is precisely the ToCM as diagnosed by Greene. Recall that the ToCM ensues from different tribes having incompatible visions of a moral society, with each tribe relying on a more or less idiosyncratic version of moral common sense. This, for Greene, explains much of the conflict, division, and violence between tribes—and it is to this phenomenon that one must provide a solution, or at least offer some means of amelioration. That tribes have their own values to which they cling, for which they fight, and which they are loathe to give up, is thus an old thought for Greene; his solution of deep pragmatism to the ToCM is, in fact, formed directly in response to it.

What is at stake, then, is quite simply how convincing Greene's argument for metamorality is. There are two main claims in the argument, namely 1) that a metamorality is necessary, and 2) that deep pragmatism is the best candidate. In response to the first claim, the HO1 does nothing to show that a metamorality is not necessary, for it does not touch upon the ToCM—if anything, it reiterates the tragedy. The more promising criticism might address the second claim; it is not clear, however, how 'meaning' as such would be a superior common currency to shared values with an aim towards maximizing happiness. To the extent that a tribe's source of meaning is based on values characteristic to that tribe, it cannot be a shared value by definition. And to the extent that a tribe's source of meaning is based on values that could potentially be shared and accepted by other tribes, it can be accommodated by Greene—in fact, this scenario feeds right into Greene's purposes. The first form of happiness objections, HO1, therefore provides no real challenge to Greene's project.

Another criticism of Greene's proposal of making happiness maximization the end of metamorality is voiced by Rosenqvist (2017). Rosenqvist rightly points out that 'happiness' is not a uniform construct, nor are its determinants always clear and predictable. This second objection may be formulated as follows:

HO2: Maximizing happiness is not as unambiguous as Greene makes it seem; disagreements between tribes over what is likely to produce happiness, let alone over what happiness entails, are likely to ensue.

For instance, tribes may selfishly argue for potential consequences to their happiness in any debate within the domain of public moral controversy. How does one decide, at the level of metamorality, the consequences for happiness, the priorities in relation to happiness-affecting decisions, and so on?

That happiness calculation is not straightforward, that it brings its own set of problems, is part and parcel of criticisms against utilitarian theories more generally.²⁷ What is important for current purposes is how this issues of calculation touches on Greene's argument specifically. Greene is advocating a way of thinking about conflict between tribes at a meta-level; as such, questions about the nature and distribution of happiness can be addressed after the fact, once deep pragmatism and its happiness maximization ideal have been accepted as a common currency of shared values. The crucial point is that deep pragmatism provides a way of speaking, a second-order discourse, when disagreement occurs. Greene's diagnosis is that disagreements are rife between tribes—but in his view, this is due to first-order differences between conflicting values to which different tribes ascribe. I think that Greene would happily accept as an improvement of the situation that disagreements move from between-values to second-order questions concerning happiness and happiness-distribution.

IV.3. Moral Truth Objections

As in the case of happiness objections, two distinct kinds of criticism have been leveled against Greene on account of his position on moral truth. The first is based on Rosenqvist (2017), who attempts to undermine Greene's goal of developing a metamorality by calling into question its very need. He does so by way of Greene's position on the absence of moral truth. The objection may be expressed as follows:

MTO1: If there is no moral truth, and if, therefore, the ToCM does not constitute a morally bad state of affairs, then there is no need for a metamorality in the first place.

In other words, that something ought to be done to resolve the ToCM hinges, for Rosenqvist and the MTO1, on whether or not that state of affairs is truly morally bad. If there is no moral truth, then the ToCM is also not a moral(ly bad) problem; this then subverts the need for a metamorality, because Greene proposes deep pragmatism directly in response to the ToCM.

²⁷ See, for instance, Williams (2012) for a related critique of utilitarianism.

Rosenqvist himself provides a clue as to what Greene's response should be when he suggests that "[i]f utilitarianism is the true moral theory, then we should perhaps avoid [the ToCM] to produce more happiness" (2017, 227). Greene has defined morality as "a set of psychological adaptations that allow otherwise selfish individuals to reap the benefits of cooperation" (2013, 23). This definition may, of course, be challenged; but Rosenqvist does not do so. Therefore, if one accepts Greene's description of morality, cast in terms of the value of cooperation, then what is 'bad' about the ToCM are its consequences in terms of noncooperation or conflict, which need not have a moral truth component. One may value cooperation without committing oneself metaphysically to a stance on moral truth—as in, for example, Bowles and Gintis's (2011) account. That it is good to cooperate among tribes or that it is bad to have intertribal conflicts need not comprise moral as such, but can be otherwise justified. One might do so in terms of, for instance, the development of the human species; one might argue that cooperation ultimately facilitates technological progress, while conflict causes developmental stagnation. That moral tribes are so deeply entrenched in the language of their own values that they cannot, and therefore do not, communicate with other tribes is a bad state of affairs not in light of moral truth, but because it impedes the settling—or even the very conditions for potentially settling—public moral controversies. Therefore, MTO1 does not undermine Greene's argument for metamorality, because the undesirability of the ToCM may rather straightforwardly be based on criteria other than moral truth.

The second objection to Greene's account concerning moral truth is found in Wielenberg (2014), who perceives a tension between two lines of argument in Greene's case for deep pragmatism.²⁸ This objection requires the most attention, as it is the subtlest and potentially the most far-ranging. It may be stated as follows:

MTO2: Greene's argument for deep pragmatism is incoherent because it appeals both to the nonexistence or unknowability of moral truth, as well as to the failure of certain kinds of moral judgments to track moral truth.

Moral Tribes is a hodgepodge of different claims; for the sake of clarity, I will reconstruct five arguments that make up Greene's overall case for deep pragmatism, and I will specify the individual claims that

²⁸ A tension, more specifically, between what I will call in what follows the Argument against Antiutilitarian Intuitions and the Argument for Shared Values.

go into each of these arguments so that the incoherence regarding moral truth will become clear.²⁹ First, there is the Argument for the Necessity of Metamorality (ANM), with which Greene seeks to show why we need a metamorality at all. It is stated as follows:

ANM:

1. Morality is a solution to intratribal problems—it solves the ToC. (26)³⁰
2. Under modern conditions, the solution to the ToC, namely morality, leads to intertribal conflict—it creates the ToCM. (26)
3. It does so because different tribes derive conflicting values from morality. (26)
4. There is no moral truth to ground any one tribe's system of values. (188)
5. Therefore, what is needed is moral common ground to resolve the ToCM—i.e., a metamorality.

Having established the need for a metamorality, Greene turns to the question of what this metamorality should entail. A second argument, the Argument against Antiutilitarian Intuitions (AAI), serves to show that metamorality should not be based on antiutilitarian (e.g., deontological) intuitions. It is specified as follows:

AAI:

6. Antiutilitarian intuitions are sometimes oversensitive—i.e., tracking morally irrelevant factors. (212)
7. Antiutilitarian intuitions are sometimes undersensitive—i.e., failing to track morally relevant factors. (212)
8. Therefore antiutilitarian intuitions are unreliable. (212)
9. Therefore metamorality should not be based on antiutilitarian intuitions.

Having demonstrated that metamorality ought not to be based on antiutilitarian intuitions, Greene turns to a third argument for what metamorality should be based upon, namely the Argument for Shared Values (ASV). It is formulated as follows:

²⁹ The separation of the following five arguments, their names, and the combination of claims that comprise them are my own (re)formulations. This was done for reasons of precision and subsequent analysis; if not all formulations are found in *Moral Tribes* as they are here, then I am nonetheless certain that nothing in that work speaks against them.

³⁰ All page numbers in parentheses refer to *Moral Tribes* (Greene, 2013) and indicate the location of each particular claim.

ASV:

10. There is no moral truth—or, if there is, we have no access to it. (188)
11. Therefore metamorality should not be based on moral truth. (188-189)
12. The goal of a metamorality is to establish moral common ground. (189)
13. Therefore, metamorality should be based on shared values. (189)

Having established that metamorality should be based on shared values, Greene specifies this further in utilitarian terms using an Argument for Utilitarianism as Shared Value (USV). It is expressed as follows:

USV:

14. If all else is equal, everyone prefers more to less happiness. (193)
15. If all else is equal, we care about intraindividual as well as interindividual levels of happiness. (193)
16. If all else is equal, therefore, we prefer to increase the total amount of happiness across people. (193)
17. Dropping the 'all else equal' qualifier leaves us with utilitarianism. (194)
18. This gives us the best, most comprehensive, metamorality, which can in principle resolve any disagreement. (194)

Greene has an additional argument for the 'sharedness' of utilitarianism, which is based on his dual-process theory and which is worth mentioning here:

19. Manual mode thinking is predisposed to utilitarian thinking. (198)
20. We can all share in this because we all have the same manual mode machinery.³¹ (194)
21. Therefore, manual mode/utilitarian thinking is a good basis for shared values.

³¹ One might object that, if we all share brain anatomy, then aside from manual mode machinery we must also all share automatic setting machinery. One therefore cannot favor one over the other on this basis. This is a good point. It is one of those cases where Greene needs additional arguments and cannot rely solely on neuroscience, for it does not help him make the case he wants to make. More can be said about this, which I cannot do here. What is important for present purposes is that, in order for his theory to work, Greene needs to link utilitarian thought with a basic human capacity; for if it should turn out that not everyone can at least in principle engage in this sort of reasoning, then it follows that it cannot be the basis of a shared value system. That a shared value system is necessary, and that it ought to be utilitarian, are of course separate arguments.

Finally, to bring it all together, there is the argument for utilitarianism as metamorality, which may also simply be called Deep Pragmatism (DP):

DP:

22. We need a metamorality (*ANM*).
23. It should not be based on antiutilitarian intuitions (*AAI*).
24. It should be based on shared values (*ASV*).
25. Utilitarianism is the best candidate for shared values (*USV*).
26. Therefore, our metamorality should be utilitarian.

Having examined the different arguments that go into Greene's case for deep pragmatism and the individual claims that enter each, the critical question is what moral truth is doing in Greene's various arguments and if what it is doing *it is doing consistently*. On the one hand, as should be clear, Greene argues that there is no (access to) moral truth; he uses this idea as part of the ANM (claim 4) and the ASV (claim 10), which is unproblematic in itself. Where he runs into trouble, however, is when he argues on the other hand that antiutilitarian intuitions sometimes fail to track the moral truth, which is a notion that forms the AAI (claims 6-8). These claims about the unreliability of antiutilitarian intuitions (6-8) are in direct opposition to claims 4 of the ANM and 10 of the ASV. For it follows that, if there is no such thing as moral truth, then antiutilitarian intuitions also cannot fail to track it. If there are no features of moral dilemmas that stand in a special relation to moral truth such that their truth-value can be tracked—because any such relation is necessarily precluded by there being no (access to) moral truth—then the claim that antiutilitarian intuitions fail to track morally relevant features is meaningless. There is quite simply nothing true to be tracked—reliably or otherwise. In metaethical terms, after having renounced moral truth, Greene nonetheless endorses a cognitivist position about moral judgments (i.e., that they are truth-apt) in his argument against antiutilitarian intuitions. He thus sneaks in an assumption concerning moral truth in his AAI that he explicitly denies in both his ASV and the ANM.

Accordingly, either the moral dilemmas that Greene considers possess features that afford themselves to be tracked for truth, so that one can fail or succeed in tracking them (as in claims 6-8), or no such features exist—as in claims 4 and 10—so that unreliability becomes inapt. One of Greene's arguments, therefore, contradicts two others: the ANM and the ASV are incompatible with the AAI,

as they are based on competing claims about the existence of moral truth. As a result, Greene's argument for deep pragmatism (DP) fails to hold, because claims 23 and 24, arising as they do from irreconcilable claims, cannot both be accepted within the same argument.

IV.4 What Is To Be Done?

Where does this leave Greene? If he wants to avail himself of the argument against the unreliability of antiutilitarian intuitions (AAI) to support deep pragmatism, then he must abandon claims 4 and 10 that rely on the nonexistence or inaccessibility of moral truth. Claim 4, however, is essential to Greene's rationale of forming a metamorality in the first place (i.e., to the ANM); he builds on the absence of moral truth to make his case for the need of a metamorality. Claim 10 is also crucial to Greene's argument, for he uses it to show (in the ASV) that shared values or a common currency are to be the foundation of metamorality. If there turns out to be accessible moral truth, then Greene's deep pragmatism will be left vulnerable, for it is unlikely that—in the face of knowable moral truth—anyone, no matter to which tribe one belongs, would forego this moral truth for a loose (i.e., non-truth-based) set of shared values. Granted that the discovery of both knowable and indisputable moral truth is highly unlikely (it has been a long time trying), Greene's deep pragmatism still departs precisely from the abandonment of such efforts, so that doing without the denial of knowledge of moral truth would be most injurious to his project. More technically, without moral truth, and therefore without claims 4 and 10 and the ANM and ASV that reply upon them in turn, Greene is left without claims 22 and 24 of his argument for deep pragmatism; he loses, respectively, both the necessity for metamorality and the argument that metamorality be based on shared values.

To illustrate this point, consider a hypothetical analogy with science. Let us say that we have certain epistemic capacities that served us well in the environment of evolutionary adaptedness and that allow us to solve the basic scientific problems of everyday life (analogous to morality as a natural solution to the ToC). These epistemic dispositions, however, were not designed for novel and complex modern scientific problems (like quantum mechanics, the origin of the universe and what it is made of, how life began, and so on). Different groups have come to harness and value different epistemic dispositions to tackle these problems, leading to various tribal sciences. There is intertribal disagreement about and conflict over these central modern problems, with different tribes favoring their own, idiosyncratic scientific explanations (analogous to the ToCM). Science is an exceedingly useful tool and imperative to the development of the human species as a whole. Intertribal conflicts

over science are thwarting scientific progress. Consequently, we need a metascience in order to settle disagreements over these new scientific problems. The question then becomes: On what should our metascience be based? And the answer: metascience should be based on shared values (analogous to the ASV). We all share an interest in being efficacious in the world, in realizing practical goals, and in making instrumental scientific advances. Therefore, metascience should be based on a pragmatic principle; that is, it should be founded on what works best for us. Deep pragmatism is, accordingly, the metascience that we need to adopt. The problem with this analogy is that there is a premise missing. This is the claim that *there is no scientific truth* or, if there is, that we have no epistemic access to it. The argument for a truth-independent metascience falls apart with the availability of scientific truth. If there *is* scientific truth to be discovered—that is, under conditions of scientific realism—the corresponding epistemic stance would, for instance, regard "mature and predictively successful scientific theories as well-confirmed and approximately true of the world," so that "the entities posited by them, or, at any rate, entities very similar to those posited, do inhabit the world" (Psillos 1999, xvii). There is no reason why, under these conditions, where scientific truth can be and is discovered, one would need to go beyond the first-order activity of science by means of a second-order metascience. That the origin of the universe has not been revealed does not mean it cannot be or will not be; whether or not it actually will depends on the course that science takes. Unless, of course, there is no scientific truth to be known.³² Bringing things back to Greene, I maintain that the metascience analogy holds for metamorality, so that the premise of the nonexistence or undiscoverability of truth is necessary to both—whether in the case of science or morality. With the possibility of finding out the truth, and of subsequently building a truth-based theory upon its foundation, there is no good reason to substitute this pursuit for a pragmatic meta-approach. The argument fails to get off the ground.

Greene therefore cannot do without the disavowal of moral truth; to bring it back to his series of arguments, he cannot abandon claims 4 of the ANM and 10 of the ASV. The alternative is to forego claims 6 to 8 of the AAI, which appeal to the imperfect tracking of moral truth for antiutilitarian intuitions. This option is, on the whole, less harmful to Greene's case for deep pragmatism than resigning the other claims. Nevertheless, this rips out a significant portion of his argument, because Greene wants to use the superiority of specifically utilitarian moral judgments to deontological moral judgments as a reason to accept a metamorality based on the former. More precisely, if Greene

³² Or, unless one does not care about the truth. This is admittedly another way out, although I doubt that anyone would favor the argument that there is truth to be known but that this does not matter for the conduct of science, human behavior and cognition, epistemology, and so on. In any case, one would need a strong argument for the devaluation of knowable truth and the principle by which it ought to be supplanted.

abandons claims 6-8, he loses his argument against antiutilitarian intuitions (AAI) upon which metamorality ostensibly ought not to be based, which also means that he has to renounce claim 23 in his argument for deep pragmatism (DP). This weakens Greene's final argument by eliminating one of its claims, and opens it up to counterarguments for metamorality to be based on antiutilitarian (e.g., deontological) intuitions, which become candidates anew as soon as Greene proves unable to discard them via appeals to unreliability.

There is a way out. Greene is wrong to hinge claims 6-8 on the unreliability of antiutilitarian intuitions, because his stance on the absence of moral truth is necessitated by his goal of devising a metamorality. What he ought to do instead is to reevaluate and reformulate antiutilitarian intuitions as *undesirable* rather than unreliable, so that he can argue from there that antiutilitarian intuitions should not form the basis of metamorality.

To see how he could do this, it is worth looking into some of the examples that Greene offers for the supposed oversensitivity/undersensitivity of moral intuitions in order to gauge how they might be alternatively explained. First, Greene argues that our automatic settings, our moral intuitions that, according to the CTP, feed into characteristically deontological judgments, can be oversensitive. Evidence of this is that they sometimes respond to "things that, upon reflection, don't seem to be morally relevant" (2013, 212). As example, he cites studies that have shown that the judgments of juries are sensitive to characteristics of the defendant such as race (Baldus et al. 1998; Eberhardt et al. 2006), which "we (participants in this conversation) today regard as morally irrelevant" (2013, 212). Greene has also argued that moral judgments concerning the permissibility of different actions in trolley problems are oversensitive to personal force, so that the directness or 'personalness' of the force applied by an agent appears to be a significant factor in how morally wrong people consider the same action to be (Greene et al. 2009; Greene 2006). Second, Greene argues that automatic settings can sometimes be oversensitive, in that they may "fail to respond to things that, upon reflection, do seem to be morally relevant" (2013, 212). Greene sticks with the judicial court for his examples, and offers as a case of undersensitivity the sometimes inadequate accounting for a defendant's age by juries.

While all these examples of oversensitivity/undersensitivity are framed by Greene in terms of unreliability when it comes to moral relevance, he could simply reframe them in terms of undesirability in order to move away from the issue of moral truth. One way to do this might be to take the theory that he values (i.e., utilitarianism) and argue against relying on automatic settings on consequentialist grounds. That is, he could use a version of his USV argument to show that these cases of unreliability are in fact undesirable in that they fail to effectively maximize happiness. However, since Greene seeks

to use both the USV and the unreliability/undesirability of automatic settings as part of his argument for deep pragmatism in order to justify his overall theory (DP), his argument hereby becomes circular. To put it in different terms, to use manual mode because reliance on automatic mode produces consequences that in manual mode appear undesirable is to value manual mode from the start, thereby begging the question. Appealing to utilitarian or manual mode thinking therefore fails to provide a way out.

It must be recalled that Greene defines morality as a set of psychological adaptations that permit selfish individuals to cooperate, which demonstrates the importance Greene attaches to cooperation. In fact, *Moral Tribes* as a whole can be read as an extended appeal to human (i.e., intertribal) cooperation. With cooperation already prevailing as an ideal for Greene, and with that notion being one that underlies his very notion of morality, it would be a small and quite natural move for him to recast unreliability as undesirability in terms of cooperation. To illustrate, take the example of juries relying on automatic intuitions about race or age. Rather than arguing for a mismatch between judgments and moral truth, Greene could argue instead that a jury that sentences based on racial attributes or with a disregard for age is one that undermines an optimally functioning, cooperative society. Singling out individuals and extending harsher punishments to them solely due to race is precisely the kind of situation in which different tribes (in this case, members of different races) are likely to clash. Greene could appeal here, for instance, to principles of fairness; in any case, the audience which he addresses in *Moral Tribes* is never going to accept a jury that sentences differentially based on physical appearance. Rather than using unreliability as a principle, and in this way contradicting his stance on moral truth, Greene could employ more *ad hoc* reasoning—for any given outcome related to reliance on automatic settings—to show how such an outcome is ultimately undesirable for cooperative purposes. In the case of personal force, for example, rather than dabble with the issue of moral truth, Greene could argue that a society that in otherwise similar cases condemns up close and personal killing but does not condemn impersonal killing from afar (say, via a drone) is likely to invite conflict in society at large. The minimization of conflict on the one hand, and the fostering of cooperation on the other, is at the same time precisely the kind of deeply pragmatic approach that Greene has in mind with his metamorality. The ToCM, after all, is a tragedy of group conflict, nondialogue, and noncooperation.

Let us forget about moral truth, Greene urges through his arguments for metamorality (ANM) and for shared values (ASV), and let us see what we can work with—that is, let us mobilize and implement a common currency. Is not the eradication of discrimination based on race or age precisely the kind of

shared value that we can feasibly employ towards the ideal of intergroup cooperation? Furthermore, consider once more the analogy with metascience. Let us assume, unlike before, that there is no scientific truth to be known. In that case, it makes good sense to have a pragmatic principle to ensure that different scientific tribes do not get bogged down in their own theories, so that, for the sake of cooperation and scientific progress for the human race altogether, a metascience is introduced to settle scientific debates. Happiness maximization, it must be noted, is a more distal natural goal here than is cooperation.

With cooperation used as a benchmark to explain the undesirability of reliance on automatic settings, Greene no longer needs to rely on unreliability when it comes to moral truth. In this way, the MTO2 is deflated. That reliance on automatic settings be undesirable for Greene is enough to sidestep this particular objection. To be consistent, however, Greene would also need to adopt cooperation as explanation for why utilitarian moral judgments, as a class, are more desirable than antiutilitarian intuitions. That is, he needs a single principle by which to evaluate intuitions or moral judgments *as such* if he wants to make meaningful comparisons between specific subclasses of intuitions and use this as a line of argument towards his metamorality. With morality already defined in terms of cooperation by Greene (2013, 14), and with the undesirable or desirable effects of antiutilitarian and utilitarian intuitions each explained in cooperative terms, one is left wondering, however, why Greene needs utilitarian happiness maximization at all. In the absence of moral truth, and without appeals to the unreliability of antiutilitarian intuitions, Greene might as well adopt cooperation as metamorality. In the end, if anything, tribes are probably more like to cooperate with enemy tribes than to account for their happiness. The end of cooperation—the achievement of shared goals—is a common currency in itself, by definition. It might be justified by way of happiness, justice, progress, or some other value. You might even say that happiness, like the realization of any other value on a societal scale, requires cooperation from the outset. If no one is willing to cooperate in the maximization of intertribal happiness, then where does this leave us? Greene might, when all is said and done, be appealing more to cooperation than to anything else.

He could even keep the name: deep pragmatism. It may be espoused by anyone seeking to move beyond the limitations of blind adherence to their tribal values. Because, just as we all 'get' utilitarianism, we all 'get' cooperation. Unless Greene can demonstrate that there is some robust element of our cognitive machinery that limits our understanding of cooperation solely to intratribal cooperation, manual mode thinking should work for cooperative gains just as well as it does for happiness gains. Bowles and Gintis (2011) epitomize us, after all, as a cooperative species. With reason.

V. Conclusion

I have shown that Greene's research program draws from and can be understood against the backdrop of a recent empirical turn in the study of morality, which has seen important developments in the fields of experimental philosophy, metaethics, moral psychology, and moral neuroscience. I have argued that neuroscience has a role to play in normative ethics, albeit a more circumscribed one than early optimism appeared to warrant. The metamorality that Greene has attempted to build out of his research is impressive and worthy of serious attention and critique. I have reconstructed in their strongest possible forms three central strands of criticism against Greene's deep pragmatism, based on the few available critical reviews of *Moral Tribes*. I have shown that, while all objections must be taken seriously, only the second form an objection to Greene's use of moral truth (i.e., the MTO2) seriously threatens Greene's project. It does so because, as I have demonstrated, Greene advocates conflicting views on the existence of moral truth; he denies it to argue that we need a metamorality (ANM) and that it ought to be based on shared values (ASV), while he affirms it in his attempt to discredit antiutilitarian intuitions (AAI). Greene needs to commit to one stance or the other and accept the consequences this has for his theory of deep pragmatism. I have argued, using an analogy with (meta)science among other things, that he must necessarily deny the knowability of moral truth; otherwise, his argument for a metamorality is redundant. I have pointed towards a way out of this bind for Greene, which lies in the recalibrating of antiutilitarian intuitions from unreliable to undesirable, which may be done with an appeal to cooperation, which is the most natural standard, by Greene's own definition of morality and which becomes clear through the metascience analogy. Were Greene to take this direction, however, as I think he should, his ideal of happiness maximization is left vulnerable. Cooperation is more parsimonious and, in the end, appears more fundamental to solving intertribal conflict.

Bibliography

- Alexander, Joshua. *Experimental Philosophy: An Introduction*. Cambridge: Polity Press, 2012.
- Alexander, Joshua, and Jonathan Weinberg. "Analytic Epistemology and Experimental Philosophy." *Philosophy Compass* 2, no. 1 (2007): 56-80.
- Baldus, David C., George Woodworth, David Zuckerman, and Neil Alan Weiner. "Racial Discrimination and the Death Penalty in the Post-Furman Era: An Empirical and Legal Overview with Recent Findings from Philadelphia." *Cornell Law Review* 83, no. 6 (1998): 1638-1821.
- Bartels, Daniel M., and David A. Pizarro. "The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas." *Cognition* 121, no. 1 (2011): 154-161.
- Bealer, George. "On the Possibility of Philosophical Knowledge." *Philosophical Perspectives* 10 (1996): 1-34.
- Bealer, George. "Intuition and the Autonomy of Philosophy." In *Rethinking Intuition*. Edited by Michael DePaul and William Ramsey. Maryland: Rowman & Littlefield Publishers, 1998. 201-239.
- Bechara, Antoine, Antonio R. Damasio, Hanna Damasio, and Steven W. Anderson. "Insensitivity to future consequences following damage to human prefrontal cortex." *Cognition* 50 (1994): 7-15.
- Bechara, Antoine, Daniel Tranel, Hanna Damasio, and Antonio R. Damasio. "Failure to Respond Autonomically to Anticipated Future Outcomes Following Damage to Prefrontal Cortex." *Cerebral Cortex* 6, no. 2 (1996): 215-225.
- Berker, Selim. "The Normative Insignificance of Neuroscience." *Philosophy & Public Affairs* 37, no. 4 (Fall 2009): 293-329.
- Bowles, Samuel, and Herbert Gintis. *A Cooperative Species: Human Reciprocity and Its Evolution*. Princeton: Princeton University Press, 2011.
- Brand, Cordula, ed. *Dual-Process Theories in Moral Psychology: Interdisciplinary Approaches to Theoretical, Empirical, and Practical Considerations*. Wiesbaden: Springer VS, 2016.
- Buchtel, Emma E., and Ara Norenzayan. "Which should you use, intuition or logic? Cultural differences in injunctive norms about reasoning." *Asian Journal of Social Psychology* 11, no. 4 (2008): 264-273.
- Butchvarov, Panayot. "Ethics Dehumanized." In *Metaethics after Moore*. Edited by Terry Horgan and

- Mark Timmons. Oxford: Oxford University Press, 2006. 367-389.
- Buckwater, Wesley. "Knowledge Isn't Closed on Saturday: A Study in Ordinary Language." *Review of Philosophy and Psychology* 1, no. 3 (2010): 395-406.
- Buckwater, Wesley, and Stephen Stich. "Gender and Philosophical Intuition." In *Experimental Philosophy: Volume 2*. Edited by Joshua Knobe and Shaun Nichols. Oxford: Oxford University Press, 2014. 307-346.
- Cappelen, Herman. *Philosophy without Intuitions*. Oxford: Oxford University Press, 2012.
- Cappelen, Herman. "X-Phi without Intuitions?" In *Intuition*. Edited by Anthony Booth and Darrell Rowbottom. Oxford: Oxford University Press, 2014. 269-286.
- Chaiken, Shelly, and Yaacov Trope, eds. *Dual-process Theories in Social Psychology*. New York: Guilford Press, 1999.
- Conning, Andrew Scott. "Moral Tribes: Emotion, Reason, and the Gap between Us and Them." *Journal of Moral Education* 44, no. 1 (2015): 119-121.
- Curry, Oliver. "Who's Afraid of the Naturalistic Fallacy?" *Evolutionary Psychology* 4 (2006): 234-247.
- Deutsch, Max. *The Myth of the Intuitive: Experimental Philosophy and Philosophical Method*. Cambridge: The MIT Press, 2015.
- Damasio, Antonio. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Vintage, 2006.
- Darwall, Stephen. "Getting Moral Wrongness in the Picture." In *Moral Brains: The Neuroscience of Morality*. Edited by Matthew S. Liao. Oxford: Oxford University Press, 2016. 159-169.
- Doris, John. "Persons, situations, and virtue ethics." *Noûs* 32 (1998): 504-530.
- Doris, John. *Lack of character: Personality and moral behavior*. Cambridge: Cambridge University Press, 2002.
- Doris, John M. "Introduction." In *The Moral Psychology Handbook*. Oxford: Oxford University Press, 2012. 1-3.
- Driver, Julia. "The Limits of the Dual-Process View." In *Moral Brains: The Neuroscience of Morality*. Edited by Matthew S. Liao. Oxford: Oxford University Press, 2016. 150-158.
- Eberhardt, J. L., P. G. Davies, V. J. Purdie-Vaughns, and S. L. Johnson. "Looking deathworthy: perceived stereotypicality of Black defendants predicts capital-sentencing outcomes." *Psychological Science* 12, no. 5 (2006): 383-386.
- Edmonds, David. *Would You Kill the Fat Man? The Trolley Problem and What Your Answer Tells Us about Right and Wrong*. Princeton: Princeton University Press, 2014.
- Evans, Jonathan St. B. T., and Keithe E. Stanovich. "Dual-Process Theories of Higher Cognition:

- Advancing the Debate." *Perspectives on Psychological Science* 8, no. 3 (2013): 223-241.
- Feltz, Adam, and Chris Zarpentine. "Do You Know More When It Matters Less?" *Philosophical Psychology* 23, no. 5 (2010): 683-706.
- Fisher, Andrew. *Metaethics: An Introduction*. Durham: Acumen, 2011.
- Foot, Philippa. "The Problem of Abortion and the Doctrine of Double Effect." *Oxford Review* 5 (1967): 5-15.
- Frankish, Keith. "Dual-Process and Dual-System Theories of Reasoning." *Philosophy Compass* 5, no. 10 (2010): 914-926.
- Gigerenzer, Gerd. "Personal Reflections on Theory and Psychology." *Theory & Psychology* 20, no. 6 (2010): 733-743.
- Gill, Michael B. "Moral Rationalism vs. Moral Sentimentalism: Is Morality More Like Math or Beauty?" *Philosophical Compass* 2, no. 1 (2007): 16-30.
- Greene, Joshua D. *The terrible, horrible, no good, very bad truth about morality and what to do about it*. Doctoral thesis. Department of Philosophy: Princeton University, 2002.
- Greene, Joshua. "From neural 'is' to moral 'ought': what are the moral implications of neuroscientific moral psychology?" *Nature Reviews Neuroscience*, 4, no. 10 (October 2003): 846-849.
- Greene, Joshua. "Emotion and Cognition in Moral Judgment: Evidence from Neuroimaging." In *Neurobiology of Human Values*. Edited by J.-P. Changeux, A.R. Damasio, W. Singer, and Y. Cristen. Berlin: Springer-Verlag, 2005a. 57-66.
- Greene, Joshua. "Cognitive Neuroscience and the Structure of the Moral Mind." In *The Innate Mind: Structure and Contents*. Edited by Peter Carruthers, Stephen Laurence, and Stephen P. Stich. Oxford: Oxford University Press, 2005b. 338-352.
- Greene, Joshua D. "Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains." *TRENDS in Cognitive Science* 11, no. 8 (2007): 322-323.
- Greene, Joshua D. "The Secret Joke of Kant's Soul." In *Moral Psychology, Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*. Edited by Walter Sinnott-Armstrong. Cambridge: The MIT Press, 2008. 35-80.
- Greene, Joshua. "Notes on 'The Normative Insignificance of Neuroscience' by Selim Berker." Published online on November 2010. Available at:
<https://static1.squarespace.com/static/54763f79e4b0c4e55ffb000c/t/54cb945ae4b001aedee69e81/1422627930781/notes-on-berker.pdf>
- Greene, Joshua D. *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. New York: Penguin

Books, 2013.

- Greene, Joshua D. "Beyond Point-and-Shoot Morality: Why Cognitive Neuro(Science) Matters for Ethics. In *Moral Brains: The Neuroscience of Morality*. Edited by Matthew S. Liao. Oxford: Oxford University Press, 2016a. 119-149.
- Greene, Joshua D. "Reply to Driver and Darwall." In *Moral Brains: The Neuroscience of Morality*. Edited by Matthew S. Liao. Oxford: Oxford University Press, 2016b. 170-181.
- Greene, Joshua D. "Solving the Trolley Problem." In *A Companion to Experimental Philosophy*. Edited by Justin Sytsma and Wesley Buckwalter. Oxford: John Wiley & Sons, Ltd., 2016c.
- Greene, Joshua D. "The rat-a-gorical imperative: Moral intuition and the limits of affective Learning." *Cognition* 167 (October 2017): 66-77.
- Greene, Joshua, and Jonathan Cohen. "For the law, neuroscience changes nothing and everything." *Philosophical Transactions of the Royal Society B: Biological Sciences* 359 (2004): 1775-1785.
- Greene, Joshua, and Jonathan Haidt. "How (and where) does moral judgment work?" *TRENDS in Cognitive Science* 6, no. 12 (December 2002): 517-523.
- Greene, Joshua D., R. Brian Sommerville, Leigh E. Nystrom, John M. Darley, and Jonathan D. Cohen. "An fMRI Investigation of Emotional Engagement in Moral Judgment." *Science* 293 (2001): 2105-2108.
- Greene, Joshua D., Leigh E. Nystrom, Andrew E. Engell, John M. Darley, and Jonathan D. Cohen. "The Neural Bases of Cognitive Conflict and Control in Moral Judgment." *Neuron* 44 (October 2004): 389-400.
- Greene, Joshua D., Sylvia A. Morelli, Kelly Lowenberg, Leigh E. Nystrom, and Jonathan D. Cohen. "Cognitive load selectively interferes with utilitarian moral judgment." *Cognition* 107 (2008): 1144-1154.
- Greene, Joshua D., Fiery A. Cushman, Lisa E. Stewart, Kelly Lowenberg, Leigh E. Nystrom, and Jonathan D. Cohen. "Pushing moral buttons: The interaction between personal force and intention in moral judgment." *Cognition* 111 (2009): 364-371.
- Haidt, Jonathan. "The emotional dog and its rational tail: A social intuitionist approach to moral judgment." *Psychological Review* 108, no. 4 (2001): 814-834.
- Haidt, Jonathan. "Moral Psychology for the twenty-first century." *Journal of Moral Education* 42, no. 3 (2013): 281-297.
- Haidt, Jonathan, and Craig Joseph. "The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules." In *The Innate*

- Mind*, Volume 3. Edited by Peter Carruthers, Stephen Laurence, and Stephen Stich. Oxford: Oxford University Press, 2007. 367-392.
- Hardin, Garrett. "The Tragedy of the Commons." *Science* 162 (December 1968): 1243-1248).
- Harman, Gilbert. "Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error." *Proceedings of the Aristotelian Society* 99 (1999): 315-331.
- Harman, Gilbert. "The nonexistence of character traits." *Proceedings of the Aristotelian Society* 100 (2000): 223-226.
- Haukioja, Jussi, ed. *Advances in Experimental Philosophy of Language*. London: Bloomsbury Academic, 2015.
- Horgan, Terry, and Mark Timmons. "Introduction." In *Metaethics after Moore*. Edited by Terry Horgan and Mark Timmons. Oxford: Oxford University Press, 2006. 1-16.
- Isen, Alice M., and Paula F. Levin. "Effect of feeling good on helping: Cookies and kindness." *Journal of Personality and Social Psychology* 21, no. 3 (1972): 384-388.
- Joyce, Richard. *The Evolution of Morality*. Cambridge: The MIT Press, 2006.
- Kahane, Guy. "Evolutionary Debunking Arguments." *Noûs* 45, no. 1 (2011): 103-125.
- Kahane, Guy. "On the Wrong Track: Process and Content in Moral Psychology." *Mind & Language* 27, no. 5 (November 2012): 519-545.
- Kahneman, Daniel. "A Perspective on Judgment and Choice: Mapping Bounded Rationality." *American Psychologist* 58 (2003): 697-720.
- Kahneman, Daniel. *Thinking, Fast and Slow*. New York: Farrar, Straus, and Giroux, 2011.
- Kauppinen, Antti. "The Rise and Fall of Experimental Philosophy." *Philosophical Explorations* 10, no. 2 (2007): 95-118.
- Kleingeld, Pauline. "Consistent egoists and situation managers: Two problems for situationism." *Philosophical Explorations* 18, no. 3 (2015): 344-362.
- Knobe, Joshua. "Experimental Philosophy and Philosophical Significance." *Philosophical Explorations* 10, no. 2 (2007): 119-121.
- Knobe, Joshua. "Person as scientist, personal as moralist." *Behavioral and Brain Sciences* 33, no. 4 (2010): 315-329.
- Knobe, Joshua, and Jesse Prinz. "Intuitions about consciousness: Experimental studies." *Phenomenology and the Cognitive Sciences* 7, no. 1 (2008): 67-83.
- Knobe, Joshua, and Shaun Nichols, eds. *Experimental Philosophy*. Oxford: Oxford University Press, 2008.

- Liao, S. Matthew, ed. *Moral Brains*. Oxford: Oxford University Press, 2016a.
- Liao, S. Matthew. "Morality and Neuroscience: Past and Future." In *Moral Brains*. Edited by Matthew S. Matthew. Oxford: Oxford University Press, 2016b. 1-44.
- Machery, Edouard, Ron Mallon, Shaun Nichols, and Stephen P. Stich. "Semantics, cross-cultural style." *Cognition* 92, no. 3 (2004): 1-12.
- Mann, Thomas. *Death in Venice and Other Stories*. London: Vintage, 1998.
- Milgram, Stanley. *Obedience to Authority: An Experimental View*. New York: HarperCollins Publishers, 2009.
- Miller, Alexander. *An Introduction to Contemporary Metaethics*. Cambridge: Polity Press, 2003.
- Nadelhoffer, Thomas, Eddy Nahmias, and Shaun Nichols, eds. *Moral Psychology: Historical and Contemporary Readings*. Oxford: Blackwell Publishing, 2010.
- Nahmias, Eddy, Stephen Morris, Thomas Nadelhoffer, and Jason Turner. "The Phenomenology of Free Will." *Journal of Consciousness Studies* 11 (2004): 162-179.
- Nahmias, Eddy, Stephen Morris, Thomas Nadelhoffer, and Jason Turner. "Surveying Freedom: Folk Intuitions about Free Will and Moral Responsibility." *Philosophical Psychology* 18 (2005): 561-584.
- Nahmias, Eddy, Stephen Morris, Thomas Nadelhoffer, and Jason Turner. "Is Incompatibilism Intuitive?" *Philosophy and Phenomenological Research* 73 (2006): 28-53.
- Nichols, Shaun. *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford: Oxford University Press, 2004a.
- Nichols, Shaun. "Folk concepts and intuitions: from philosophy to cognitive science." *Trends in Cognitive Sciences* 8, no. 11 (2004b): 514-518.
- Nichols, Shaun. "Process Debunking and Ethics." *Ethics* 124 (July 2014): 727-749.
- Nichols, Shaun. "Sentimentalism naturalized." In *Moral Psychology, Volume 2. The cognitive science of morality: Intuition and diversity*. Cambridge: The MIT Press, 2008. 255-274.
- Nichols, Shaun, and Joshua Knobe. "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions." *Noûs* 41, no. 4 (2007): 663-685.
- Nichols, Shaun, and Michael Bruno. "Intuitions about personal identity: An empirical study." *Philosophical Psychology* 23, no. 3 (2010): 293-312.
- Prinz, Jesse. "The emotional basis of moral judgments." *Philosophical Explorations* 9, no. 1 (29-43): 2006.
- Prinz, Jesse. "Sentimentalism and the Moral Brain." In *Moral Brains: The Neuroscience of Morality*. Edited

- by Matthew S. Liao. Oxford: Oxford University Press, 2016. 45-73.
- Psillos, Stathis. *Scientific Realism: How science tracks truth*. London: Routledge, 1999.
- Railton, Peter. "The Affective Dog and Its Rational Tail: Intuition and Attunement." *Ethics* (July 2014): 813-859.
- Rosenqvist, Simon. "Moral Tribes: Emotion, Reason, and the Gap between Us and Them." *Journal of Moral Philosophy* 44, no. 2 (2017): 225-228.
- Sabini, John, and Maury Silver. "Lack of Character? Situationism Critiqued." *Ethics* 115, no. 3 (2005): 535-562.
- Schwitzgebel, Eric, and Fiery Cushman. "Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers." *Mind and Language* 27, no. 2 (2012): 135-153.
- Singer, Peter. "Famine, affluence and morality." *Philosophy and Public Affairs* 1 (1972): 229-243.
- Singer, Peter. "Ethics and Sociobiology." *Philosophy and Public Affairs* 11 (1982): 40-64.
- Singer, Peter. "Ethics and Intuitions." *The Journal of Ethics* 9, no. 3-4 (January 2005): 331-352.
- Singer, Peter. *The Expanding Circle*. Princeton: Princeton University Press, 2011.
- Sinnott-Armstrong, Walter. "Introduction." In *Moral Psychology, Volume 1: The Evolution of Morality: Adaptions and Innateness*. Edited by Walter Sinnott-Armstrong. Cambridge: The MIT Press, 2008a.
- Sinnott-Armstrong, Walter. "Introduction." In *Moral Psychology, Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*. Edited by Walter Sinnott-Armstrong. Cambridge: The MIT Press, 2008b.
- Street, Sharon. "A Darwinian Dilemma for Realist Theories of Values." *Philosophical Studies* 127 (2006): 109-166.
- Sytsma, Justin, ed. *Advances in Experimental Philosophy of Mind*. London: Bloomsbury Academic, 2014.
- Sytsma, Justin and Jonathan Livengood. *The Theory and Practice of Experimental Philosophy*. Ontario: Broadview Press, 2016.
- Thomson, Judith Jarvis. "The Trolley Problem." *Yale Law Journal* 94 (1985): 1395-1415.
- Tiberius, Valerie. *Moral Psychology: A Contemporary Introduction*. New York: Routledge, 2015.
- Tobia, Kevin Patrick. "Moral Tribes: Emotion, Reason, and the Gap between Us and Them." *Philosophical Psychology* 28, no. 5 (2015): 746-750.
- Upton, Candace L. "Virtue Ethics and Moral Psychology: The Situationism Debate." *The Journal of Ethics* 12 (2009): 103-115.

- van Roojen, Mark. *Metaethics: A Contemporary Introduction*. New York: Routledge, 2015.
- Weinberg, Jonathan M., Chad Gonnerman, Cameron Buckner, and Joshua Alexander. "Are philosophers expert intuiters?" *Philosophical Psychology* 23, no. 3 (2010).
- Wielenberg, Erik J. "On the Evolutionary Debunking of Morality." *Ethics* 120 (April 2010): 441-464.
- Wiegmann, Alex, Yasmina Okan, and Jonas Nagel. "Order effects in moral judgment." *Philosophical Philosophy* 25, no. 6 (December 2012): 813-836.
- Wielenberg, Erik J. "Review: Greene Joshua, *Moral Tribes: Emotion, Reason, and the Gap between Us and Them*." *Ethics* 124, no. 4 (July 2014): 910-916.
- Wielenberg, Erik J. "Ethics and Evolutionary Theory." *Analysis Reviews* 76, no. 4 (October 2016): 502-515.
- Williams, Bernard. "A Critique of Utilitarianism." In *Ethics: Essential Readings in Moral Theory*. Edited by George Sher. New York: Routledge, 2012. 253-261.
- Williamson, Timothy. "Philosophical 'Intuitions' and Scepticism about Judgment." *Dialectica* 58, no. 1 (2004):109-153.
- Zimbardo, Philip. *The Lucifer Effect: How Good People Turn Evil*. London: