Neuroethics (2019) 12:119–131
https://doi.org/10.1007/s12152-018-9378-3

ORIGINAL PAPER

Metamorality without Moral Truth

Steven R. Kraaijeveld · **Hanno Sauer**

Received: 26 February 2018 / Accepted: 19 June 2018 / Published online: 21 July 2018
© The Author(s) 2018

Abstract Recently, Joshua Greene has argued that we need a metamorality to solve moral problems for which evolution has not prepared us. The metamorality that he proposes is a utilitarian account that he calls deep pragmatism. Deep pragmatism is supposed to arbitrate when the values espoused by different groups clash. To date, no systematic appraisal of this argument for a metamorality exists. We reconstruct Greene's case for deep pragmatism as a metamorality and consider three lines of objection to it. We argue that, in the end, only one of these objections seriously threatens Greene's position. Greene has to commit to the nonexistence of moral truth in order for his argument for the need of a metamorality to get off the ground. This, however, leads to a tension in his overall argument for deep pragmatism: ultimately, it casts his rejection of antiutilitarian moral intuitions into doubt.

Keywords Metamorality · Moral truth · Deep pragmatism · Dual-process theory · Metaethics · Utilitarianism

Introduction

Empirical findings are increasingly brought to bear on long-standing philosophical claims and theories [1, 2]. Joshua Greene has been at the forefront of this movement, introducing empirical evidence into debates that had been reserved for traditional moral philosophy. In his earlier work, Greene examined the underlying human brain structures involved in making different kinds of moral judgments [3–5]. Some of this research informs his version of a dual-process model of moral judgment [3, 6–8], which holds that 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" [9]. One negative argument that Greene has maintained, in tandem with his dual-process model, is that deontological moral judgments are primarily driven by knee-jerk emotional responses and therefore ought to be mistrusted [10]. This line of research, and Greene's dual-process model especially, has received much critical attention in the literature [11–14]. Much less attention has been devoted, on the other hand, to the positive argument that Greene makes to complement his dual-process model, which he develops at length in his monograph *Moral Tribes* [15]. Here, he argues for the need of a metamorality that he calls deep pragmatism, which is meant to arbitrate when the values of different groups clash—analogous to the way morality serves to temper selfish impulses within tribes.

To date, there have been no article-length discussions of deep pragmatism. That Greene's argument should be sound 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

S. R. Kraaijeveld (✉)
Wageningen University & Research, Wageningen, The Netherlands
e-mail: steven.kraaijeveld@wur.nl

H. Sauer
Utrecht University, Utrecht, The Netherlands

moral problems that divide us" [15]. The ideal of constructing a metamorality to settle public moral controversy is eminently worth pursuing, and has potential ramifications that extend beyond academia. Our main objective 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.

The structure of this paper is as follows. We will first consider in detail the culmination of Greene's research program and intellectual development thus far, namely his theory of deep pragmatism as it is formulated in *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, we will show why Greene thinks we need a metamorality. We will also assess the different lines of argument that Greene employs towards his specific rendering of a metamorality, namely deep pragmatism.

Next we develop five lines of argument against Greene's reasoning behind deep pragmatism as metamorality. First, there is what we call the Automatic Mode Objection, which suggests that, rather than manual mode reasoning, automatic judgments may ultimately serve deep pragmatism better. Second, there is what we refer to as the Motivation Objection, which casts doubt on why tribes should renounce established systems of meaning for the sake of happiness as defined by Greene's deep pragmatism. Third, there is the Happiness Objection, which calls into question the pragmatic goal of happiness maximization. Fourth, there is the Redundancy Objection, which questions why we ought to adopt a metamorality when Greene does not commit himself to moral truth. Finally, there is the Incoherence Objection, which reveals a contradiction in Greene's argument for deep pragmatism based on his use of competing claims about moral truth.

We will show that, while all objections must be taken seriously, only the Incoherence Objection seriously threatens Greene's project. In fact, it puts pressure on Greene to give up one of his arguments for deep pragmatism. In the end, we argue that Greene's case for the unreliability of antiutilitarian intuitions is untenable, and we consider some ways out of this bind.

Moral Tribes and Deep Pragmatism

Greene argues 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 we will focus on the arguments from that book. Where needed, however, we will bring in additional formulations and arguments that Greene makes outside of his monograph.[1]

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 pastures[2] to illustrate and differentiate between the two tragedies. The first, the Tragedy of the Commons (ToC), occurs as in the classic case[3] when there is intratribal conflict—when there is conflict between, as Greene characterizes it, "Me versus Us," or selfishness set against concern for others [15]. 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 [p.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" [p.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 [p.23].

While morality thus evolved for intragroup cooperation, providing a way out of the bind posed by the ToC,

[1] To avoid redundancy, this part will serve as a general introduction to deep pragmatism, in order to set the stage for the critical part that follows. There we will explore several strands of Greene's argument in greater depth and detail.

[2] 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.

[3] As originally formulated by Hardin in 1968 [16].

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 themselves [pp. 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" [p.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" [p.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 [p.4]. This tragedy of conflicting commonsense morality is therefore "the larger problem behind the moral controversies that divide us" [p.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 [p.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" [p.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 in terms of moral truth; he writes, somewhat playfully, that "[i]f we're feeling metaphysically ambitious, we may seek the moral truth[4]: universal principles that tell us how we … ought to live, what rights and duties we truly have" [p.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" [p.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" [p.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 [p.178]. Greene declares to remain agnostic about whether there is such a thing as moral truth; in any case, he denies that there is "direct, reliable, non-question-begging access to [it]" [p.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 we will demonstrate more formally later. 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 [17].

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" [15]. 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" [p.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 [15]. 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 metamorality.

[4] Emphasis in the original.

Greene associates the automatic setting of the brain with deontological moral judgments and manual mode with utilitarian moral judgments. 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 calls the No Cognitive Miracles Principle (NCMP):

> NCMP: When we are dealing with unfamiliar[5] 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 [9].

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.

Five Arguments against Deep Pragmatism

What we 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 pragmatism[6] and to offer respective counterarguments.

We will focus on five lines of objection against Greene's reasoning for deep pragmatism as favored metamorality. First, we will examine the Automatic Mode Objection (AMO), which is based on Tobia [18]. Second, we will focus on another criticism, namely the Motivation Objection (MO), which is derived from Conning [19] and which questions why tribes should surrender established systems of meaning for happiness as such. Third, we will explore the Happiness Objection (HO), which is based on Rosenqvist [20] and which casts doubt on the pragmatic aspect 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. Fourth, we will develop the Redundancy Objection (RO), which is voiced by Rosenqvist [20] when he questions why we ought to adopt a metamorality in order to solve the TCM in the first place when Greene does not commit himself to moral truth. If the TCM is not morally bad, then why do we need to solve it? Finally, we discuss the Incoherence Objection (IC) that is based on Wielenberg's [21] exposition of incoherent commitments to moral truth in Greene's argument for deep pragmatism.

We will argue that, while all objections must be taken seriously, only the Incoherence Objection seriously threatens Greene's project. Greene has to drop one of his arguments concerning moral truth. We will explore where this leaves Greene and deep pragmatism in the final section.

The Automatic Mode Objection

In his review of Moral Tribes, Tobia offers a few critical points in response to Greene's case for metamorality [18]. 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" [15, p163], 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 later on. What matters right now is what Tobia does with this concern, which is—perhaps somewhat

[5] In evolutionary terms.

[6] To date, 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 nature of these publications, the critical points are undeveloped and underexplored; furthermore, the issues that are touched on in the reviews appear not to have been taken up either by the respective reviewers or by others, evidence of which is the lack of subsequent critical scholarship.

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 intertribal cooperation problems (in the *long* term for *lots* of future people)[7] may be to develop and inculcate automatic processes adept to handle these problems [18].

Greene is, after all, concerned with solving moral problems on the 'new pastures', and the question of whether, and to what extent, moral concern should be extended to (potential) agents outside of one's kin and other members of one's tribe (like, e.g., other animals, persons very far away, and future persons). This is a decidedly modern issue from an evolutionary perspective [22]. 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" [18]. 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.

However, 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. What is needed, Greene would nevertheless 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 these 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 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" [9]. Reexamining the AMO, one might argue that either of these two—cultural learning or individual experience—may impart the 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, who 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 [23]. Greene 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 [24].

Greene is ultimately 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

[7] Emphasis in the original.

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 little against Greene's account of deep pragmatism that he cannot agree with or accommodate.

The Motivation Objection

The next criticism revolves 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" [15]. The key question here is why tribes should be motivated to trade existing sources of meaning for the sake of happiness ideals as construed by Greene [19]. The objection may be articulated in the following way:

> MO: 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" [19]. 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 arises 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 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 MO 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.

However, from the perspective from particular tribes, it must be said that even if it were true that deep pragmatism provides the best metamorality, some tribes may simply reject it because they are unwilling even in theory to abandon (any one of) their values to its cause. Greene's response to this concern is worth quoting at length:

> For our purposes, shared values need not be perfectly universal. They just need to be shared widely, shared by members of different tribes whose disagreements we might hope to resolve by appeal to a common moral standard. If you're so selfish that you're not willing to lift a finger to spare another human from serious suffering, then you're simply not part of this conversation [15, 191].

Greene's proposal of deep pragmatism is for those who seek to transcend the conflictual state of clashing values. For those who are happy to stubbornly fight for their idiosyncratic values without any concern for conflict resolution, he admits that he can do little to persuade. This must count as a limitation of Greene's theory, for the deepest inter-tribal conflicts tend to be over the most entrenched and therefore the most difficultly foregone values.[8] In Greene's defense, however, no moral theory enjoys universal acceptance, let alone on its introduction. Part of Greene's goal is to persuade those who are willing at least *in principle* to find a way out of the ToCM. And tribes can always choose to accept deep pragmatism and cast the preservation of

[8] We would like to thank an anonymous reviewer for raising this point.

their particular values in terms of happiness, so as to have a stake in the debate—a debate that they have nonetheless, in this way, joined.

The MO, although it does raise some concerns, ultimately provides no real challenge to Greene's project.

The Happiness Objection

Another criticism of Greene's proposal of making happiness maximization the end of metamorality is expressed by Rosenqvist, who rightly points out that 'happiness' is not a uniform construct, nor are its determinants always clear and predictable [20]. The objection may be formulated as follows:

> HO: 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, and that it brings about its own set of problems, is part and parcel of criticisms against utilitarian theories more generally.[9] 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. Greene would probably be happy to accept as an improvement of the situation that disagreements move from between-values to second-order questions concerning happiness and happiness-distribution.

Nevertheless, the question remains whether the move in discourse from first-order conflicts between values to second-order conflicts over happiness measurements and consequences would actually constitute an improvement. The worry here is that there would remain as much conflict as before, except that this would now center on different tribal happiness considerations rather than tribal values.[10] For Greene, the advantage appears to be primarily that of discourse gains. Conflict will always arise where different tribes meet, debate, and coordinate action. If tribes can be motivated to be deeply pragmatic—that is, if they can be persuaded to take into account overall happiness considerations rather than the absolute realization of their particular cherished values—then, at least in principle, they should be open to a distribution of happiness that is not entirely unilateral. Perhaps this is an overly optimistic expectation. At the same time, if one takes seriously the 'tragedy' that is the ToCM, it may be worth taking the leap. Greene certainty thinks so.

The Redundancy Objection

Another criticism seeks to undermine the very need for a metamorality. Rosenqvist [20] argues that Greene's position on the absence of moral truth makes a metamorality redundant. The objection may be expressed as follows:

> RO: 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 for the RO, 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 subverts the need for a metamorality, because Greene proposes deep pragmatism directly in response to the ToCM.

[9] See, for instance, Williams [25] for a related critique of utilitarianism.

[10] We would like to thank an anonymous reviewer for rightfully pressing this point.

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" [20]. Greene has defined morality as "a set of psychological adaptations that allow otherwise selfish individuals to reap the benefits of cooperation" [15]. 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 to a metaphysical stance on moral truth—as in, for example, Bowles and Gintis's account [26]. That it is good to cooperate among tribes or that it is bad to have intertribal conflicts need not comprise moral truth 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. Of course, in these cases, moral assumptions may still be lurking in the background and may become visible only when the argument is pushed back.[11] To give another example, intertribal conflicts may result in physical and emotional pain, yet the undesirability of this pain does not necessarily hinge on moral truth. That having pain is a bad experience is a physiological and psychological fact for most people, which may be moralized but which does not have to be. Even if there is no moral truth, so that inflicting pain on another person is not morally wrong, one may still not want to experience pain and act to avoid it. The same goes for wellbeing. One may want to live free of intertribal conflict not because this is a morally good state of affairs, but because one values one's wellbeing *in its own right*. In fact, this is precisely the kind of shared value that Greene proposes—happiness involving, at least, the absence of pain and the presence of wellbeing.

[11] We would like to thank an anonymous reviewer for pushing us on this point.

The RO therefore does not undermine Greene's argument for metamorality, because the undesirability of the ToCM may more straightforwardly be based on criteria other than moral truth.

The Incoherence Objection

The final line of objection homes in on Greene's account of moral truth and is found in Wielenberg [21], who perceives a tension between two lines of argument in Greene's case for deep pragmatism.[12] This objection requires the most attention, as it is the subtlest and potentially the most far-ranging. It may be stated as follows:

> IO: 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, we will reconstruct five arguments that make up Greene's overall case for deep pragmatism, and we will further specify the individual claims that go into each of these arguments.[13] 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. [p.26][14]
2. Under modern conditions, the solution to the ToC, namely morality, leads to intertribal conflict—it creates the ToCM. [p.26]

[12] A tension, more specifically, between what we will call in what follows the Argument against Antiutilitarian Intuitions (AAI) and the Argument for Shared Values (ASV).

[13] The separation of the following five arguments, their names, and the combination of claims that comprise them are our 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 we are nonetheless certain that nothing in that work speaks against them.

[14] All page numbers in what follows refer to *Moral Tribes* [15].

3. It does so because different tribes derive conflicting values from morality. [p.26]
4. There is no moral truth to ground any one tribe's system of values. [p.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 stated as follows:

AAI:

6. Antiutilitarian intuitions are sometimes oversensitive—i.e., tracking morally irrelevant factors. [p.212]
7. Antiutilitarian intuitions are sometimes undersensitive—i.e., failing to track morally relevant factors. [p.212]
8. Therefore antiutilitarian intuitions are unreliable. [p.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:

ASV:

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

Having established that metamorality should be based on shared values, Greene further specifies it 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. [p.193]
15. If all else is equal, we care about intraindividual as well as interindividual levels of happiness. [p.193]
16. If all else is equal, therefore, we prefer to increase the total amount of happiness across people. [p.193]
17. Dropping the 'all else equal' qualifier leaves us with utilitarianism. [p.194]
18. This gives us the best, most comprehensive, metamorality, which can in principle resolve any disagreement. [p.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. [p.198]
20. We can all share in this because we all have the same manual mode machinery.[15] [p.194]
21. Therefore, manual mode/utilitarian thinking is a good basis for shared values.

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*).

[15] 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 could be said about this, which we 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.

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 into each, the crucial 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. The way that Greene uses moral relevance has to be understood in terms of features which, when one pays attention to them, increase or decrease the likelihood that one arrives at the moral truth. Otherwise, the conclusions that Greene draws from these observations do not make sense. Yet without moral truth, 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.

What is to Be Done?

Where does this leave Greene? If he wants to avail himself of the argument against 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 moral truth is 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 should 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 thwart 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 a premise is 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" [27, p.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 be depends on the course that science takes. Unless, of course, there is no scientific truth to be known.[16] Bringing things back to Greene, we maintain that the metascience analogy holds for metamorality, so that the premise of the nonexistence or indiscoverability 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 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 return 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 removes a significant portion of his argument, because Greene wants to use the superiority of utilitarian moral judgments to deontological moral judgments as a reason to accept a metamorality based on the former. More precisely, if Greene 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 basing metamorality on antiutilitarian (e.g., deontological) intuitions, which become candidates anew as soon as Greene proves unable to discard them via appeals to unreliability.

One possible way out is the following.[17] Rather than deny the existence of moral truth, Greene could simply assert the truth of deep pragmatism. If moral truth exists, and if deep pragmatism were the true moral theory, then this would meet several of the objections previously raised. It would solve the RO, for it would make the ToCM a morally bad state of affairs requiring a metamorality—deep pragmatism—to settle it. It would also solve the IO, for with the existence of moral truth, the failure of certain kinds of moral judgments to track moral truth would be unproblematic. It would not necessarily solve the HO, since there may be other theories superior to deep pragmatism for settling moral conflicts, but it would at least contribute a motivating factor to adopting deep pragmatism—that it is based on the truth. All the same, it is unlikely that

[16] Or, unless one does not care about the truth. This is admittedly another way out, although we 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.

[17] We would like to thank an anonymous reviewer for this suggestion.

Greene would be comfortable with this metaphysical move, since it would open up deep pragmatism to (many) competing claims for moral truth. Moreover, in terms of persuasion, it is unlikely that simply asserting the truth of deep pragmatism is going to win any favor from tribes.

Another potential way out is as follows. Greene is wrong to tie claims 6–8 to 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 [28, 29], which "we … today regard as morally irrelevant" [15]. In the same vein, Greene has 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 [9, 30]. 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" [15]. Greene sticks with the judicial court for his examples, offering 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.[18] 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 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.

Conclusion

The metamorality that Greene builds in *Moral Tribes* and that he calls deep pragmatism is worthy of systematic attention and critique, neither of which it has thus far received. We have reconstructed in their strongest possible forms five strands of criticism against deep pragmatism, based on the few available critical reviews that have appeared in the literature. We have shown that, while all objections must be taken seriously, only the objection to Greene's incoherent use of moral truth (IO) seriously threatens Greene's project. It does so because, as we 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 this ought to be based on shared values (ASV), while he affirms it to discredit antiutilitarian intuitions (AAI). Greene needs to commit to one of the two mutually exclusive stances and accept the consequences that this commitment entails for his theory of deep pragmatism. We have argued that he must necessarily deny the knowability of moral truth, because otherwise his argument for a metamorality is redundant. We have also shown that Greene cannot rely on a manual mode type of utilitarian reasoning to make a case for the

[18] Another strategy that Greene might adopt, suggested by an anonymous reviewer, is to invoke utilitarian reasoning where deontological intuitions conflict. This may be fruitful in some cases, for example when a moral problem elicits various incompatible deontological judgments. However, it is not clear how this approach might work in many of the cases that Greene discusses, for instance in classic trolley dilemmas where the tension is principally if not exclusively between deontological (rights-based) and utilitarian (happiness-maximization-based) intuitions. It is presumably in cases of this kind that we would profit most from a metamorality. What is certain is that this strategy, if endorsed, would decrease the reach of Greene's approach.

undesirability of relying on antiutilitarian automatic settings, since this begs the question. This leaves him in a position where he is free to argue for a utilitarian kind of manual mode reasoning any way he likes, just not on the basis of findings about antiutilitarian and utilitarian moral judgments. As long as he also denies the existence of moral truth.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Sinnott-Armstrong, Walter. 2008. Introduction. In *Moral psychology, volume 1: The evolution of morality: Adaptions and innateness*, ed. Walter Sinnott-Armstrong. Cambridge: The MIT Press.
2. Alexander, Joshua. 2012. *Experimental philosophy: An introduction*. Cambridge: Polity Press.
3. Greene, Joshua D., R. Brian Sommerville, Leigh E. Nystrom, John M. Darley, and Jonathan D. Cohen. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293: 2105–2108.
4. Greene, Joshua. 2005a. Emotion and Cognition in Moral Judgment: Evidence from Neuroimaging. In *Neurobiology of Human Values*, ed. J.-P. Changeux, A.R. Damasio, W. Singer, and Y. Cristen, 57–66. Berlin: Springer-Verlag.
5. Greene, Joshua. 2005b. Cognitive Neuroscience and the Structure of the Moral Mind. In *The Innate Mind: Structure and Contents*, ed. Peter Carruthers, Stephen Laurence, and Stephen P. Stitch, 338–352. Oxford: Oxford University Press.
6. Greene, Joshua D., Leigh E. Nystrom, Andrew E. Engell, John M. Darley, and Jonathan D. Cohen. 2004. The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44: 389–400.
7. Greene, Joshua D. 2007. Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Science* 11 (8): 322–323.
8. Greene, Joshua D., Sylvia A. Morelli, Kelly Lowenberg, Leigh E. Nystrom, and Jonathan D. Cohen. 2008. Cognitive load selectively interferes with utilitarian moral judgment. *Cognition* 107: 1144–1154.
9. 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, 2016. 119–149.
10. Greene, Joshua D. 2008. The Secret Joke of Kant's Soul. In *Moral Psychology, Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, ed. Walter Sinnott-Armstrong, 35–80. Cambridge: The MIT Press.
11. Berker, Selim. 2009. The normative insignificance of neuroscience. *Philosophy & Public Affairs* 37 (4): 293–329.
12. Kahane, Guy. 2012. On the wrong track: Process and content in moral psychology. *Mind & Language* 27 (5): 519–545.
13. Darwall, Stephen. 2016. Getting moral wrongness in the picture. In *Moral Brains: The Neuroscience of Morality*, ed. Matthew S. Liao, 159–169. Oxford: Oxford University Press.
14. Driver, Julia. 2016. The limits of the dual-process view. In *Moral Brains: The Neuroscience of Morality*, ed. Matthew S. Liao, 150–158. Oxford: Oxford University Press.
15. Greene, Joshua D. 2013. *Moral tribes: Emotion, reason, and the gap between us and them*. New York: Penguin Books.
16. Hardin, Garrett. 1968. The tragedy of the commons. *Science* 162: 1243–1248.
17. Greene, Joshua D. 2002. *The terrible, horrible, no good, very bad truth about morality and what to do about it*. Doctoral thesis. Department of Philosophy: Princeton university.
18. Tobia, Kevin Patrick. 2015. Moral tribes: Emotion, reason, and the gap between us and them. *Philosophical Psychology* 28 (5): 746–750.
19. Conning, Andrew Scott. 2015. Moral tribes: Emotion, reason, and the gap between us and them. *Journal of Moral Education* 44 (1): 119–121.
20. Rosenqvist, Simon. 2017. Moral tribes: Emotion, reason, and the gap between us and them. *Journal of Moral Philosophy* 44 (2): 225–228.
21. Wielenberg, Erik J. 2014. Review: Greene Joshua, moral tribes: Emotion, reason, and the gap between us and them. *Ethics* 124 (4): 910–916.
22. Singer, Peter. 2011. *The expanding circle*. Princeton: Princeton University Press.
23. Railton, Peter. 2014. The affective dog and its rational tail: Intuition and attunement. *Ethics* 124 (4): 813–859.
24. Greene, Joshua D. 2017. The rat-a-gorical imperative: Moral intuition and the limits of affective learning. *Cognition* 167: 66–77.
25. Williams, Bernard. 2012. A critique of utilitarianism. In *Ethics: Essential Readings in Moral Theory*, ed. George Sher, 253–261. New York: Routledge.
26. Bowles, Samuel, and Herbert Gintis. 2011. *A cooperative species: Human reciprocity and its evolution*. Princeton: Princeton University Press.
27. Psillos, Stathis. 1999. *Scientific realism: How science tracks truth*. London: Routledge.
28. Baldus, David C., George Woodworth, David Zuckerman, and Neil Alan Weiner. 1998. 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 (6): 1638–1821.
29. Eberhardt, J.L., P.G. Davies, V.J. Purdie-Vaughns, and S.L. Johnson. 2006. Looking deathworthy: Perceived stereotypicality of black defendants predicts capital-sentencing outcomes. *Psychological Science* 12 (5): 383–386.
30. Greene, Joshua D., Fiery A. Cushman, Lisa E. Stewart, Kelly Lowenberg, Leigh E. Nystrom, and Jonathan D. Cohen. 2009. Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition* 111: 364–371.